WHICH NOUN PHRASES IS THE VERB SUPPOSED TO AGREE WITH? OBJECT AGREEMENT IN AMERICAN ENGLISH

Brian Dillon
University of Massachusetts

Adrian Staub
University of Massachusetts

Joshua Levy
University of Massachusetts

Charles Clifton, Jr.
University of Massachusetts

We investigate a noncanonical agreement pattern in American English in which a fronted wh-phrase appears to control agreement on an inflected auxiliary, as in *Which flowers are the gardener planting?* (Kimball & Aissen 1971). We explore this phenomenon with five acceptability-judgment experiments and interpret the resulting data with the aid of a quantitative model of the judgment process. Our study suggests that fronted wh-phrases interfere with agreement primarily as a function of their linear and structural position, and that this effect is not significantly modulated by overt case or thematic cues in off-line judgments. We suggest that our findings support a model of agreement processing in which syntactic phrases compete to control agreement on the basis of their structural and linear position with respect to the inflected verb.*

Keywords: agreement, syntax, psycholinguistics, agreement attraction, experimental linguistics, grammatical illusions

1. Introduction. Over the last several decades, the study of agreement phenomena in natural language has proven fruitful for both psycholinguistics and theoretical linguistics. In recent psycholinguistic traditions, agreement has been influential for the development of theories of language production (see Bock & Middleton 2011 and Franck 2011 for reviews) and more recently for the development of theories of syntactic comprehension (Clifton et al. 1999, Dillon et al. 2013, Pearlmutter et al. 1999, Tanner et al. 2014, Wagers et al. 2009). In the present article we contribute to this broad research program by investigating the effects that a fronted wh-object can exert on the perception of agreement dependencies within a sentence. Our study begins with an observation, first theoretically described by Kimball and Aissen (1971), that a fronted object in a wh-question appears to be able to license agreement on an inverted auxiliary. For example, the sentence in 1 seems acceptable (at least to some degree), despite the fact that the auxiliary *do* bears the same number specification as the object *What people*, instead of the subject *Muttonhead*.

(1) What people do Muttonhead address his words to? (Kimball & Aissen 1971:245)

Kimball and Aissen noted that similar effects arise in object relative clauses (e.g. *Mark knows the people who Clark think are in the garden*; Kimball & Aissen 1971:241). For object relative clause configurations, subsequent experimental research has verified and extended this observation using both production and comprehension tasks (Bock & Miller 1991, Clifton et al. 1999, Franck et al. 2010, Staub 2009, 2010, Wagers et al. 2009). There is relatively less work that has investigated how speakers treat configurations like 1.

In this article we evaluate a number of hypotheses about the mechanism by which a moved object comes to play a role in the computation of verbal agreement, focusing on root wh-question configurations like 1. There are three classes of explanation that we entertain. First, we consider whether this effect reflects a process that allows the fronted

* We would like to thank Lyn Frazier for extensive comments on an earlier draft of this manuscript, Colin Phillips and Jon Sprouse for insightful discussion of the modeling work presented here, and Lap-Ching Keung for his assistance in data collection.

WH-object to transmit its number feature to the subject (FEATURE-TRANSMISSION accounts). A second possibility is that agreement errors arise in 1 because speakers experience some degree of uncertainty about which phrase is the subject (SUBJECT-CONFUSION accounts), which might arise if speakers initially take the fronted WH-phrase to be a subject (Gorrell 1998).

The results we report lead us to doubt both feature-transmission and subject-confusion accounts of the object agreement in 1 and to advocate for a third possibility. Our results suggest that speakers determine the phrase that controls agreement by jointly considering the structural prominence and linear configuration of possible candidate controllers. On this view, the selection of an agreement controller is essentially a discrete and configurationally determined process, but in certain structural configurations speakers fall prey to a high degree of syntactic interference that prevents them from reliably identifying the target agreement controller, despite having an appropriate parse of the sentence. The results presented here thus support and extend the proposals of Franck and colleagues (Franck 2011, Franck et al. 2006, Franck et al. 2010), who have argued that the linear and syntactic arrangement of phrases in a sentence is a prime determinant of agreement errors, as well as accounts that emphasize the role of syntactic interference in working memory as a source of error in agreement processing (Badecker & Kuminiak 2007, Wagers et al. 2009).

The article is structured as follows. First we introduce feature-transmission accounts of agreement attraction. Next, we turn to a discussion of subject-confusion accounts of agreement attraction and show how these accounts have been offered as an explanation for the apparent object agreement in 1. We then present an alternative view of the source of noncanonical agreement in 1. We present five acceptability-judgment experiments that map out the conditions under which a fronted WH-object appears to control agreement in sentences like 1 and interpret these results with the aid of a quantitative model of the judgment process. We conclude that an in-depth look at the object-agreement effects in 1 supports the existence of a variable-choice agreement-controller selection process that is highly sensitive to configurational information, but relatively insensitive to overt thematic or case cues.

1.1. FEATURE-TRANSMISSION ACCOUNTS OF AGREEMENT ATTRACTION. Properties of the production and/or comprehension systems lead to characteristic deviations from canonical subject-verb agreement in English and other languages, a phenomenon broadly referred to as AGREEMENT ATTRACTION (Bock & Miller 1991). A widely cited example from Bock and Miller (1991) is given in 2.

(2) a. The key to the cabinets are getting rusty.
  b. The keys to the cabinet is getting rusty.

Bock and Miller (1991) showed that when a singular subject is modified by a prepositional phrase that contains a plural 'local' noun, speakers produce plural agreement at a rate that is higher than expected (see also Bock & Cutting 1992, Eberhard et al. 2005, Solomon & Pearlmutter 2004, Vigliocco et al. 1995, among others). One well-attested feature of this phenomenon is the plural asymmetry effect: plural local nouns as in 2a lead to a relatively high rate of attraction errors, whereas unmarked singular local nouns as in 2b do not reliably cause attraction errors (Bock & Miller 1991).

Some theories of this phenomenon ascribe the error in 2a to a feature-transmission process that takes the morphological number feature of the local noun *cabinets* and transmits it to the head noun *key*, where it then in turn controls verbal agreement. The MARKING-AND-MORPHING model of agreement production is one such theory (Eberhard et al. 2005; see also Bock & Middleton 2011). On this model, attraction effects arise because the subject NP in configurations like 2 bears a real-valued number feature that represents a weighted combination of the number values of all NPs contained within it. This SINGULAR-AND-PLURAL (SAP) value represents a graded degree of morphological plural marking on the subject NP, and this value controls the probability of selecting a plural or singular form of the verb during the course of production. A closely related account of the attraction errors in 2 is the feature-percolation account. This account holds that features may optionally percolate throughout a syntactic tree, and thus in 2a the plural feature of the local noun may probabilistically overwrite the number features of the head noun (Bock & Eberhard 1993, Franck et al. 2002, Vigliocco & Nicol 1998).

Attraction is not limited to local nouns contained within the subject NP: Hartsuiker, Antón-Méndez, and van Zee (2001) observed attraction from an object NP in SOV configurations in Dutch. This finding has led both percolation and gradient-number accounts of agreement attraction to allow for number features to spread to the subject NP from anywhere in the structure (Eberhard et al. 2005, Hartsuiker et al. 2001). Because existing feature-transmission accounts already admit the possibility that an object may pass its features to the subject, these accounts predict attraction in WH-object configurations in 1.

This model provides an explanation of some key features of agreement attraction. Attraction is primarily sensitive to the morphological specification of the local noun: notionally plural local nouns such as *group* do not generate attraction errors (see reviews in Bock & Middleton 2011, Eberhard et al. 2005). This is expected if attraction errors are the result of the spreading of morphosyntactic features. In addition, the attraction effect is not modulated by the linear distance between the local noun and the inflected verb. Instead, the structural distance between the local noun and the head noun determines the rate of attraction in these configurations (Bock & Cutting 1992, Bock & Miller 1991, Franck et al. 2002, Franck et al. 2010, Vigliocco & Nicol 1998; but cf. Gillespie & Pearlmutter 2013). This finding is expected if feature transmission occurs over a syntactic structure, rather than over a linear string of words.[1] Note that in feature-transmission models, it is generally assumed that the identification of the subject occurs without error.

To our knowledge, all feature-transmission accounts of attraction explain the number-markedness asymmetry effect in the same way. They assume a privative feature-coding scheme, whereby marked NPs (*the cabinets*) bear an overt [PLURAL] specification and unmarked NPs (*the cabinet*) simply have no featural representation of number (Eberhard 1997, Eberhard et al. 2005). Only plural nouns cause attraction, then, because only plural nouns are represented with a morphological number specification that can overwrite the features of the head noun.

If we extend these accounts to model the apparent object-agreement configuration in 1, then several predictions follow. The first is a strong plural asymmetry effect. The pri-

[1] Another variety of feature-transmission model attributes the attraction effects in 2 to the degree of semantic integration of the head and local noun during the planning process (Gillespie & Pearlmutter 2011, 2013, Solomon & Pearlmutter 2004). For present purposes, we note that these models have a number of properties similar to percolation and marking-and-morphing models. Specifically, semantic-integration models assume that the identification of the agreement controller is largely error free, and that agreement errors reflect a process whereby the local noun overwrites the number features of the subject. The semantic-integration account denies a critical role for syntactic structure in this process. Instead, feature overwriting is due to semantic integration of the two relevant nouns. The more integrated two nouns are, the more likely they are to be coactive in production, and thus the more opportunity there is for features of the local noun to corrupt the number representation of the subject.

vative number-coding mechanism invoked by these accounts predicts that object agreement should be possible with plural objects (3a), but not with singular objects (3b).

(3) a. Which flowers are the gardener planting?
b. Which flower is the gardeners planting?

Because the singular object in 3b bears no morphological representation of number, it should be unable to interfere in the agreement process. These models thus make the strong prediction that no attraction should arise in 3b. Kimball and Aissen's (1971) theoretical model makes a similar prediction, for essentially the same reason. Because they also used privative coding of number in their account, their model shares the prediction that singular WH-objects should not interfere with the agreement process at all.

Further predictions about the effect in 1 depend on the specifics of the feature-spreading mechanism. In the marking-and-morphing model, all NPs deterministically contribute to a real-valued SAP number that is marked on every subject NP. On this model, the subject in the object-agreement configurations like 1 or 3 would bear an intermediate number value that represents a blending of the number features of the subject and the object. Importantly, the subject receives such an intermediate value for number every time such a configuration is processed. Staub (2009, 2010) argued that this intermediate representation of number predicts that speakers should be characteristically slower or less confident in computing agreement in attraction configurations. This is because as the representation of the number on the subject becomes more ambiguous between singular and plural, the decision about which verbal inflection to select should grow more difficult. Staub showed that this prediction was borne out for intervening attraction configurations as in 2a.

If a gradient mechanism like SAP is responsible for the object agreement in 1 and 3, then we predict that the acceptability of these sentences should be gradiently impacted by feature mismatch between the object and subject. In the context of an acceptability-judgment task, this means that participants should perceive 1 as neither fully acceptable nor fully unacceptable. Instead, it should receive a rating closer to the middle of the scale. If, however, the feature-transmission process in 1 is better modeled with a discrete mechanism like feature percolation, substantial trial-to-trial variation should be in evidence. Since feature percolation may or may not happen on any given presentation of a sentence, there should be a mixture of responses to sentences like 1. When the object's plural feature has percolated to the subject, it should be perceived as fully unacceptable. When it has not, it should be perceived as fully acceptable.

1.2. SUBJECT-CONFUSION ACCOUNTS. Feature-transmission accounts of agreement attraction focus on how the subject NP could come to bear an erroneous number specification. In general, these models implicitly assume that the selection of the agreement controller in attraction configurations proceeds without error. However, it is also possible that errors in 1 and 2 could arise because speakers are momentarily confused about which phrase is the subject. We refer to theories that explain attraction errors in this way as subject-confusion accounts.

For example, Gorrell (1998) offers a treatment of this effect in terms of misanalysis of the initial WH-object. In an incremental parse, the initial WH-object may well be treated as a subject. If perceivers initially take the WH-phrase as a subject, then they should enforce an agreement relation between the WH-phrase and the auxiliary. Temporary misanalysis, coupled with incremental processing, will lead to a perception of ill-formedness whenever the WH-phrase and the auxiliary fail to agree in number. On this view, the controller for agreement is misidentified because the grammatical roles for each of the arguments in the sentence have been incorrectly assigned.

Bock and Miller (1991) suggest that subject confusion of a similar sort is responsible for a subset of attraction errors. They suggested that temporary confusion about which NP is the subject of the sentence gives rise to apparent agreement attraction effects in nonintervening attractor configurations like 4.

(4) The politicians that the flag were …

Bock and Miller argued that in sentences like this, the agreement error reflected 'subject misidentification'. Their argument was based on the finding that the nonintervening agreement error in 4 is sensitive to the animacy of the nouns involved: when the head of the relative clause is animate, and thereby more subject-like, more attraction errors are observed. In contrast, animacy does not impact the amount of agreement attraction contributed by the local noun in configurations like 2 above. Furthermore, they observed that continuations of sentence preambles as in 4 suggested that subjects had incorrectly interpreted the politicians as the subject of the embedded clause when an agreement error was made. An inspection of the continuations for canonical attraction configurations (2a) found no such relation. Staub (2010) also argued that agreement errors in configurations like 4 reflected subject misidentification. This argument was based on a distributional analysis of response times (RTs) in a timed agreement-selection task (see Staub 2010 for details). He observed that RTs for correct agreement responses in the configuration in 2 involved a shift in the location and the skew of the RT distribution, whereas nonintervening attractors as in 4 resulted in an effect almost exclusively in the tail of the RT distribution. This pattern suggests that the attraction in configurations like 4 is driven by a subset of difficult trials, exactly the kind of trial-to-trial variation predicted by the 'subject misidentification' account.

Subject-confusion accounts share a key central prediction: the less subject-like the local noun is, the less likely it is to interfere with agreement processes. This is because making the attractor less subject-like decreases the probability of misanalyzing the structure, which should in turn reduce or eliminate the possibility of agreement errors.

1.3. SYNTACTIC-INTERFERENCE ACCOUNTS. A third possibility is that agreement errors arise not because the structure has been misanalyzed, or because the true agreement controller bears an erroneous number specification, but because an error arises when producers or comprehenders search a syntactic parse for a phrase to license verbal agreement. Badecker and Kuminiak (2007; see also Badecker & Lewis 2007) offer one such account. They proposed that agreement-controller selection proceeds using a cue-based memory-retrieval mechanism. On this view, in order to determine the number specification for an inflected verb, a producer must search working memory for a syntactic representation of the subject in order to copy its features onto the verb. A cue-based retrieval mechanism searches working memory by matching retrieval cues (i.e. the features of the desired representation) against the contents of memory and selecting a representation for processing on the basis of its match with the retrieval cues. This leads to agreement attraction in configurations like 2, because the local noun bears some subject-like features, such as its preverbal position and its lack of overt objective case. Because it partially matches the retrieval cues for an agreement controller, the retrieval process may erroneously identify the local noun as the controller. Badecker and Kuminiak supported this hypothesis by showing that in Slovak, only local nouns that were homophonous with nominative case caused agreement attraction; local nouns that were unambiguously marked as accusative case did not. This is predicted on the retrieval-based account, because the local nouns that were superficially compatible with nominative case are arguably more subject-like than those with unambiguous objective case.

Wagers, Lau, and Phillips (2009) provided further arguments in favor of memory-retrieval models of agreement attraction. They looked at agreement attraction in comprehension, where it is possible to investigate the influence a local noun has in normatively grammatical sentences (*the key to the cabinets is rusty*), as well as normatively ungrammatical sentences (*the key to the cabinets are rusty*). One central finding of their study is that the influence of local nouns is limited to ungrammatical sentences in reading time and speeded judgment measures: ungrammatical agreement can be reliably 'rescued' by a feature-matched local noun, but grammatical agreement is not made less acceptable or more difficult by a feature-mismatched local noun. For example, self-paced reading times at *was* in examples like 5a were not affected by the number of the local noun, and in a speeded acceptability task, the rate of 'acceptable' responses was not impacted by the number of the local noun. In contrast, the number of the local noun in examples like 5b had a significant impact on reading times and acceptability rates. When the local noun was plural, reading times were faster and the sentence was judged more acceptable (for similar findings with eye-tracking, see Dillon et al. 2013; for similar findings with event-related potentials (ERPs), see Tanner et al. 2014).

(5) a. The key to the cabinet(s) was rusty from years of disuse.
 b. *The key to the cabinet(s) were rusty from years of disuse.

The grammatical asymmetry effect has been offered as evidence for a cue-based retrieval process to identify the agreement controller (Dillon et al. 2013, Tanner et al. 2014, Wagers et al. 2009). In grammatical sentences like 5a, the cues contributed by the verb *was* provide a very good match to *key*, leading to a high probability that it will be retrieved as an agreement controller. The local noun *cabinets* is a very poor match to the cues in the verb and thus is very unlikely to interfere in the agreement process. In ungrammatical sentences like 5b, however, the cues contained in the verb *were* are an imperfect fit both to the head noun *key* and to the local noun *cabinets*: syntactic-configurational cues match *key*, but number-feature cues match *cabinets*. Agreement attraction thus arises in this configuration because the retrieval cues underdetermine which phrase is the agreement controller, creating the possibility that the local noun *cabinets* is incorrectly selected to control agreement.

Finally, Franck and colleagues (Franck 2011, Franck et al. 2006, Franck et al. 2010) offer a similar account of agreement attraction errors in terms of syntactic interference, broadly construed. Franck and colleagues argue that the notions of c-command and linear precedence are the prime determinants of agreement attraction. In their account, agreement attraction in examples comparable to 4 is the result of a syntactic trace associated with the moved object that intervenes hierarchically between the subject and the verb. They argue that the object trace is in a specifier-head relationship with the inflected verb at an intermediate level of representation. Because there is an intermediate level of representation that puts the object in a position that is structurally quite close to the verb (the object both c-commands and is in a specifier-head relationship with the verb), the likelihood of increasing syntactic interference is increased, and agreement-attraction errors are more likely.

1.4. The current study: agreement with wh-objects. As the preceding review makes clear, existing theories of agreement processing offer many different explanations for how the wh-object could come to influence agreement in configurations like 1. It is possible that the wh-object is structurally close enough to the subject to transmit its feature values to it, as in, for example, marking-and-morphing or feature-percolation models. It is possible that the wh-object influences agreement processing simply because of the temporary ambiguity of the initial wh-object: if it is initially parsed as a subject, it should then control agreement. Lastly, the fronted WH-object bears many subject-like features: it is in a preverbal specifier position and does not have overt object case. So the effect in 1 may arise because the WH-object has many subject-like features, and so has a high probability of being retrieved when comprehenders search a syntactic parse for a phrase to control agreement.

In the present article we explore the effect in 1 using acceptability judgments to better understand the nature of this phenomenon. To preview, the results of our investigation lend broad support to the syntactic-interference view of agreement errors, although the details of our results suggest some amendments to existing models of syntactic interference. Specifically, we argue that object agreement arises in 1 during the stage where a controller is identified, a claim we share with memory-retrieval accounts of agreement attraction phenomena. However, our results suggest that controller selection depends primarily on configurational factors—the linear and structural arrangement of NPs in the sentence—and is relatively less sensitive to thematic or case cues to subjecthood. The WH-object in 1 is a tempting controller for agreement because it is linearly adjacent to the auxiliary and is in a close syntactic relationship with this auxiliary (e.g. a specifier-head relationship).

In experiment 1, we establish the effect using a simple off-line acceptability-judgment task. One goal of this first experiment was to determine whether acceptability judgments provide an appropriate tool for investigating these effects. In addition, we develop an explicit model of the judgment process to derive quantitative predictions about the distribution of acceptability judgments in our experiment. To preview the results of our study, experiment 1 does demonstrate an effect of a mismatching WH-object in both grammatical and ungrammatical sentences. Experiments 2 and 3 then focus on the extent to which this effect is alleviated when unambiguous cues to the grammatical roles of the object (in experiment 2) or the subject (in experiment 3) are provided. Experiments 4 and 5 focus on the potential role of linear adjacency or proximity between the WH-object and the verb.

2. Experiment 1. Experiment 1 tests whether a mismatch in features between a fronted WH-object and the subject influences acceptability in off-line acceptability judgments using a seven-point rating scale. The primary goal of experiment 1 is to establish the utility of off-line judgment methodology for investigating agreement-error phenomena. Specifically, we ask whether a number mismatch between the subject NP and a fronted WH-object causes a drop in acceptability ratings in sentences that are structurally identical to 1. Second, we ask whether the magnitude of the mismatch effect is the same for singular and plural subjects. In production and comprehension studies on agreement attraction phenomena, it is widely observed that mismatch effects are more pronounced for singular subjects than for plural subjects. Lastly, we ask whether any mismatch effects impact ungrammatical and grammatical sentences to the same degree. In reading studies and speeded judgment studies on agreement attraction phenomena, mismatch effects seem to be more pronounced in ungrammatical sentences (Dillon et al. 2013, Tanner et al. 2014, Wagers et al. 2009). However, the grammatical asymmetry effect seems to be less robust for off-line judgment methodology (Häussler & Bader 2009, Nicol et al. 1997), and even for on-line comprehension measures, some studies have shown a reliable mismatch effect for grammatical sentences (Franck et al. 2015).

2.1. Method.

Participants. For all five experiments, speakers of American English were recruited using Amazon's Mechanical Turk (https://www.mturk.com). Participation was restricted to workers with IP addresses in the United States. Prior to participation in the experiment, participants filled out a questionnaire aimed at assessing their language background. Participants gave informed consent and were paid $1 for their participation in the experiment. In experiment 1, forty-four participants were recruited this way. All participants reported that English was their native language. Completion of the survey took approximately eight minutes on average.

Materials. We developed sixteen sets of experimental items that consisted of the eight experimental conditions shown in Table 1. Experiment 1 had a fully crossed 2 × 2 × 2 design with the factors subject/object number match (match or mismatch), subject number (singular or plural), and grammaticality (grammatical or ungrammatical), resulting in two observations per condition per subject. All experimental items were object wh-questions in the present progressive or in a present form of a passive structure. All items thus had a present-tense form of *be* as the agreement-bearing auxiliary in the question. Across items, there was variation in the grammatical role of the wh-object (direct or indirect object), as well as variation in the animacy of the head of the wh-phrase. Of our sixteen items, one was removed after data collection in experiment 1 due to an unforeseen ambiguity, leaving fifteen item sets for analysis.

In all five experiments reported here, the sixteen critical experimental items were combined with the same sixteen filler sentences. Fillers included both grammatical and ungrammatical sentences. Unacceptable items included unlicensed negative polarity item dependencies and ungrammatical sequences of tense.

match	subject	gram	label	
Match	Singular	Gram	SSS	Which tree is the gardener planting?
Mismatch	Singular	Gram	PSS	Which trees is the gardener planting?
Match	Plural	Gram	PPP	Which trees are the gardeners planting?
Mismatch	Plural	Gram	SPP	Which tree are the gardeners planting?
Mismatch	Singular	Ungram	*PPS	*Which trees are the gardener planting?
Match	Singular	Ungram	*SPS	*Which tree are the gardener planting?
Match	Plural	Ungram	*SSP	*Which tree is the gardeners planting?
Mismatch	Plural	Ungram	*PSP	*Which trees is the gardeners planting?

Table 1. Illustration of material in experiment 1. ‘Label’ refers to number (Singular/Plural) of object, auxiliary, and subject respectively.

Procedure. The critical sentences were distributed into eight Latin-square lists, and each participant was assigned to a different list. For each participant, the order of presentation was randomized, subject to the constraint that no two critical experimental items appeared adjacent to each other. The questionnaire was administered over the internet using the IbexFarm experimental software (http://spellout.net/ibexfarm). Participants were instructed to rate sentences on a 1–7 Likert scale according to what sounded like ‘acceptable’ English and were given three examples to introduce them to the task. Each trial consisted of a single sentence presented on the screen, and the seven response options were listed below the sentence. Participants responded by choosing their desired rating using either the mouse or the number keys on the keyboard. To remind participants of the interpretation of the Likert scale, the left end of the scale was labeled ‘Bad’, and the right end of the scale was labeled ‘Good’. Participants were allowed to take as long as they liked to judge the sentences. Reaction times were recorded, but not analyzed.

2.2. Results. Average ratings for all experimental conditions are presented in Table 2, along with by-participant standard errors corrected for between-participant variance

(Bakeman & McArthur 1996). The rating scores were analyzed using separate by-participant (F_1) and by-item (F_2) repeated-measures ANOVAs.

	GRAMMATICAL		UNGRAMMATICAL	
	singular subject	plural subject	singular subject	plural subject
MATCH	6.7 (±0.13)	6.6 (±0.15)	2.6 (±0.14)	3.2 (±0.17)
MISMATCH	4.2 (±0.23)	5.6 (±0.21)	4.6 (±0.21)	3.6 (±0.22)

TABLE 2. Means of ratings, experiment 1. Parentheses represent standard error corrected for between-participant variance, following Bakeman and McArthur (1996).

Statistical analysis revealed significant main effects of grammaticality ($F_1(1,43) = 187.2$, $p < 0.05$; $F_2(1,14) = 172.1$, $p < 0.05$), subject number ($F_1(1,43) = 5.7$, $p < 0.05$; $F_2(1,14) = 3.68$, $p < 0.1$), and subject/object mismatch ($F_1(1,43) = 12.9$, $p < 0.05$; $F_2(1,14) = 13.0$, $p < 0.05$). These main effects were qualified by a three-way interaction ($F_1(1,43) = 51.4$, $p < 0.05$; $F_2(1,14) = 32.5$, $p < 0.05$) and by significant interactions of subject/object mismatch with grammaticality ($F_1(1,43) = 79.6$, $p < 0.05$; $F_2(1,14) = 136.4$, $p < 0.05$) and subject number with grammaticality ($F_1(1,43) = 8.4$, $p < 0.05$; $F_2(1,14) = 10.8$, $p < 0.05$).

To understand the nature of our three-way interaction, we computed planned comparisons on the effect of subject/object mismatch (the MISMATCH EFFECT) for each level of subject number and grammaticality. Planned comparisons revealed that subject/object mismatch led to significant decreases in acceptability for grammatical sentences. This was significant both for singular subjects ($t_1(43) = -8.6$, $p < 0.05$; $t_2(14) = -9.6$, $p < 0.05$; 95% CI = [−3.1, −2.0])[2] and for plural subjects ($t_1(43) = -3.9$, $p < 0.05$; $t_2(14) = -4.2$, $p < 0.05$; 95% CI = [−1.6, −0.5]). In contrast, subject/object mismatch led to a reliable increase in acceptability for ungrammatical sentences. In pairwise comparisons this was significant only for singular subjects ($t_1(43) = 7.2$, $p < 0.05$; $t_2(14) = 5.7$, $p < 0.05$; 95% CI = [1.4, 2.5]); there was no statistically reliable mismatch effect for ungrammatical sentences with plural subjects.

2.3. DISCUSSION. In experiment 1, we observed reliable mismatch effects in off-line acceptability ratings. For grammatical sentences, a mismatch in number features between the number of the subject and of the object led to a decrement in acceptability. For ungrammatical sentences, a mismatch in the features of the subject and object led to an increase in acceptability. Overall, this pattern suggests that the number features of the object exert a reliable effect on intuitive acceptability of these sentences, lowering acceptability when they mismatch the number features of the verb, and increasing acceptability when they match the features of the verb. This confirms the intuitive observations of Kimball and Aissen (1971). In addition, we observed the expected effects of grammaticality, that is, lower ratings for ungrammatical sentences.

Planned comparisons revealed that not all of the mismatch effects were of equal magnitude. In particular, the absolute magnitude of the mismatch effect is larger when the subject is singular than when it is plural, for both grammatical and ungrammatical sentences alike. In fact, for plural subjects we observed a significant mismatch effect only in grammatical sentences; this effect did not reach significance in ungrammatical sentences. Interestingly, we also observed a significant interaction of subject number and grammaticality, reflecting a smaller effect of grammaticality for plural subjects than for singular subjects.

[2] All confidence intervals (CIs) reported in this article are by participant.

The results of experiment 1 show mismatch effects for grammatical and ungrammatical sentences. Subject/object mismatch degrades the acceptability of otherwise grammatical sentences, and improves the acceptability of otherwise ungrammatical sentences. This finding distinguishes the off-line judgments of object-auxiliary mismatch from other well-studied cases of agreement attraction in real-time comprehension measures, which have been argued to exhibit substantial mismatch effects only in ungrammatical sentences (Dillon et al. 2013, Tanner et al. 2014, Wagers et al. 2009). Like agreement attraction, the number of the subject plays a role in gating the mismatch effect: plural subjects are associated with smaller mismatch effects both in grammatical and ungrammatical sentences. However, there was a statistically reliable mismatch effect for plural subjects for grammatical sentences, suggesting that plural subjects merely reduce, rather than eliminate, the mismatch effect.

The results of experiment 1 establish that the acceptability of WH-object questions is adversely impacted if either the subject or the object in the sentence does not match in number features with the verb. However, we have thus far only shown this to hold at the level of mean ratings per condition. This finding is in fact compatible with at least two distinct distributions of acceptability ratings. One possibility is that participants in our experiment were categorically classifying examples such as 1 as either essentially grammatical or essentially ungrammatical on any individual trial. On this model, the intermediate mean rating would reflect a mixture of grammatical and ungrammatical responses whose average lies in between these two extremes. However, it is also possible that object mismatch has a more gradient effect on acceptability judgments, such that the acceptability of a sentence like 1 is truly perceived as intermediate between fully grammatical and ungrammatical. In this case, we expect the distribution of responses to reflect an overall shift in the location of the acceptability responses, with individual acceptability judgments in mismatch conditions being pushed toward the middle of the response scale. Because the nature of the response distribution can place strong constraints on a possible account of the mismatch effect, it is of theoretical interest for present purposes. In the next section we turn to the development of a computational model of acceptability judgments in order to gain a more fine-grained understanding of the distribution of response scores in experiment 1.

3. A computational model of the judgment process.

3.1. Method.

In this section we develop and evaluate quantitative models to attempt to answer the following question: Does the subject/object mismatch effect reflect discrete or gradient agreement processes? A priori, it seems that the central critical and empirical criterion for distinguishing these two possibilities is the proportion of 'extreme' responses on the acceptability scale provided in experiment 1. As a first approximation, it seems that a discrete model predicts relatively many extreme responses (1s and 2s, and 6s and 7s), reflecting judgments that categorically alternate between 'grammatical' and 'ungrammatical'. In contrast, a gradient model predicts many more in-between responses and few extreme responses (e.g. mostly 3s, 4s, and 5s).

To make firm predictions about these two different types of response behavior, we formulate them as quantitative models in the following fashion. The DISCRETE MODEL of the responses assumes that every response to partial match conditions is an independent draw from either an ungrammatical reference distribution OR a grammatical reference distribution. This model reflects categorical response behavior, with participants treating every sentence in the experiment as either 'grammatical' or 'ungrammatical' on any given trial. In contrast, the GRADIENT MODEL assumes that each individual response in the partial match conditions is a weighted combination of draws from the grammatical and the ungrammatical reference distributions. On this model, each experimental observation in mismatch conditions is drawn from a distribution that lies between the grammatical or ungrammatical extremes. In other words, this model implements the response behavior expected if the mismatch conditions reflect 'gradient grammaticality', or a continuous downward shift in acceptability that affects all trials. To determine which model better captures the response behavior of our participants, we generated quantitative predictions about the expected distribution of responses for each model and compared them against the distributions observed in experiment 1. Full model details are available in Appendix A.

3.2. RESULTS. We apply these models to the group-level data from experiment 1 by pooling observations across participants within conditions. The resulting empirical response distributions across conditions are plotted in Figure 1, which plots the grammatical and ungrammatical reference distributions. For conditions with a singular verb, the responses to SSS form the grammatical reference distribution, and the responses to *PSP form the ungrammatical one. Likewise, for conditions with a plural verb, we take PPP as the grammatical reference distribution, and *SPS as the ungrammatical one.

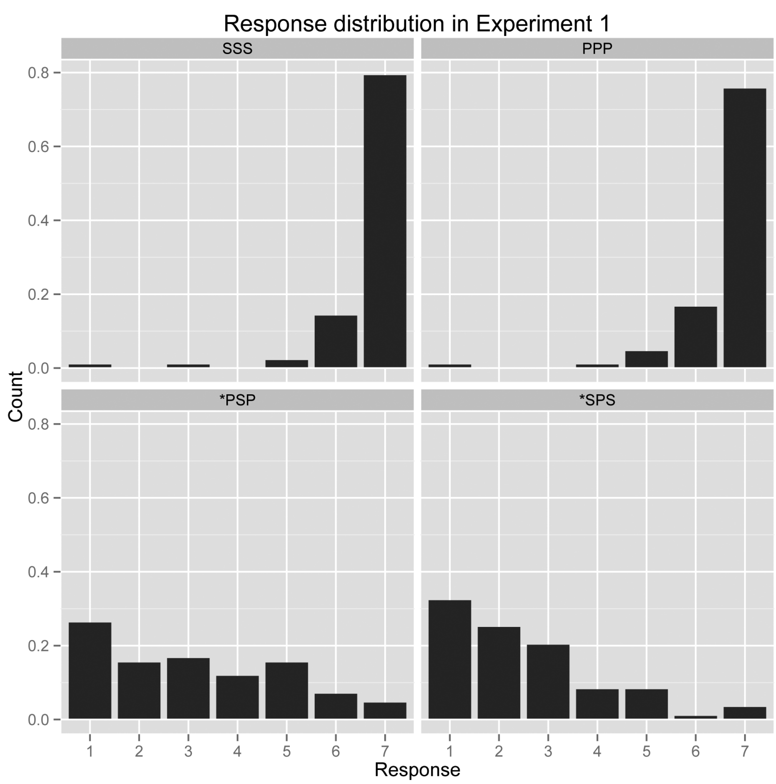

FIGURE 1. Group-level response distribution in experiment 1 for the four reference distributions. The top two panels are the grammatical reference distributions, and the bottom two panels are the ungrammatical reference distributions.

We fit discrete and gradient models to each of the subject/object mismatch conditions (PSS, SPP, *PPS, *SSP) separately. An evaluation of the model fits is presented in Table 3. For each modeling comparison reported in the text, we ran 100 Monte Carlo simulations of our models, and we report summary statistics calculated over these simulations. Despite having very different response distributions, both the discrete and the gradient models achieved a close fit between the expected mean response value on the theoretical distribution and the observed mean rating for each of the four test conditions. Figure 2 presents each mismatch condition's empirical response distribution, along with the average predicted response distributions across all Monte Carlo trials. In all cases, the discrete model provides a closer fit to the observed response distribution than does the gradient model. Surprisingly, the gradient model does predict response patterns that show hints of bimodality: its predicted response direction for the PSS condition, for example, presents relative modes at 3 and 6, although the reason for this is unclear. Nonetheless, this surprising finding provides useful methodological insight. In general, it may not be sufficient to search for multiple modes in the response distribution to diagnose bimodal response behavior. Our results suggest that the response distribution may not be directly interpretable without a quantitative response model.

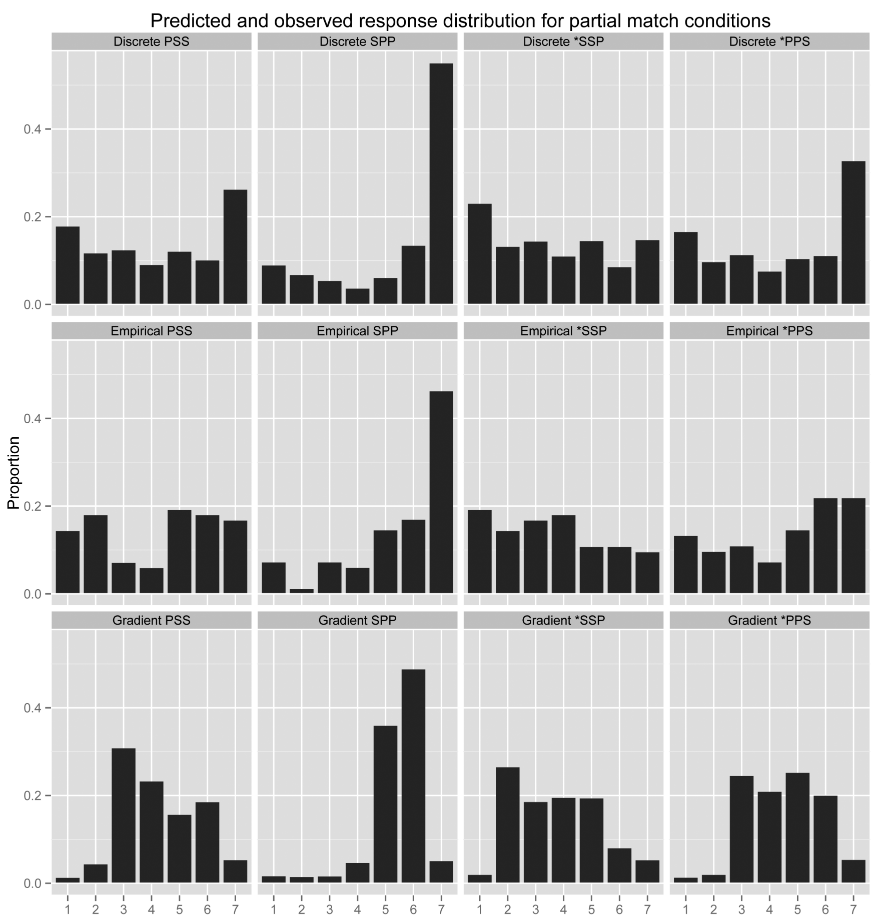

FIGURE 2. Predicted and observed response proportions for each subject/object mismatch condition. Predicted distributions represent the average over all Monte Carlo trials.

Model-comparison metrics reveal a distinct advantage for discrete response models over gradient response models. Here we present the difference in the BAYESIAN INFORmation criterion (BIC; see Appendix A), a model-selection metric that allows us to evaluate the goodness of fit for the discrete model and the gradient model. Negative BIC difference scores indicate an advantage for the discrete model over the gradient model. A difference in BIC scores greater than 10 provides strong evidence in favor of one model over another (Kass & Raftery 1995). In all comparisons, the average BIC advantage for the mixture model is clearly greater than this suggested criterial value. The Bayes factor (BF) model comparisons also reveal a substantial advantage for the discrete model. A Bayes factor of 100 is considered to be 'decisive' evidence in favor of the better-fit model (Jeffreys 1961). In our case, the Bayes factors in favor of the discrete model are considerably in excess of this threshold.

		gradient model		discrete model			
	empirical mean	π	predicted mean	π	predicted mean	$BIC_{Discrete} - BIC_{Gradient}$	BF
PSS	4.2	.31	4.3	.30	4.2	−94.8 [−152.3, −47.6]	1.2×10^{31}
SPP	5.6	.73	5.4	.77	5.5	−145.5 [−263.1, −63.9]	1.4×10^{55}
*SSP	3.6	.16	3.7	.13	3.7	−53.8 [−104.9, −8.9]	6.1×10^{20}
*PPS	4.5	.39	4.5	.40	4.5	−79.2 [−138.9, −29.8]	1.4×10^{28}

Table 3. Quantitative analysis of two response models. Values represent the average over 100 Monte Carlo simulations of model. Square brackets represent the full range of values observed across Monte Carlo trials.

3.3. Discussion. The quantitative models presented here suggest that for partial matching conditions—those conditions where either the subject or the object mismatched the number of the verb—the responses observed in the experiment are better fit by a discrete response model than by a gradient response model. We take this result to mean that in experiment 1, participants behaved in an essentially categorical fashion on a trial-by-trial basis: they registered responses to these conditions that could be classified as 'ungrammatical' or 'grammatical'. The modeling results provide strong evidence against a gradient grammaticality view of our mismatch conditions.

This result places significant constraints on any possible account of the mismatch effect. In particular, it provides some evidence against a subject-confusion (syntactic-misanalysis) view of the mismatch effect. To see this, it is necessary to further specify a mechanism by which an incorrect parse of the sentence can influence acceptability judgments. One plausible way this could happen is if on some significant proportion of trials, comprehenders misanalyze an object mismatch configuration as a subject wh-question, recover, and are left with a lingering perception of unacceptability or difficulty from the reanalysis process, which causes them to assign a lower acceptability score to that token. On this model, most mismatching trials would reflect an overall downward shift in the response distribution when compared to the relevant grammatical reference distribution. In other words, this model makes the same distributional predictions as the gradient model. For this reason, this implementation of the misanalysis model is not supported by the distributional evidence.

Instead, in order to maintain a temporary-misanalysis account of the mismatch effect that is consistent with the modeling evidence here, one would have to stipulate that on some proportion of trials, comprehenders completely fail to reanalyze their initial parse and so judge the sentence as if it were completely ungrammatical. On the remaining portion of trials, reanalysis is complete, and the failed parse exerts no influence on the eventual judgment. On this model, the rate of failed reanalysis must be equal to $1 - \pi$ in the discrete model in order to account for the data. Thus, in order to fit the data from the

PSS condition, one needs to assume that comprehenders fail to obtain the correct parse in these configurations on 70% of trials. We believe that it is highly implausible that comprehenders fail to recover a grammatical parse of our stimulus sentences on roughly three quarters of trials. For this reason, we take the modeling results to constitute an argument against the misanalysis account.

Of course, we cannot categorically rule out the misanalysis account on the basis of this result. It is possible that a more complicated model of our response distribution, such as a mixture of ungrammatical, grammatical, and 'reanalyzed' trials near the middle of the scale, could capture our results. However, without making unmotivated assumptions about the distribution of response scores in the case of reanalysis, such a model is difficult to build and evaluate. It therefore remains to be shown that a reanalysis account could be made to fit these data, although we do not believe it likely.

Although we believe that the present modeling results are compelling evidence in favor of a discrete response model, there are some limitations of these results that must be acknowledged. First, although the results indicate discrete variation in the response distributions in experiment 1, we cannot yet say whether this variation reflects inter- or intra-individual variation. That is, the results may reflect differences in the individual grammars of the participants in our experiment, or they may reflect stochastic variation in the agreement choices made by each speaker. Further experimentation is necessary to tease apart these two possibilities.

A second limitation of our modeling approach is that it assumes a linear relationship between the underlying acceptability percept and the response scale: the acceptability of the mismatching conditions is modeled as a linear combination of samples from the response distributions. It is not obvious that there is such a linear mapping between acceptability and response. If responses are nonlinearly related to acceptability, then the strength of these results is unclear. To our minds, there is one plausible nonlinear mapping that would be fatal to the argument presented here: a strong bias toward extreme response options even with sentences of intermediate acceptability. This could arise if participants are simply reluctant to give ratings in the middle of the response scale. However, we think it is unlikely that such a nonlinear mapping between acceptability and response holds generally of judgments on Likert scales. An inspection of data from other acceptability judgments run in our labs, as well as the data reported by Sprouse, Schütze, and Almeida (2013), reveals many response distributions that are characterized by a clear mode in the center of the response scale, with few extreme responses, similar to the distributional predictions of the gradient model. Because it is in principle possible to observe distributions such as those predicted by the gradient model, it is unlikely that a simple response bias is driving the results here. Of course, we cannot categorically rule out the possibility that another nonlinear mapping between response and acceptability exists. However, in the absence of a plausible nonlinear mapping between acceptability and response that would undermine the argument presented here, we take the modeling results as initial confirmation of a discrete model of the acceptability judgments to these sentences.

4. Experiment 2. In experiment 1, we observed that off-line acceptability judgments reliably index the mismatch effect. This investigation revealed several important findings. One, there was no reliable grammatical asymmetry in this measure; indeed, the mismatch effect was smaller for ungrammatical sentences than for grammatical sentences, the opposite pattern of what is seen in reading time and speeded acceptability measures (e.g. Wagers et al. 2009). Further, we observed that the mismatch effect was attenuated, but not eliminated, for plural subjects. Lastly, the results of our modeling support the view that the mismatch phenomenon is essentially discrete, since a sentence with a mismatching WH-object was treated as either essentially grammatical or essentially ungrammatical on each trial. The discrete nature of the phenomenon is evidence against graded feature-transmission accounts of the mismatch effect (e.g. Eberhard et al. 2005). Further, we argued that the modeling results constitute an argument against the hypothesis that the mismatch effect is due in large part to confusion about which phrase is the subject.

In experiment 2 we provide a further test of such subject-confusion accounts of the mismatch effect. These accounts predict that when the grammatical role of the WH-filler is unambiguous, the mismatch effect should be eliminated. We test this prediction by using preposition pied-piping in an acceptability-judgment task. When a displaced WH-phrase is fronted along with its preposition, as in *To which players is the coach speaking?*, there is no ambiguity about the grammatical role of the initial WH-phrase. The ambiguity of the initial WH-phrase is restored when the preposition is stranded in its original position, as in *Which players is the coach speaking to?*. Thus subject-misidentification accounts of the mismatch effect predict that it should not survive when the preposition is fronted along with the WH-phrase.

4.1. Method.

Participants. In experiment 2, forty participants were recruited. Completion of the survey took approximately nine minutes on average.

Materials. Based on the materials of experiment 1, we developed sixteen sets of experimental items that consisted of the eight experimental conditions shown in Table 4. Experiment 2 had a fully crossed 2 × 2 × 2 design within subjects, with the factors subject/object match (MATCH or MISMATCH), pied-piping (FRONTED or IN SITU), and grammaticality (GRAMMATICAL or UNGRAMMATICAL). In addition, we manipulated the number of the subject (SINGULAR or PLURAL) between items, such that half of the experimental sentences in any given list had singular subjects, and half had plural subjects. One item was removed after data collection due to an unforeseen ambiguity.

MATCH	PIED-PIPING	GRAM	
Match	In situ	Gram	Which tree is the hiker resting under?
Mismatch	In situ	Gram	Which trees is the hiker resting under?
Match	Fronted	Gram	Under which tree is the hiker resting?
Mismatch	Fronted	Gram	Under which trees is the hiker resting?
Mismatch	In situ	Ungram	*Which trees are the hiker resting under?
Match	In situ	Ungram	*Which tree are the hiker resting under?
Mismatch	Fronted	Ungram	*Under which trees are the hiker resting?
Match	Fronted	Ungram	*Under which tree are the hiker resting?

Table 4. Illustration of material in experiment 2 with a singular subject.

Procedure. The procedure was identical to experiment 1.

4.2. Results. By-condition average ratings for all experimental conditions are presented in Table 5. Repeated-measures ANOVAs revealed a significant main effect of grammaticality ($F_1(1,39) = 70.2$, $p < 0.05$; $F_2(1,14) = 124.5$, $p < 0.05$), and an effect of subject/object mismatch that was significant by participants only ($F_1(1,39) = 4.4$, $p < 0.05$; $F_2(1,14) = 2.4$, $p = 0.14$). These main effects were qualified by significant two-way interactions of grammaticality and mismatch ($F_1(1,39) = 21.7$, $p < 0.05$; $F_2(1,14) = 16.9$, $p < 0.05$) and subject number and mismatch ($F_1(1,39) = 33.0$, $p < 0.05$;

$F_2(1,14) = 12.3$, $p < 0.05$). There was an interaction of subject number and grammaticality that was significant by participants only ($F_1(1,39) = 6.2$, $p < 0.05$; $F_2(1,14) = 2.8$, $p = 0.12$). Lastly, we observed a significant three-way interaction of subject number, grammaticality, and mismatch ($F_1(1,39) = 11.2$, $p < 0.05$; $F_2(1,14) = 6.0$, $p < 0.05$). No other effects were significant.

SINGULAR SUBJECTS	GRAMMATICAL		UNGRAMMATICAL	
	in-situ P	fronted P	in-situ P	fronted P
MATCH	6.5 (±0.18)	6.0 (±0.21)	3.8 (±0.26)	3.3 (±0.25)
MISMATCH	4.3 (±0.24)	4.6 (±0.26)	4.0 (±0.27)	3.6 (±0.29)
PLURAL SUBJECTS	GRAMMATICAL		UNGRAMMATICAL	
	in-situ P	fronted P	in-situ P	fronted P
MATCH	6.1 (±0.18)	5.6 (±0.23)	3.3 (±0.26)	3.3 (±0.25)
MISMATCH	6.0 (±0.24)	5.6 (±0.23)	3.7 (±0.27)	4.0 (±0.29)

TABLE 5. Means of ratings, experiment 2. Parentheses represent standard error corrected for between-participant variance, following Bakeman and McArthur (1996).

4.3. DISCUSSION. In experiment 2, we failed to find any reliable effect of preposition fronting on our ratings. Preposition fronting did not significantly interact with any of the other experimental factors. Our experimental results thus confirm the intuitions reported by Kimball and Aissen (1971), who claimed that the effect was not modulated by preposition fronting. Although these results provide evidence that the mismatch effect survives preposition fronting, we cannot make the stronger claim that it has no effect whatsoever on the magnitude of the mismatch effect. We note that the manipulation of preposition fronting does appear to have impacted the size of the mismatch effect for grammatical conditions with singular subjects, such that the mismatch effect was slightly smaller for configurations with fronted prepositions (95% CI = [0.73, 2.1]) than for configurations with in-situ prepositions (95% CI = [1.6, 2.9]). Despite this, our results show that preposition fronting does not have a consistent effect on the size of the mismatch effect.

We further note that experiment 2 replicates some of the core features of the object mismatch effect that were observed in experiment 1. First, we observed mismatch effects for both grammatical and ungrammatical sentences, though the mismatch effect was on average smaller for ungrammatical sentences. We also observed that the mismatch effect was smaller for plural subjects than it was for singular subjects, as in experiment 1.

One surprising finding in experiment 2 was the three-way interaction of subject number, mismatch, and grammaticality. This interaction appears to be driven by a complete neutralization of the mismatch effect for the grammatical plural-subject conditions. This contrasts with experiment 1, where there was a mismatch penalty of a full point in the analogous conditions. Comparison with the experiment 1 results and with the experiment 2 singular-subject conditions suggests that the interaction is due in large part to the low ratings for the grammatical plural-subject match conditions, together with the relatively high ratings for the grammatical plural-subject mismatch conditions (which were also observed in experiment 1). The latter may be a real effect of subject number, but we have no clear account of the former effect.

Despite this uncertainty, experiment 2 provides further evidence against subject-confusion accounts of the mismatch effect. Fronted prepositions should provide early cues to the grammatical role of the WH-object, and so should be particularly effective in blocking any misanalysis of that phrase as a subject. That this manipulation failed to result in a consistent modulation of the mismatch effect suggests that misidentification of the initial wh-object does not play a large part in the mismatch effect. These results also provide an interesting insight into syntactic-interference accounts of this effect, which attribute the agreement error to a problem in identifying the agreement controller in a correctly parsed structure. Variants of this account that attribute the effect to cue-based retrieval mechanisms (Badecker & Kuminiak 2007, Badecker & Lewis 2007) predict that mismatch effects should be reliably attenuated when the object bears fewer subject-like features; this prediction was not borne out in this experiment. Instead, these results seem to lend more support to syntactic-interference accounts which hold that configurational factors, such as c-command and linear precedence, are what primarily control the rate of agreement attraction from a given phrase (e.g. Franck et al. 2006, Franck et al. 2010).

5. Experiment 3. Experiment 2's results suggest that overt cues to objecthood do not substantially reduce the magnitude of the mismatch effect. In experiment 3, we test a related question: whether overt cues to subjecthood can reduce the magnitude of the mismatch effect. To test this, we manipulate whether the subject NP is pronominal or lexical. Because pronouns in English overtly mark grammatical case, the overt case cues contributed by a pronominal subject should lead to a reduction in uncertainty about which phrase is the subject. If speakers identify an agreement controller using a cue-based process sensitive to the overt features of NPs in working memory (Badecker & Kuminiak 2007), or if the mismatch effect is driven by uncertainty about the correct syntactic analysis of the string, then we expect to see smaller mismatch effects for case-unambiguous pronominal subjects than for case-ambiguous lexical subjects.

5.1. Method.

Participants. In experiment 3, forty participants were recruited. Completion of the survey took approximately eight minutes on average.

Materials. We developed sixteen sets of experimental items that consisted of the eight experimental conditions shown in Table 6. Experiment 3 had a fully crossed 2 × 2 × 2 design within subjects, with the factors subject type (lexical or pronominal), subject/object match (match or mismatch), and grammaticality (grammatical or ungrammatical). Note that in experiment 3, all wh-objects were plural. Thus, unlike experiments 1 and 2, the grammaticality factor covaried with the number of the subject in this experiment. The gender of the singular pronominal subject (masculine or feminine) was manipulated so that it varied between counterbalancing lists. One item was removed prior to analysis due to a coding error.

Procedure. The procedure was identical to experiments 1 and 2.

5.2. Results. By-condition average ratings for all experimental conditions are presented in Table 7. Repeated-measures ANOVAs revealed significant main effects of grammaticality ($F_1(1,39) = 159.1$, $p < 0.05$; $F_2(1,15) = 545.6$, $p < 0.05$), subject/object match ($F_1(1,39) = 20.2$, $p < 0.05$; $F_2(1,15) = 17.5$, $p < 0.05$), and subject type ($F_1(1,39) = 25.6$, $p < 0.05$; $F_2(1,15) = 24.6$, $p < 0.05$). As in previous experiments, these main effects were qualified by a significant two-way interaction of subject/object match and grammaticality ($F_1(1,39) = 50.1$, $p < 0.05$; $F_2(1,15) = 125.6$, $p < 0.05$). In addition, there was a significant interaction of subject type and grammaticality ($F_1(1,39) = 56.9$, $p < 0.05$; $F_2(1,15) = 24.4$, $p < 0.05$), reflecting a larger grammaticality effect for pronominal subjects than for lexical subjects. No other main effects or interactions reached significance (all Fs < 3).

MATCH	SUBJECT	GRAM	
Match	Lexical	Gram	Which basketball players are the coaches planning to use this season?
Mismatch	Lexical	Gram	Which basketball players is the coach planning to use this season?
Match	Pronominal	Gram	Which basketball players are they planning to use this season?
Mismatch	Pronominal	Gram	Which basketball players is he planning to use this season?
Mismatch	Lexical	Ungram	*Which basketball players are the coach planning to use this season?
Match	Lexical	Ungram	*Which basketball players is the coaches planning to use this season?
Mismatch	Pronominal	Ungram	*Which basketball players are he planning to use this season?
Match	Pronominal	Ungram	*Which basketball players is they planning to use this season?

TABLE 6. Illustration of material in experiment 3.

	GRAMMATICAL		UNGRAMMATICAL	
	lexical	pronominal	lexical	pronominal
MATCH	6.6 (±0.13)	6.7 (±0.13)	3.1 (±0.17)	2.2 (±0.13)
MISMATCH	4.7 (±0.22)	4.9 (±0.21)	4.4 (±0.22)	2.9 (±0.22)

TABLE 7. Means of ratings, experiment 3. Parentheses represent standard error corrected for between-participant variance, following Bakeman and McArthur (1996).

5.3. DISCUSSION. In experiment 3, we again replicated the mismatch effect, and the interaction of the subject/object match and grammaticality. However, we failed to observe any interaction of subject type and the mismatch effect. Inspection of the means indicates that overt case marking on the subject did not substantially reduce the magnitude of the mismatch effect. This finding runs counter to the predictions of subject-confusion models of the mismatch effect, as these models predict that reduced uncertainty about which phrase is the subject should lead to a diminished mismatch effect. We note that we do observe a nonsignificant numerical trend toward the predicted interaction of subject type with object-auxiliary mismatch for ungrammatical sentences. Thus, as with experiment 2, we do not endorse the strong conclusion that overt case cues play absolutely no role in the mismatch effect. The results of experiment 3 do suggest, however, that overt case cues do not have a substantial influence on the size of the effect.

Interestingly, we did observe a significant interaction of subject type and grammaticality. Inspection of the means suggests that this pattern reflects a larger grammaticality effect for pronominal subjects than for lexical subjects. Since participants were more sensitive to subject-auxiliary mismatch for pronominal subjects, it seems that our pronominal manipulation did in fact have the desired effect of helping participants more reliably identify the subject of the sentence. But crucially, this effect did not interact with the mismatch effect.

6. EXPERIMENT 4. The results of experiments 1–3 appear to rule out subject-confusion accounts of the mismatch effect. Perhaps the clearest evidence in this regard is the finding that neither overt cues to objecthood (experiment 2) nor overt cues to subjecthood (experiment 3) substantially reduced the magnitude of the mismatch effect. In experiment 4 we turn to a different question and ask whether string adjacency plays a role in the mismatch effect. In experiments 1–3, the WH-object always immediately preceded the inflected auxiliary, leading to the possibility that the mismatch effect reflects sheer linear adjacency rather than the relative syntactic prominence of the WH-object. To discriminate between these possibilities, we test the mismatch effect generated by WH-objects that have prepositional postmodifiers, as in 6.

(6) Which players on the bench is the coach going to put in the game?

The addition of a PP modifier linearly separates the head of the WH-phrase *players* from the auxiliary. Instead, a local noun *bench* now stands in a linear-adjacency relationship with the auxiliary. The results of Franck et al. 2015 offer reason to believe that, in general, the head of the WH-phrase will generate more interference than the linearly preceding local noun. In a speeded acceptability-judgment study on attraction with moved WH-phrases, Franck and colleagues observed more reliable mismatch effects for the head noun than for the local noun. However, in their materials the local noun was not immediately linearly adjacent to the critical verb. If the mismatch effect is partially driven by pure linear adjacency between a fronted NP and the auxiliary, then we expect the local noun position (e.g. *bench*) to generate stronger mismatch effects than the head noun position. If, by contrast, the mismatch effect is driven by the relative syntactic prominence of the WH-object, then we expect the head of the WH-phrase to generate a larger mismatch effect.

6.1. Method.

Participants. In experiment 4, eighty participants were recruited. Because the design of experiment 4 only yields a single observation per condition per subject, the number of participants was doubled in order to increase statistical power. Completion of the survey took approximately nine minutes on average.

Materials. We developed sixteen sets of experimental items that consisted of the sixteen experimental conditions shown in Table 8. Experiment 4 had a fully crossed $2 \times 2 \times 2 \times 2$ design within subjects, with the factors subject/object match (MATCH or MISMATCH), subject/local noun match (MATCH or MISMATCH), subject number (SINGULAR or PLURAL), and grammaticality (GRAMMATICAL or UNGRAMMATICAL). A sample item is shown in Table 8.[3]

Procedure. The procedure was identical to previous experiments.

6.2. Results. By-condition average ratings for all experimental conditions are presented in Table 9. Repeated-measures ANOVAs revealed a significant main effect of grammaticality ($F_1(1,79) = 146.1, p < 0.05$; $F_2(1,15) = 179.3, p < 0.05$). As before, there was a significant two-way interaction of grammaticality and subject/head noun match ($F_1(1,79) = 29.1, p < 0.05$; $F_2(1,15) = 36.8, p < 0.05$), and a two-way interaction of subject number and subject/head noun match that was significant by participants only ($F_1(1,79) = 4.0, p < 0.05$; $F_2(1,15) = 2.1, p = 0.17$). In addition, we observed a significant two-way interaction of grammaticality and subject number ($F_1(1,79) = 8.1, p < 0.05$; $F_2(1,15) = 6.8, p < 0.05$). Finally, we observed a three-way interaction of grammaticality, subject number, and subject/head noun mismatch ($F_1(1,79) = 9.7, p < 0.05$; $F_2(1,15) = 8.4, p < 0.05$) and a three-way interaction of grammaticality, subject number, and subject/local noun mismatch ($F_1(1,79) = 9.9, p < 0.05$; $F_2(1,15) = 5.1, p < 0.05$).

To resolve our three-way interactions, we used paired *t*-tests to evaluate all subject/head noun mismatch effects, and all subject/local noun mismatch effects. For grammat-

[3] One item in experiment 4 contained the same ambiguity as the item removed in experiments 1 and 2 (#10; see Appendix B). However, we include it in the present experiment because it would be impossible to perform a by-participant analysis after removing it, given that subjects only saw one item in each condition. A by-item analysis that excluded this item revealed no substantial differences from those reported here.

HEAD MATCH	LOCAL N MATCH	SUBJECT	GRAM	
Match	Match	Singular	Gram	Which basketball player on the sideline is the coach planning to use this season?
Mismatch	Match	Singular	Gram	Which basketball players on the sideline is the coach planning to use this season?
Match	Mismatch	Singular	Gram	Which basketball player on the sidelines is the coach planning to use this season?
Mismatch	Mismatch	Singular	Gram	Which basketball players on the sidelines is the coach planning to use this season?
Mismatch	Mismatch	Singular	Ungram	*Which basketball players on the sidelines are the coach planning to use this season?
Match	Mismatch	Singular	Ungram	*Which basketball player on the sidelines are the coach planning to use this season?
Mismatch	Match	Singular	Ungram	*Which basketball players on the sideline are the coach planning to use this season?
Match	Match	Singular	Ungram	*Which basketball player on the sideline are the coach planning to use this season?
Match	Match	Plural	Gram	Which basketball players on the sidelines are the coaches planning to use this season?
Mismatch	Match	Plural	Gram	Which basketball player on the sidelines are the coaches planning to use this season?
Match	Mismatch	Plural	Gram	Which basketball players on the sideline are the coaches planning to use this season?
Mismatch	Mismatch	Plural	Gram	Which basketball player on the sideline are the coaches planning to use this season?
Mismatch	Mismatch	Plural	Ungram	*Which basketball player on the sideline is the coaches planning to use this season?
Match	Mismatch	Plural	Ungram	*Which basketball players on the sideline is the coaches planning to use this season?
Mismatch	Match	Plural	Ungram	*Which basketball player on the sidelines is the coaches planning to use this season?
Match	Match	Plural	Ungram	*Which basketball players on the sidelines is the coaches planning to use this season?

TABLE 8. Illustration of material in experiment 4.

SINGULAR SUBJECTS	GRAMMATICAL		UNGRAMMATICAL	
	local match	local mismatch	local match	local mismatch
HEAD MATCH	6.0 (±0.16)	5.4 (±0.17)	3.1 (±0.15)	3.6 (±0.18)
HEAD MISMATCH	5.1 (±0.17)	4.8 (±0.19)	4.2 (±0.21)	4.4 (±0.18)
PLURAL SUBJECTS	GRAMMATICAL		UNGRAMMATICAL	
	local match	local mismatch	local match	local mismatch
HEAD MATCH	5.8 (±0.15)	6.1 (±0.13)	3.6 (±0.16)	3.5 (±0.16)
HEAD MISMATCH	5.3 (±0.17)	5.5 (±0.19)	3.6 (±0.17)	3.6 (±0.15)

TABLE 9. Means of ratings, experiment 4. Parentheses represent standard error corrected for between-participant variance, following Bakeman and McArthur (1996).

ical sentences, subject/head noun mismatch led to significant decreases in acceptability for singular subjects ($t(79) = 3.8$, $p < 0.05$; 95% CI = [0.35, 1.15]) and for plural subjects ($t(79) = 3.3$, $p < 0.05$; 95% CI = [0.22, 0.88]). For ungrammatical sentences, subject/head noun mismatch led to significant increases in acceptability for singular subjects only ($t(79) = -5.1$, $p < 0.05$; 95% CI = [−1.30, −0.57]); this effect was not significant for plural subjects ($t(79) = -0.6$, $p = 0.55$; 95% CI = [−0.37, 0.20]).

Subject/local noun mismatch had a similar effect on ratings but was less statistically reliable. For grammatical sentences, subject/local noun mismatch decreased acceptability for singular subjects ($t(79) = 2.9$, $p < 0.05$; 95% CI = [0.15, 0.75]), but not for plural subjects ($t(79) = -1.3$, $p = 0.18$; 95% CI = [−0.55, 0.11]), where the direction of the numerical trend is actually toward increased acceptability with a mismatched local noun. Similarly, subject/local noun mismatch marginally increased acceptability of ungrammatical sentences with singular subjects ($t(79) = -1.8$, $p < 0.1$; 95% CI = [−0.69, 0.03]), but not with plural subjects ($t(79) = 0.18$, $p = 0.86$; 95% CI = [−0.25, 0.30]).

6.3. Discussion. Experiment 4 sought to test whether linear adjacency or structural prominence drove the mismatch effect in object wh-object questions. Overall, the results of experiment 4 suggest that subject/head noun mismatch has a greater effect on acceptability than does subject/local noun mismatch: in all comparisons, the magnitude of the subject/head noun mismatch effect was greater than the subject/local noun mismatch effect, and the latter was less statistically reliable overall. This pattern suggests that the mismatch effect is driven to a greater degree by the syntactic prominence of the object wh-phrase than it is by its linear adjacency to the auxiliary. For example, in grammatical sentences with singular subjects, the subject/local noun mismatch effect was 0.45 points, while the subject/head noun mismatch effect was 0.75 points. Similarly, in grammatical sentences with plural subjects, subject/head noun mismatch resulted in a 0.55 point effect on judgments, whereas subject/local noun mismatch did not reliably reduce acceptability (rather, there was a trend toward an effect in the opposite direction). These results suggest string adjacency between a verb and a preceding noun is neither necessary nor sufficient to generate a mismatch effect.

However, comparison of experiment 4 with the preceding experiments suggests that interposing material between the head of the fronted wh-object and the auxiliary may reduce the size of the mismatch effect. Across the board, effects of mismatch between the auxiliary and the head were smaller in experiment 4 than in the analogous conditions of experiment 1, which is the most closely comparable experiment. Thus, while it is clear that linear adjacency is not necessary for a mismatch effect to appear, there is an indication that nonadjacency may reduce the size of the effect. We test this more directly in experiment 5.

7. Experiment 5. Experiment 4 showed that syntactic prominence of the wh-object led to greater mismatch effects than did linear adjacency. However, a comparison of the overall size of the mismatch effect in experiment 4 with that in experiment 1 suggests a role for linear adjacency; in experiment 1 (when the wh-object's head noun was linearly adjacent to the auxiliary), mismatch effects were approximately twice as large as in experiment 4 (when it was not). Thus, while syntactic prominence appears to be a primary factor is generating mismatch effects, linear adjacency may still play a role. In experiment 5, we directly test this possibility by manipulating the position of a parenthetical phrase, as in 7.

(7) a. According to Tom, which flowers is the gardener planting?
    b. Which flowers, according to Tom, is the gardener planting?

The parenthetical-phrase manipulation holds constant the structural position of the wh-object, as well as its complexity. If linear adjacency (or recency) plays a role in generating mismatch effects, we expect to see larger mismatch effects when the parenthetical does not intervene between the object and the auxiliary (7a) than when it does (7b).

7.1. Method.

Participants. In experiment 5, forty participants were recruited. Completion of the survey took approximately nine minutes on average.

Materials. We developed twenty sets of experimental items that consisted of the four experimental conditions shown in Table 10. Experiment 5 had a fully crossed 2 × 2 design within subjects, with the factors subject/object match (MATCH or MISMATCH) and parenthetical position (INTERVENING or NONINTERVENING). Subject number was counterbalanced across items, such that half of the items had a singular subject, and half had a plural subject. All items in this experiment were grammatical. Every parenthetical contained a noun phrase, and the number of this noun phrase was counterbalanced across items.

MATCH	PARENTHETICAL	
Match	Nonintervening	According to the schedule, which monkey is the zookeeper training this week?
Match	Intervening	Which monkey, according to the schedule, is the zookeeper training this week?
Mismatch	Nonintervening	According to the schedule, which monkeys is the zookeeper training this week?
Mismatch	Intervening	Which monkeys, according to the schedule, is the zookeeper training this week?

TABLE 10. Illustration of material in experiment 5.

PROCEDURE. The procedure was identical to the previous judgment experiments.

7.2. RESULTS. By-condition average ratings for all experimental conditions are presented in Table 11. Repeated-measures ANOVAs revealed a significant main effect of object-auxiliary number match ($F_1(1,39) = 44.0$, $p < 0.05$; $F_2(1,19) = 22.5$, $p < 0.05$) and a significant two-way interaction of object-auxiliary mismatch and parenthetical position ($F_1(1,39) = 13.6$, $p < 0.05$; $F_2(1,19) = 9.6$, $p < 0.05$).

	NONINTERVENING	INTERVENING
MATCH	6.2 (±0.10)	5.8 (±0.07)
MISMATCH	5.1 (±0.10)	5.4 (±0.09)

TABLE 11. Means of ratings, experiment 5. Parentheses represent standard error corrected for between-participant variance, following Bakeman and McArthur (1996).

To resolve the significant interaction of object-auxiliary match and parenthetical position, we used paired t-tests to test for an effect of number match within each level of the parenthetical position number factor. This analysis revealed significant effects of object-auxiliary mismatch both for nonintervening parentheticals (95% CI of mismatch effect = [0.70, 1.4], $t(39) = 6.4$, $p < 0.05$) and for intervening parentheticals (95% CI of mismatch effect = [0.20, 0.64], $t(39) = 3.9$, $p < 0.05$).

7.3. DISCUSSION. Experiment 5 confirmed that linear adjacency of the fronted WH-phrase and the auxiliary has a significant influence on the magnitude of the mismatch effect. Positioning a parenthetical between the auxiliary and the WH-object effectively halved the size of the mismatch effect, although it did not eliminate it. On the basis of these results alone, we cannot say if this effect is strictly due to linear adjacency of the WH-object and the auxiliary, or if it instead reflects the relative recency of the WH-object. Nonetheless, the finding that linear order has an effect on the mismatch effect is theoretically interesting.

From the point of view of models that attribute the mismatch effects to an error during working-memory retrieval, it is perhaps surprising that mismatch effects are largest when the WH-object is adjacent to the verb. This is because this is not a position where memory retrieval is likely to be necessary, as the object and the auxiliary are likely to be under active processing at the same time (see McElree 2006 for discussion). For this reason it is not clear that the mismatch effect in our examples should be attributed to working-memory retrieval per se, but the present results do support the view that general memory variables such as recency codetermine how likely the WH-object is to interfere in agreement processing.

Furthermore, the finding that linear proximity to the verb allows the WH-object to interfere more strongly with the agreement relation sets our mismatch effects apart from agreement attraction in PP postmodifier configurations: to the extent that the linear position of the local noun in these configurations creates interference, it seems that linear distance to the head noun, not the agreeing verb, is critical. We take up this observation in more detail in the general discussion.

8. GENERAL DISCUSSION. Across five off-line acceptability-judgment experiments, we examined the effect that a fronted WH-object exerts on the agreement relations inside a sentence. In experiment 1, we saw mismatch effects for both grammatical and ungrammatical sentences. When the fronted WH-object was plural, the mismatch effect was roughly twice as large as when it was singular, though both plural and singular objects created reliable mismatch effects. We provided a computational model of the judgments that suggests that this result reflects discrete 'grammatical' or 'ungrammatical' responses to mismatching conditions, rather than a gradient effect of grammaticality. Experiment 2 investigated the role of explicit disambiguation of the initial WH-object using fronted prepositions. It was found that preposition fronting did not reliably reduce the size of the mismatch effect. Experiment 3 investigated the role that case cues played in the controller-selection process by comparing pronominal versus lexical subjects. We failed to observe any effect of subject type on the size of the mismatch effect, suggesting that overt case marking of the subject does not substantially diminish the likelihood of choosing the object as the agreement controller. Taken together, experiments 2 and 3 show that using case or thematic cues to disambiguate the subject and object does not significantly modulate the probability of forming an agreement relation between the verb and the fronted object. This argument is based on noneffects of preposition fronting and case marking on the subject, and so we cannot conclusively rule out the possibility that case and thematic information modulate the mismatch effect to some extent. However, it appears that the effect of these variables is relatively minimal compared to other factors we manipulated.

In experiments 4 and 5, we examined whether strict linear adjacency between the object and the auxiliary was a necessary component of the mismatch effect. We found that it was not, although the size of the mismatch effect was smaller in experiment 4, in which a PP modifier intervened between the head of the WH-phrase and the auxiliary. Experiment 5 found a similar pattern: the magnitude of the mismatch effect was reduced when a parenthetical linearly intervened between the object and the auxiliary.

8.1. TOWARD AN ACCOUNT OF THE OBJECT MISMATCH EFFECT. There are several features of our data that any account of the object mismatch effect must accommodate. They are as follows:

(i) ALL OBJECTS INTERFERE: Singular and plural WH-objects both cause mismatch effects.
(ii) NUMBER ASYMMETRY: The mismatch effect is smaller with a plural subject.
(iii) LINEAR ADJACENCY: The mismatch effect is strongest when the object head noun is linearly adjacent to the auxiliary.
(iv) NONEFFECT OF CASE OR THEMATIC CUES: The mismatch effect is not substantially reduced by preposition fronting of the object, or overt case marking on the subject.
(v) DISCRETE RESPONSES: The mismatch effect reflects discrete choices.

Taken together, we believe that features (i)–(v) provide strong constraints on possible accounts of the mismatch effect in the WH-object constructions we studied. Perhaps most clearly, the finding that mismatch effects remain robust even when grammatical role is explicitly disambiguated (iv) rules out accounts of this effect that turn on subject confusion or syntactic misanalysis of some sort (Gorrell 1998; see also Bock & Miller 1991). We also note that the effect of linear adjacency of the object and the auxiliary is not straightforwardly compatible with the temporary-misanalysis account. Given the observation that lengthening ambiguous regions is known to make reanalysis more difficult (Ferreira & Henderson 1991, Tabor & Hutchins 2004), we would expect the addition of length to increase the severity of the mismatch effect, rather than decrease it, on a subject-confusion account. The results of experiment 5 disconfirm this prediction.

We believe it is also difficult to straightforwardly account for our results using existing feature-transmission accounts of agreement errors, which are often invoked to explain attraction from PP postmodifiers. For example, the finding of discrete response patterns (v) is incompatible with graded representations of number on the subject, ruling out a mechanism such as the SAP number-marking value in Eberhard et al. 2005. In contrast, this finding is compatible with a discrete feature-transmission mechanism, such as feature percolation. However, there are two features of our data that may be taken as arguments against this interpretation. First, consider our findings that all objects interfere (i) and the number asymmetry effect (ii). Feature-transmission accounts of agreement attraction often invoke markedness to account for the widely observed number asymmetry in agreement attraction (e.g. Eberhard 1997). If one adopts the view that singular NPs do not (normally) carry a number feature, then these accounts make the strong prediction that there should be no interference at all from singular objects, because the singular object does not have a [SINGULAR] value to transmit to the subject. Thus, the simple observation of attraction from singular fronted objects prevents these models from providing a satisfactory account of the current data. Although we suggest that these theories do not straightforwardly extend to our configurations, we hasten to add that the finding of interference from singular distractors does not empirically distinguish the mismatch effect in WH-questions from the mismatch effect in PP attraction. Franck and colleagues (2002) observed reliable mismatch effects from singular distractors inside PP postmodifiers in French, as well as in English. Their findings suggest that attraction from unmarked singular distractors is reliably observed in configurations that have an overall high rate of agreement errors. Other studies have also reported some effect of a singular distractor in a PP postmodifier in English (e.g. Staub 2009), though this effect, as in the present data, is smaller than the corresponding plural effect.

A second difficulty in our data for feature-transmission models of agreement attraction is the effect of linear order (iii). Linear distance from the local noun to the agreeing verb does not appear to be a strong factor in agreement attraction from PP postmodifiers (Bock & Cutting 1992, Franck et al. 2002, Vigliocco & Nicol 1998). Gillespie and Pearlmutter (2011) provide evidence that it may be linear distance from the distractor to the head noun, rather than the linear distance from the distractor to the verb, that contributes to agreement attraction with PP postmodifiers. However, Haskell and MacDonald (2005) provide evidence for a strong effect of linear adjacency on agreement with conjoined subjects, as in 8.

(8) a. Can you tell me whether the horses or the clock is/are red?
b. Can you tell me whether the clock or the horses are/is red?

Haskell and MacDonald observed almost 100% plural production in 8b, compared with approximately 25% plural production in 8a. When they reversed the linear order of verb and the controlling disjunction by prompting participants to ask yes/no questions (e.g. 'Are the horses or the clock red?'), this pattern was reversed. They concluded that speakers consider linear adjacency between an NP and an inflected verb in addition to syntactic prominence when selecting an agreement controller. Finally, Franck and colleagues (2006) observed significantly more attraction from moved objects in cleft structures in French when the subject was postverbal, creating an OVS structure, than when the subject preceded the verb, in a surface SVO structure. Overall, in terms of the effect that linear order has on agreement errors, errors in English WH-object questions pattern more like agreement in disjunction and agreement from moved objects in cleft constructions, rather than attraction from PP postmodifiers, which instead appears sensitive to the distance between the attractor and the head noun of the subject. This differential sensitivity to linear order supports the view that distinct mechanisms may underlie the mismatch effect we observe in our studies and agreement attraction from PP postmodifiers.

If this line of argumentation is correct, we are left with syntactic-interference models as the most promising account of erroneous agreement in WH-object questions. On this view, the syntactic structure of the sentence has been correctly identified, but agreement errors arise during the process of identifying an agreement controller over that structure. The finding that moved objects in English constituent questions create a particularly potent form of agreement attraction is predicted on the model advocated by Franck and colleagues (2002, 2006, 2010). This is because both the subject and the object in this configuration are in a specifier-head relationship with the auxiliary, albeit at different levels of representation. This follows from the standard assumption that the WH-phrase in our stimuli has been moved to occupy a position in the specifier of the highest clausal projection (e.g. CP), and that the process of auxiliary inversion involves moving the inflectional head of the clause (e.g. T) to occupy the head of the highest clausal projection (CP). On the Franck et al. model, this tight syntactic relationship between the moved object and the inflected auxiliary should allow the object to create a high degree of interference in the search for an agreement controller (see also Franck 2011).

Our results suggest, however, that more than the spec-head relationship is at play in determining the likelihood of object agreement. There is a substantial linear component: the size of the mismatch effect decreases the further the head of the object NP is, in string linear terms, from the agreeing auxiliary. This fits neatly with accounts of syntactic interference that explain attraction effects in terms of working-memory retrieval errors (Badecker & Kuminiak 2007, Wagers et al. 2009). These models can explain the linear-adjacency effects by appealing to standard memory variables such as recency and decay, which could make more distant distractors less available in working memory, and thereby less available to create interference in the search for an agreement controller. In addition, working memory-oriented explanations are compatible with a discrete response distribution. However, the mismatch effect we observed in our experiments creates both theoretical and empirical difficulties for existing models of syntactic interference based on working-memory retrieval. From a theoretical point of view, it is unclear what role retrieval from working memory would play in generating the mismatch effect when the distractor noun is linearly adjacent to the critical verb, since this configuration may obviate the need for any retrieval of the object phrase from working memory (McElree 2006). Empirically, the results of experiments 2 and 3 fail to confirm the predictions of a working-memory retrieval account, because neither overt cues to the fronted object's grammatical position (experiment 2) nor overt case cues on the subject (experiment 3) significantly reduced the mismatch effect. Overall, our results suggest that both structural factors (viz. specifier-head configurations) and linear adjacency conspire to determine how strongly a phrase will compete to be selected as an agreement controller, with relatively less influence from overt case or thematic cues. This simple model accounts for features (i) and (iii)–(v) of our data.

The account we have described does not, on its own, provide any explanation of the number asymmetry effect. One way in which the model may be extended to capture this effect is to stipulate that phrases that bear a marked value for number (on our assumptions, plural NPs) outcompete phrases that do not bear marked values for number. Badecker and Kuminiak (2007:68) suggest that this scenario follows from an optimality-theoretic approach to agreement computation. They suppose that in English there is a highly ranked faithfulness constraint that prefers outputs that express marked number values in the input (i.e. FaithAgr(PLURAL)), which is necessary to overcome the markedness constraint that weighs against having marked features in the output (e.g. *PLURAL). In any scenario with multiple potential agreement controllers, then, a faithfulness constraint of this sort will provide a bias toward selecting phrases with marked features to control agreement; choosing the unmarked controller will result in a violation of the FaithAgr(PLURAL) constraint, and so will be dispreferred (see Badecker & Lewis 2007 for a computational implementation of this idea).

This model predicts that highly marked controllers should outcompete less marked controllers, minimizing the amount of interference they contribute. Intuition suggests that this prediction is correct. Consider the effect of introducing into a wh-object question a first-person subject, which is assumed to bear a marked person value relative to third person NPs (Harley & Ritter 2002, Preminger 2014).

(9) *Which flowers are I planting in the garden today?
*Which flower are I planting in the garden today?

To our judgment, the mismatch effect does not arise in this configuration: 9a and 9b seem equally ill-formed (see also Baker 2008). If this intuition is correct, it suggests that configurational and markedness factors jointly determine the fitness of a phrase as an agreement controller.

Finally, we note an informal observation about native speakers' reactions to sentences such as *Which trees is the gardener planting?*. For at least some speakers, the subjective impression that this sentence is degraded survives the explicit recognition that the verb does agree with the subject. We think that this is a qualitatively different state of affairs from what obtains in the case of syntactic misanalysis, or in the case of traditional agreement attraction. Naive subjects generally find *The horse raced past the barn fell* to be completely unacceptable, until, often quite suddenly, it is not. Once the subject is able to construct the correct parse, there is no remaining sense of difficulty or unacceptability. By contrast, a sentence such as *The key to the cabinets is on the table* does not generate much of a subjective impression of unacceptability in the first place. But a number mismatch between an initial wh-object and the following auxiliary seems (again, for at least some speakers) to have an effect akin to the effect exerted by a visual illusion such as the Müller-Lyer illusion. In this case, the subjective impression that two lines are of different lengths is not alleviated, let alone eliminated, by measuring the lines and determining that they are in fact the same length. We think that this (admittedly informal) observation provides an argument that the current phenomenon stands apart from other forms of attraction, either in strength or in kind. It further suggests that the sentence-processing mechanism does sometimes establish an agreement relation between the auxiliary and the preceding wh-object.

8.2. Conclusion. We conducted a series of experiments aimed at exploring the unusual agreement pattern observed in sentences like *Which flowers are the gardener planting?*. Our investigations suggest that this phenomenon resists explanation in terms of existing models of agreement attraction. Instead, we argued for a model whereby syntactic phrases compete to control agreement on the verb as a function of their markedness and of their linear and structural relationship to the verb (see also Franck 2011 and references therein). This model predicts that moved WH-objects in constituent questions in English should create a high degree of syntactic interference, leading to a high degree of uncertainty about which phrase actually controls agreement in this configuration.

APPENDIX A: DETAILS OF THE JUDGMENT MODEL

This appendix describes how we generated the distributional predictions of the discrete and gradient response models of experiment 1.

The DISCRETE MODEL of the responses assumes that every response to partial match conditions is an independent draw from EITHER an ungrammatical reference distribution (*Ungram*) OR a grammatical reference distribution (*Gram*). The probability of sampling from either distribution is controlled by a mixing parameter π, which here represents the probability of drawing from a sample from the grammatical distribution. We thus draw a sample judgment *j* from the mixture model in the following fashion.

(A1) $$\begin{cases} j \sim Gram & \pi \\ j \sim Ungram & (1-\pi) \end{cases}$$

A sample from the GRADIENT MODEL is generated by independently drawing a sample *g* and a sample *u* from the grammatical and ungrammatical reference distributions, respectively. A sample judgment *j* from the admixture model is then computed by combining these samples in a weighted fashion.

(A2) $$g \sim Gram \qquad u \sim Ungram$$

(A3) $$j = \pi g + (1-\pi)u$$

In this model, π is the relative weight given to samples drawn from the grammatical distribution. In order to make the resulting samples *j* conform to the response scale, noninteger values of *j* were rounded to the nearest integer.

Prior to fitting the models, we estimated the model-free parameter π by minimizing the distance between a given model's predicted mean response value and the observed empirical mean for a given test distribution. To avoid local minima, we estimated π five times independently, and chose from these five runs the value of π that had the smallest value on our objective function.

Next we modeled the response distribution for a given test condition by drawing *n* random samples from each of the discrete and gradient response models, where *n* is the number of experimental observations in the test condition to be modeled (in all cases $n = 83$). We smoothed the resulting counts using Laplace add-one smoothing, which is equivalent to placing a uniform Dirichlet prior over the outcomes. The resulting smoothed counts were then used to estimate the parameters of a multinomial distribution over each of the seven possible response categories using maximum likelihood estimation.

To evaluate each model's fit to the observed data in each of the four test conditions, we used the Bayesian information criterion (BIC; Kass & Raftery 1995, Schwarz 1978). The BIC is calculated as follows.

(A4) $$\mathrm{BIC} = -2\ln L + k \ln(n)$$

Where *L* is the data likelihood under the model, *k* is the number of model parameters, and *n* is the number of data points.[4] The BIC is a goodness-of-fit metric based on the data likelihood that can be used to compare nonnested models. Here, the lower the BIC value, the higher the probability that the model assigns to the data. Because raw BIC scores do not yield an easily interpretable measure of model performance, we computed Bayes factors on the basis of the BIC scores. Here, the Bayes factors were the ratio of the likelihood of the data given the discrete model to the likelihood of the test data given the gradient model. These ratios reflect the odds in favor of the discrete model over the gradient model, and can be approximated from the difference in BIC scores between two models by:

(A5) $$\mathrm{BF} = \exp(-\tfrac{1}{2} * \Delta\mathrm{BIC})$$ (Kass & Raftery 1995)

[4] Note that for both admixture and mixture models, $k = 6$ and $n = 83$, making the model-complexity term $k \ln(n)$ constant across both models to be compared.

Appendix B: Experimental materials

Experiment 1

1. Which tree(s) is/are the gardener(s) planting?
2. Which basketball player(s) is/are the coach(es) planning to use this season?
3. Which contestant(s) is/are the reality show judge(s) going to send home tonight?
4. Which painting(s) is/are the art critic(s) talking about in the article?
5. Which spice(s) is/are the Mexican chef(s) going to add to the dish?
6. Which computer(s) is/are the technician(s) trying to repair before Monday?
7. Which barn(s) is/are the farmer(s) painting bright red?
8. Which wrench(es) is/are the plumber(s) supposed to bring today?
9. Which car(s) is/are the mechanic(s) rebuilding now?
10. Which window(s) is/are the janitor(s) supposed to wash?
11. Which TV program(s) is/are the kid(s) going to watch tonight?
12. Which vegetable(s) is/are the chef(s) using on the show?
13. Which bicycle(s) is/are the mechanic(s) working on today?
14. Which monkey(s) is/are the zookeeper(s) training this week?
15. Which teacher(s) is/are the administrator(s) firing at the end of the year?

Experiment 2

1. Which tree(s) is/are the hiker resting under? / Under which tree(s) is/are the hiker resting?
2. Which sandbox(es) is/are the kid planning to play in? / In which sandbox(es) is/are the kid planning to play?
3. Which passenger(s) is/are the flight attendant yelling at? / At which passenger(s) is/are the flight attendant yelling?
4. Which tourist(s) is/are the guide traveling with? / With which tourist(s) is/are the guide traveling?
5. Which arena(s) is/are the boxer fighting in? / In which arena(s) is/are the boxer fighting?
6. Which hostage(s) is/are the diplomat speaking to? / To which hostage(s) is/are the diplomat speaking?
7. Which state(s) is/are the plane flying over? / Over which state(s) is/are the plane flying?
8. Which bar(s) is/are the monkey hanging from? / From which bar(s) is/are the monkey hanging?
9. Which contestant(s) are/is the reality show judges speaking to? / To which contestant(s) are/is the reality show judges speaking?
10. Which customer(s) are/is the Mexican chefs cooking for? / For which customer(s) are/is the Mexican chefs cooking?
11. Which lake(s) are/is the girls swimming in? / In which lake(s) are/is the girls swimming?
12. Which river(s) are/is the boys fishing in? / In which river(s) are/is the boys fishing?
13. Which machine(s) are/is the mechanics working with? / With which machine(s) are/is the mechanics working?
14. Which suspect(s) are/is the witnesses referring to? / To which suspect(s) are/is the witnesses referring?
15. Which car(s) are/is the burglars sleeping in? / In which car(s) are/is the burglars sleeping?
16. Which bed(s) are/is the kids jumping on? / On which bed(s) are/is the kids jumping?

Experiment 3

1. Which basketball players are/is the coach(es) planning to use this season? / Which basketball players are/is they/he planning to use this season?
2. Which contestants are/is the reality show judge(s) going to send home tonight? / Which contestants are/is they/she going to send home tonight?
3. Which paintings are/is the art critic(s) talking about in the article? / Which paintings are/is they/he talking about in the article?
4. Which spices are/is the Mexican chef(s) going to add to the dish? / Which spices are/is they/she going to add to the dish?
5. Which computers are/is the technician(s) trying to repair before Monday? / Which computers are/is they/he trying to repair before Monday?
6. Which trees are/is the gardener(s) planting? / Which trees are/is they/she planting?
7. Which barns are/is the farmer(s) painting bright red? / Which barns are/is they/he painting bright red?
8. Which wrenches are/is the plumber(s) supposed to bring today? / Which wrenches are/is they/she supposed to bring today?
9. Which cars are/is the mechanic(s) rebuilding now? / Which cars are/is they/he rebuilding now?
10. Which roads are/is the bicyclist(s) using? / Which roads are/is they/she using?
11. Which TV programs are/is the kid(s) going to watch tonight? / Which TV programs are/is they/she going to watch tonight?

12. Which vegetables are/is the chef(s) using on the show? / Which vegetables are/is they/he using on the show?
13. Which bicycles are/is the mechanic(s) working on today? / Which bicycles are/is they/she working on today?
14. Which monkeys are/is the zookeeper(s) training this week? / Which monkeys are/is they/he training this week?
15. Which teachers are/is the administrator(s) firing at the end of the year? / Which teachers are/is they/she firing at the end of the year?

Experiment 4

1. Which basketball player(s) on the sideline(s) is/are the coach(es) planning to use this season?
2. Which athlete(s) in the tournament(s) is/are the referee(s) going to disqualify?
3. Which painting(s) near the sculpture(s) is/are the art critic(s) talking about in the article?
4. Which spice(s) on the shelf(s) is/are the Mexican chef(s) going to add to the dish?
5. Which computer(s) on the counter(s) is/are the technician(s) trying to repair before Monday?
6. Which tree(s) from the landscaper(s) is/are the gardener(s) planting?
7. Which barn(s) near the field(s) is/are the farmer(s) painting bright red?
8. Which tool(s) from the cabinet(s) is/are the plumber(s) supposed to bring today?
9. Which car(s) near the pedestrian(s) is/are the police officer(s) going to pull over?
10. Which road(s) through the park(s) is/are the pedestrian(s) crossing?
11. Which stamp(s) in the display case(s) is/are the postal worker(s) selling to the customer?
12. Which reptile(s) from the mountain(s) is/are the scientist(s) studying?
13. Which vegetable(s) from the garden(s) is the chef(s) using in the kitchen?
14. Which recipe(s) from the cookbook(s) is/are the chef(s) preparing?
15. Which monkey(s) in the tree(s) is/are the zookeeper(s) trying to find?
16. Which program(s) in the school(s) is/are the administrator(s) cutting at the end of the year?

Experiment 5

1. According to Mom, which tree(s) is the gardener planting today? / Which tree(s), according to Mom, is the gardener planting today?
2. By the way, which TV program(s) is the kid going to watch tonight? / Which TV program(s), by the way, is the kid going to watch tonight?
3. According to the landlord, which pipe(s) is the plumber supposed to fix next? / Which pipe(s), according to the landlord, is the plumber supposed to fix next?
4. By the way, which road(s) is the biker afraid to travel on? / Which road(s), by the way, is the biker afraid to travel on?
5. According to the schedule, which monkey(s) is the zookeeper training this week? / Which monkey(s), according to the schedule, is the zookeeper training this week?
6. According to the teachers, which window(s) is the janitor supposed to wash? / Which window(s), according to the teachers, is the janitor supposed to wash?
7. As far as you know, which car(s) is the mechanic rebuilding now? / Which car(s), as far as you know, is the mechanic rebuilding now?
8. According to the ranchers, which barn(s) is the farmer painting bright red? / Which barn(s), according to the ranchers, is the farmer painting bright red?
9. As far as you know, which vegetable(s) is the chef using in the dish? / Which vegetable(s), as far as you know, is the chef using in the dish?
10. According to the students, which bicycle(s) is the mechanic working on today? / Which bicycle(s), according to the students, is the mechanic working on today?
11. According to the news, which teacher(s) are the administrators firing this year? / Which teacher(s), according to the news, are the administrators firing this year?
12. By the way, which computer(s) are the technicians trying to repair before Monday? / Which computer(s), by the way, are the technicians trying to repair before Monday?
13. In your opinion, which contestant(s) are the reality show judges sending home tonight? / Which contestant(s), in your opinion, are the reality show judges sending home tonight?
14. By the way, which house(s) are the contractors being hired to work on? / Which house(s), by the way, are the contractors being hired to work on?
15. For this dish anyway, which spice(s) are the Mexican chefs using in the sauce? / Which spice(s), for this dish anyway, are the Mexican chefs using in the sauce?
16. As far as you know, which basketball player(s) are the coaches recruiting this season? / Which basketball player(s), as far as you know, are the coaches recruiting this season?

17. According to CNN and MSNBC, which politician(s) are the voters going to kick out of office? / Which politicians, according to CNN and MSNBC, are the voters going to kick out of office?
18. According to the professors, which problem(s) are the students going to get wrong? / Which problem(s), according to the professors, are the students going to get wrong?
19. As far as you know, which painting(s) are the art critics talking about in the article? / Which painting(s), as far as you know, are the art critics talking about in the article?
20. According to the blogs, which book(s) are the cool kids going to read this summer? / Which book(s), according to the blogs, are the cool kids going to read this summer?

REFERENCES

Badecker, William, and Frantisek Kuminiak. 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. DOI: 10.1016/j.jml.2006.08.004.

Badecker, William, and Richard Lewis. 2007. A new theory and computational model of working memory in sentence production: Agreement errors as failures of cue-based retrieval. Paper presented at the 20th annual CUNY Conference on Human Sentence Processing, University of California, San Diego.

Bakeman, Roger, and Duncan McArthur. 1996. Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods, Instruments, & Computers* 28(4).584–89. DOI: 10.3758/BF03200546.

Baker, Mark C. 2008. *The syntax of agreement and concord*. Cambridge: Cambridge University Press.

Bock, Kathryn, and J. Cooper Cutting. 1992. Regulating mental energy: Performance units in language production. *Journal of Memory and Language* 31(1).99–127. DOI: 10.1016/0749-596X(92)90007-K.

Bock, Kathryn, and Kathleen M. Eberhard. 1993. Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes* 8(1).57–99. DOI: 10.1080/01690969308406949.

Bock, Kathryn, and Erica L. Middleton. 2011. Reaching agreement. *Natural Language and Linguistic Theory* 29(4).1033–69. DOI: 10.1007/s11049-011-9148-y.

Bock, Kathryn, and Carol A. Miller. 1991. Broken agreement. *Cognitive Psychology* 23(1).45–93. DOI: 10.1016/0010-0285(91)90003-7.

Clifton, Charles, Jr.; Lyn Frazier; and Patricia Deevy. 1999. Feature manipulation in sentence comprehension. *Rivista di Linguistica* 11(1).11–40. Online: http://linguistica.sns.it/RdL/11.1/Clifton-Frazier-Deevy.pdf.

Dillon, Brian; Alan Mishler; Shayne Sloggett; and Colin Phillips. 2013. Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language* 69(2).85–103. DOI: 10.1016/j.jml.2013.04.003.

Eberhard, Kathleen M. 1997. The marked effect of number on subject–verb agreement. *Journal of Memory and Language* 36(2).147–64. DOI: 10.1006/jmla.1996.2484.

Eberhard, Kathleen M.; J. Cooper Cutting; and Kathryn Bock. 2005. Making syntax of sense: Number agreement in sentence production. *Psychological Review* 112(3).531–59. DOI: 10.1037/0033-295X.112.3.531.

Ferreira, Fernanda, and John M. Henderson. 1991. Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language* 30(6).725–45. DOI: 10.1016/0749-596X(91)90034-H.

Franck, Julie. 2011. Reaching agreement as a core syntactic process: Commentary of Bock & Middleton *Reaching agreement*. *Natural Language and Linguistic Theory* 29(4).1071–86. DOI: 10.1007/s11049-011-9153-1.

Franck, Julie; Saveria Colonna; and Luigi Rizzi. 2015. Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in Psychology* 6:349. DOI: 10.3389/fpsyg.2015.00349.

Franck, Julie; Glenda Lassi; Ulrich H. Frauenfelder; and Luigi Rizzi. 2006. Agreement and movement: A syntactic analysis of attraction. *Cognition* 101(1).173–216. DOI: 10.1016/j.cognition.2005.10.003.

Franck, Julie; Gabriela Soare; Ulrich H. Frauenfelder; and Luigi Rizzi. 2010. Object interference in subject–verb agreement: The role of intermediate traces of movement. *Journal of Memory and Language* 62(2).166–82. DOI: 10.1016/j.jml.2009.11.001.

Franck, Julie; Gabriella Vigliocco; and Janet Nicol. 2002. Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes* 17(4).371–404. DOI: 10.1080/01690960143000254.

Gillespie, Maureen, and Neal J. Pearlmutter. 2011. Hierarchy and scope of planning in subject–verb agreement production. *Cognition* 118(3).377–97. DOI: 10.1016/j.cognition.2010.10.008.

Gillespie, Maureen, and Neal J. Pearlmutter. 2013. Against structural constraints in subject–verb agreement production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(2).515–28. DOI: 10.1037/a0029005.

Gorrell, Paul. 1998. Syntactic analysis and reanalysis in sentence processing. *Reanalysis in sentence processing*, ed. by Janet Dean Fodor and Fernanda Ferreira, 201–45. Dordrecht: Springer. DOI: 10.1007/978-94-015-9070-9_7.

Harley, Heidi, and Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78(3).482–526. DOI: 10.1353/lan.2002.0158.

Hartsuiker, Robert J.; Inés Antón-Méndez; and Marije van Zee. 2001. Object attraction in subject-verb agreement construction. *Journal of Memory and Language* 45(4).546–72. DOI: 10.1006/jmla.2000.2787.

Haskell, Todd R., and Maryellen C. MacDonald. 2005. Constituent structure and linear order in language production: Evidence from subject-verb agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(5).891–904. DOI: 10.1037/0278-7393.31.5.891.

Häussler, Jana, and Markus Bader. 2009. Agreement checking and number attraction in sentence comprehension: Insights from German relative clauses. *Travaux du cercle linguistique de Prague* 7.

Jeffreys, Harold. 1961. *Theory of probability*. Oxford: Oxford University Press.

Kass, Robert E., and Adrian E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* 90(430).773–95. DOI: 10.2307/2291091.

Kimball, John, and Judith Aissen. 1971. I think, you think, he think. *Linguistic Inquiry* 2(2).241–46. Online: http://www.jstor.org/stable/4177629.

McElree, Brian. 2006. Accessing recent events. *Psychology of Learning and Motivation* 46.155–200. DOI: 10.1016/S0079-7421(06)46005-9.

Nicol, Janet L.; Kenneth I. Forster; and Csaba Veres. 1997. Subject–verb agreement processes in comprehension. *Journal of Memory and Language* 36(4).569–87. DOI: 10.1006/jmla.1996.2497.

Pearlmutter, Neal J.; Susan M. Garnsey; and Kathryn Bock. 1999. Agreement processes in sentence comprehension. *Journal of Memory and Language* 41(3).427–56. DOI: 10.1006/jmla.1999.2653.

Preminger, Omer. 2014. *Agreement and its failures*. (Linguistic Inquiry monographs 68.) Cambridge, MA: MIT Press.

Schwarz, Gideon. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6(2).461–64. Online: http://www.jstor.org/stable/2958889.

Solomon, Eric S., and Neal J. Pearlmutter. 2004. Semantic integration and syntactic planning in language production. *Cognitive Psychology* 49(1).1–46. DOI: 10.1016/j.cogpsych.2003.10.001.

Sprouse, Jon; Carson T. Schütze; and Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from *Linguistic Inquiry* 2001–2010. *Lingua* 134.219–48. DOI: 10.1016/j.lingua.2013.07.002.

Staub, Adrian. 2009. On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language* 60(2).308–27. DOI: 10.1016/j.jml.2008.11.002.

Staub, Adrian. 2010. Response time distributional evidence for distinct varieties of number attraction. *Cognition* 114(3).447–54. DOI: 10.1016/j.cognition.2009.11.003.

Tabor, Whitney, and Sean Hutchins. 2004. Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(2).431–50. DOI: 10.1037/0278-7393.30.2.431.

Tanner, Darren; Janet Nicol; and Laurel Brehm. 2014. The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language* 76.195–215. DOI: 10.1016/j.jml.2014.07.003.

Vigliocco, Gabriella; Brian Butterworth; and Carlo Semenza. 1995. Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language* 34(2).186–215. DOI: 10.1006/jmla.1995.1009.

Vigliocco, Gabriella, and Janet Nicol. 1998. Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition* 68(1).B13–B29. DOI: 10.1016/S0010-0277(98)00041-9.

Wagers, Matthew W.; Ellen F. Lau; and Colin Phillips. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61(2).206–37. DOI: 10.1016/j.jml.2009.04.002.

Dillon
Department of Linguistics
University of Massachusetts
Amherst, MA 01003
[brian@linguist.umass.edu]

[Received 16 December 2014;
revision invited 12 January 2016;
revision received 8 February 2016;
accepted 4 March 2016]

Staub, Levy, Clifton
Department of Psychological and Brain Sciences
University of Massachusetts
Amherst, MA 01003
[astaub@psych.umass.edu]
[cec@psych.umass.edu]
[jwlevy@psych.umass.edu]