Journal of Semantics 25: 1–44
doi:10.1093/jos/ffm018

Coherence and Coreference Revisited

ANDREW KEHLER, LAURA KERTZ, HANNAH ROHDE AND JEFFREY L. ELMAN
University of California, San Diego

Abstract

For more than three decades, research into the psycholinguistics of pronoun interpretation has argued that hearers use various interpretation 'preferences' or 'strategies' that are associated with specific linguistic properties of antecedent expressions. This focus is a departure from the type of approach outlined in Hobbs (1979), who argues that the mechanisms supporting pronoun interpretation are driven predominantly by semantics, world knowledge and inference, with particular attention to how these are used to establish the *coherence* of a discourse. On the basis of three new experimental studies, we evaluate a coherence-driven analysis with respect to four previously proposed interpretation biases—based on grammatical role parallelism, thematic roles, implicit causality, and subjecthood—and argue that the coherence-driven analysis can explain the underlying source of the biases and predict in what contexts evidence for each will surface. The results further suggest that pronoun interpretation is incrementally influenced by probabilistic expectations that hearers have regarding what coherence relations are likely to ensue, together with their expectations about what entities will be mentioned next, which, crucially, are conditioned on those coherence relations.

1 INTRODUCTION

More than three decades of research has sought to uncover the principles that determine how hearers interpret pronouns in context.[1] This work, which has predominantly been carried out in the psycholinguistics and computational linguistics communities, has focused to a large extent on identifying preferences or heuristics that hearers utilize to interpret a pronoun; these preferences are often based on linguistic properties of possible antecedent expressions, such as the grammatical and thematic roles that they fill within a sentence. As a collection, these preferences are often in conflict, and no clear consensus has emerged with respect to how they are utilized or how conflicts among them are reconciled during the interpretation process.

[1] Throughout the discussion, we will use *pronouns* to mean unaccented, third-person pronouns, unless otherwise specified. Accented third-person pronouns are considered in section 3.5.

The emphasis on such factors may partially explain Beaver's (2004) observation of a 'curious near absence of work within [the formal semantics and pragmatics] tradition on anaphora resolution', particularly with respect to its concentration on absolute semantic constraints rather than semantically relevant factors that cause some interpretations to be favoured over others. Indeed, pronouns provide a textbook case of an underspecified linguistic form that must be semantically interpreted within a context, and as such, we would argue, the study of their behaviour offers a window into the larger questions concerning the semantic and discourse interpretation processes that go on around them. Yet, with limited exceptions, the semanticist will find a striking lack of emphasis on meaning in the existing literature on the topic.

The current focus on preference-driven theories is in fact a departure from the type of approach outlined in Hobbs (1979), who, working in the artificial intelligence tradition, argued that the mechanisms that drive pronoun interpretation are driven predominantly by semantics, world knowledge and inference, with particular reference to how these are used to establish the *coherence* of discourses. That is, in his account the same types of inference processes that semanticists commonly appeal to for computing implicatures, accommodating presuppositions, and the like are also those used for computing, using his term, the 'petty implicatures' associated with assigning pronouns to their referents. Hobbs's approach thus gives us a starting point for an attempt to bridge the gap between semantics and psycholinguistic research as they pertain to pronoun interpretation. In previous work, Kehler (2002) argued that the preferences commonly cited in the psycholinguistic and computational linguistics literatures are to some extent epiphenomena of the methods by which discourse coherence is established, although he offered no new empirical data to support this position.

In this paper, we present new evidence in support of a coherence analysis (sketched in section 2), and describe how it can accommodate a range of previous findings suggestive of conflicting preferences and biases.[2] We start section 3 by examining the grammatical subject preference (Crawley *et al*. 1990, *inter alia*), which favours referents that occupy the grammatical subject position of the previous clause, and the grammatical role parallelism preference (Sheldon 1974; Smyth 1994; Chambers and Smyth 1998, *inter alia*),[3] which favours referents that

[2] To be clear, we will not ultimately conclude that coherence establishment is the root cause of *all* biases in pronoun interpretation. See section 7 for further discussion.

[3] Smyth does not characterize parallelism effects as the result of an independent preference, but instead as a by-product of the structure of the coreference processor. We discuss his analysis in more detail in sections 3.1 and 3.4.

occupy the same grammatical role as the pronoun. We present the results of our first experiment that show that both preferences can be neutralized when coherence is carefully controlled for, and furthermore argue that the grammatical role parallelism preference is an epiphenomenon of an independent interaction between information structure and accent placement in a particular class of coherence relations. We follow in section 4 with the results of a second experiment designed to distinguish two types of bias proposed by Stevenson *et al*. (1994): a thematic role preference, according to which the occupants of the Goal thematic role are preferred to those that occupy the Source, and an event-structure bias, according to which hearers focus on the end state of the previous eventuality when interpreting an utterance. The results support the event-structure bias, and further show that the bias is limited primarily to those coherence relations which implicate event structure in their formulation. In section 5, we address the ramifications of our analysis for the time course of pronoun interpretation during incremental processing, and offer a model that captures how a hearer's coherence-driven expectations about how the discourse is likely to proceed can predict online measurements of pronoun interpretation difficulty. In section 6, we examine a case study with respect to *implicit causality* biases that have been well-studied in the psycholinguistics literature, and argue on the basis of a third experiment that they represent one instance of a more comprehensive set of biases that drive predictive discourse interpretation. In section 7, we revisit the grammatical subject preference and offer reasons against interpreting the results of Crawley *et al*. (1990) as support for an independent subject assignment strategy. We also argue, however, that data from Stevenson *et al*. (1994) offer more convincing support for the existence of a subject bias beyond what can be explained solely by coherence-driven expectations, and suggest a way in which these data can still be explained without appeal to overlaid interpretation heuristics or preferences. We conclude in section 8 by summarizing the ways in which our analysis provides alternative explanations of previous results and suggests areas for future work.

2 COHERENCE AND COREFERENCE

Hobbs (1979) presents what in some respects could be considered to be the most parsimonious theory offered to date of how pronouns are interpreted. In his account, pronoun interpretation is not even an independent process, but instead results as a by-product of more general reasoning about the most likely interpretation of an utterance,

including the establishment of discourse coherence. Pronouns are modelled as free variables in logical representations which become bound during these inference processes; potential referents of pronouns are therefore those which result in valid proofs of coherence.

To illustrate, consider passages (1a) and (1b), adapted from an example from Winograd (1972).

(1) The city council denied the demonstrators a permit because...
  a. ... they *feared* violence.
  b. ... they *advocated* violence.

Hearers appear to have little difficulty resolving the pronoun *they* in each case, despite the fact that it refers to *the city council* in sentence (1a) and *the demonstrators* in sentence (1b). Note that the only difference is the verb used in the second clause, which suggests that semantics and world knowledge are responsible for determining the correct referents. The Explanation coherence relation, as signalled by *because*, is operative in each case (the variables S_1 and S_2 represent the first and second sentences being related, respectively):

> **Explanation:** Infer P from the assertion of S_1 and Q from the assertion of S_2, where normally $Q \rightarrow P$.

Oversimplifying a bit, we encode the world knowledge necessary to establish Explanation for (1) within a single axiom, given in (2).

(2) $fear(X, V) \wedge advocate(Y, V) \wedge enable\_to\_cause(Z, Y, V) \rightarrow deny(X, Y, Z)$

If we assume that the variables X, Y, V and Z are bound to the city council, the demonstrators, violence, and the permit, respectively, axiom (2) says that if the city council fears violence, the demonstrators advocate violence, and a permit would enable the demonstrators to bring about violence, then it might 'plausibly follow' that the city council would deny the demonstrators a permit.

The first sentence in (1) is represented as in (3).

(3) $deny(city\_council, demonstrators, permit)$

This representation matches the consequent of axiom (2), triggering a process of abductive inference that can be used to establish Explanation. At this point, X will become bound to *city_council*, Y to *demonstrators* and Z to *permit*.

Both the follow-ons (1a,b) provide information that can be used to establish one of the conjuncts in the antecedent of the axiom, thereby establishing a causal connection between the clauses. Clause (1a) is

represented as in (4), in which the unbound variable *T* represents the pronoun *they*.

(4) *fear*(*T*, *violence*)

This predication unifies with the first conjunct in the antecedent of axiom (2), forcing the unification of the variables *T* and *X*. Since *X* is already bound to *city_council*, the variable *T* representing *they* also receives this binding, and the pronoun is therefore resolved.

Likewise, clause (1b) is represented in (5).

(5) *advocate*(*T*, *violence*)

This predication also matches a conjunct in the antecedent of axiom (2), but in this case it is the second conjunct, which will necessitate the unification of the variables *T* and *Y*. Since *Y* is already bound to *demonstrators*, the representation of *they* also receives this binding. Thus, identification of the correct referent for the pronoun in both (1a) and (1b) is a by-product of establishing an Explanation relation.

Despite the appeal of this example, the literature has largely rejected Hobbs's approach in favour of methods that rely on more surface-level aspects of linguistic representation, such as the grammatical and thematic roles that potential antecedents occupy. There are no doubt reasons for this; for one, there are statistical tendencies in support of such preferences (e.g. a bias towards references to the previous subject as compared to other grammatical positions) that do not receive obvious explanations from a purely coherence-driven theory. Further, it is unclear how facts concerning incremental processing can be predicted by a coherence-driven account that relies on information that may not become available until well after the pronoun is encountered.[4]

Kehler (2002) extended Hobbs's work by presenting a typology of coherence relations, most taken or adapted from those in Hobbs (1990), based on three general classes of 'connection among ideas' first articulated by Hume in his *Inquiry Concerning Human Understanding*—namely *Resemblance*, *Contiguity* in time or place and *Cause* or *Effect* [Hume 1955: 32 (1748)]. Kehler argues that these categories differ in the types of inference processes used to establish them; this distinction in turn affects how pronouns are interpreted. This, Kehler claims, explains why different heuristic preferences appear to dominate in different contextual circumstances. We will describe exemplar relations in each of the three categories (particularly *Occasion*,

[4] Both these concerns will be addressed later in the paper.

Parallel and *Result*) as well as the manner in which establishing each interacts with pronoun interpretation by considering examples (6a–d):

(6) a. Bush narrowly defeated Kerry, and special interests promptly began lobbying him. [=Bush]
b. Kerry was narrowly defeated by Bush, and special interests promptly began lobbying him. [=Kerry][5]
c. Bush narrowly defeated Kerry, and Romney absolutely trounced him. [=Kerry]
d. Bush narrowly defeated Kerry, and he quickly demanded a recount. [=Kerry]

The alternation found in examples (6a,b) can be used to argue for the existence of a grammatical subject preference. The difference in voice in the first clause results in different entities being realized in subject position, and most informants find that the favoured interpretation for the pronoun shifts accordingly. Kehler argues that the subject preference is most closely associated with examples that participate in the Contiguity relation Occasion (such as (6a,b)), which is defined as follows (adapted from definitions in Hobbs 1990):

> **Occasion:** Infer a change of state for a system of entities from the assertion of S_2, establishing the initial state for this system from the final state of the assertion of S_1.

Occasion allows one to express a situation centered around a system of entities by using intermediate states of affairs as points of connection between partial descriptions of that situation. As such, the inference process that underlies Occasion attempts to equate the initial state of the second utterance with the final state of the first, performing inferences as necessary. Biases in pronoun interpretation in Occasion are therefore predicted to correspond to the relative degrees of salience of the event participants with respect to (the hearer's mental representation of) the event's end state. As the grammatical subject is the canonical place to mention the topic of a sentence—in the sense that, information structurally, (6a) highlights what Bush did, whereas (6b) highlights what happened to Kerry—it stands to reason that the degree of salience accorded to Bush and Kerry would differ between (6a,b), and with it, the preferred referent for the pronoun.[6]

Thematic role biases discussed in the literature can also be linked to Occasion. In a passage-completion experiment, Stevenson *et al.* (1994)

[5] The preference for Kerry in this case may rely to some degree on the hearer knowing that he is a US Senator, and thus, like Bush, is able to be lobbied.

[6] Section 7 will present a refinement of this characterization.

found evidence for both a grammatical subject preference and a bias in favour of entities that occupy the Goal thematic role over those that occupy the Source. Whereas participants were considerably more likely to complete passages like (7a) in a way that requires *he* to refer to John rather than Bill (here John is both the subject and the Goal), they are equally likely to complete passages like (7b) in a way that requires that *he* refer to Bill (a non-subject Goal) as John (a subject Source).

(7) a. John seized the comic from Bill. He_____
b. John passed the comic to Bill. He_____

Stevenson *et al*. thus conclude that there is both a subject assignment strategy and a Goal preference at work—which agree on a referent in (7a), but disagree in (7b)—and that the Goal preference may result from a bias towards focusing on end states. We will return to this topic in section 4, where we argue that the end-state bias is in part a by-product of the manner in which Occasion relations are established.

Example (6c) provides counterevidence to the subject preference we witnessed in (6a,b) since the preferred referent is the object of the first clause rather than the subject. Such examples have been used to argue for a grammatical role parallelism preference, which favours entities that occupy the same grammatical role as the pronoun (Sheldon 1974; Smyth 1994; Chambers and Smyth 1998, cf. footnote 3). Kehler argues that this preference is closely associated with Resemblance coherence relations such as Parallel, in which commonalities and contrasts among corresponding sets of parallel relations and entities are established:

Parallel: Infer $P(a_1, a_2, \ldots)$ from the assertion of S_1 and $P(b_1, b_2, \ldots)$ from the assertion of S_2, for a common P and similar a_i and b_i.

In (6c), the entities Bush and Romney are parallel, as are Kerry and the referent assigned to *him*. We will henceforth refer to such pairs a_i and b_i for some i as *parallel elements*. Examples cited to support a grammatical role parallelism preference are often characterized by Parallel relations, as are the typical stimuli found in psycholinguistic research in support of the preference (Smyth 1994; Chambers and Smyth 1998). The bias towards a pronoun's parallel element in these constructions is very strong; informants are almost unanimous in judging the pronoun in (6c) to refer to Kerry (assuming that the pronoun is not contrastively accented; more on this in section 3.5). The question, then, is why the same effect is not seen in example (6a). Note that this preference is not straightforwardly predicted on a coherence-driven theory, since assigning either referent in (6c) would result in a perfectly coherent Parallel relation. Kehler (2002) offers a rationale for this association, but that position will be revised in section 3.5.

Finally, example (6d), repeated below as (8), is an instance of the Result relation, which, like the previously discussed Explanation relation, is in the Cause–Effect category.

(8) Bush narrowly defeated Kerry, and he quickly demanded a recount. [=Kerry]

Establishing a Cause–Effect relation requires that a causal link be identified between the propositions denoted by the utterances in a passage. The Result relation is essentially the same as Explanation except that the cause precedes the effect:

Result: Infer P from the assertion of S_1 and Q from the assertion of S_2, where normally $P \rightarrow Q$.

The analysis of example (8) would follow the spirit of the analysis of examples (1a,b). As this example violates both the subject and grammatical parallelism preferences, it argues instead for a 'common sense' preference, since the interpretation of the pronoun appears to be determined by the same world knowledge that is used to establish the coherence of the passage, specifically that one would expect the loser of an election to demand a recount rather than the winner.

To sum, we have described three categories of coherence relation that are associated with three underlying inference processes, which in turn appear to be correlated with different types of pronoun interpretation biases. In the sections that follow, we describe psycholinguistic experiments intended to evaluate the evidence for these biases in the context of a coherence-driven analysis.

3 GRAMMATICAL ROLE PREFERENCES

The first of our studies addresses the conflict between the subject and grammatical parallelism preferences in light of the coherence analysis. Much of the motivation for the experiment and its design draw from the work of Smyth (1994), and additional aspects are motivated by the work of Wolf *et al*. (2004). We briefly describe these two works in turn, and then follow with a discussion of our experiment.

3.1 *Smyth (1994)*

Smyth (1994) posits an *Extended Feature Match Hypothesis* (EFMH), which characterizes pronoun assignment as a search process based on feature matching that predicts that a 'pronoun with two or more grammatically and pragmatically possible antecedents in a preceding clause will be interpreted as coreferential with the candidate that has the same grammatical role' (p. 197). Whereas we have thus far cast the

parallelism preference as a heuristic, it is worth noting that Smyth explicitly denies this view, stating that 'PF [=parallel function] is not a special default strategy, but rather an epiphenomenon arising from the structure of the coreference processor', and thus 'there is no sense in which it is an independent rule or strategy to be acquired'. Instead, coreference is established by a feature-match process, and due to a priming effect, the identity of the grammatical role filled by the referent is available as one of the criteria for matching, along with other features (e.g. number, gender). A lack of full syntactic parallelism between the clauses—such as when one clause contains an adjunct and the other does not—is predicted to prevent syntactic priming and reactivation, resulting in fewer parallel interpretations (pp. 206–7).

We will focus on the two of Smyth's four experiments that are central to our analysis, his Experiments 2 and 3.[7] Both are argued to provide evidence for the EFMH, and hence to contradict the claim of Crawley *et al*. (1990) that parallel function 'is not important for understanding pronouns in text'. The Experiment 2 materials were constructed by taking 20 of the stimuli of Crawley *et al*. and modifying them so that the clauses were fully parallel syntactically. The non-subject roles were varied between direct, indirect and prepositional objects. A sample passage is given in (9).

(9) Mary helped Julie change the tire and then she helped Peter change the oil.

Participants were asked to fill a blank by writing the name of the person that they understood the pronoun to refer to. The results overwhelmingly favoured parallel assignment; 100% of the subject pronouns were assigned to the preceding subject and 88.12% of the non-subject pronouns were assigned to the non-subject referent.

Experiment 3 tested the prediction that a reduction in the parallelism between the clauses should reduce the number of parallel responses. It varied three factors: grammatical role parallelism for the non-subjects (parallel v. not parallel), full syntactic parallelism (no adjunct v. adjunct), and pronoun position (subject or non-subject). The results further supported parallel assignment, as the percentage of parallel assignments ranged from 64% to 90% across conditions. There were also main effects of adjunct parallelism and grammatical role parallelism: Cases in the non-parallel adjunct condition received fewer parallel assignments than those in

[7] The first experiment was a small study to test the role of context sentences in the experiments of Crawley *et al*. on their results. The fourth tested a variety of effects in cases involving subordinate structures, which will not concern us here.

the parallel condition, and similarly cases in the non-parallel role condition received fewer parallel assignments than those in the parallel condition.

As pointed out by Kehler (2002), however, an examination of Smyth's syntactically parallel stimuli suggests that his modifications to the examples of Crawley *et al*. may have introduced a confound, in that in some cases they also changed the operative coherence relation from Occasion to Parallel, whereas Occasion appears to be more highly represented in his non-parallel stimuli. Hence, our first experiment controls for and manipulates syntactic parallelism and coherence separately. So as to keep the results as directly comparable as possible, our design will otherwise follow Smyth's fairly closely, particularly with respect to being an offline task in which readers are explicitly asked for their pronoun assignments.

3.2 *Wolf et al. (2004)*

Wolf *et al*. (2004) previously tested the predictions of Kehler (2002) against both the grammatical subject and grammatical role parallelism preferences in a reading time experiment that manipulated coherence frame (Parallel/Result) and pronoun gender (masculine/feminine). Coherence was signalled by manipulating the verb in the context sentence as well as a connective between the first two clauses, specifically *and similarly* (which signals a Parallel relation) and *and so* (which signals a Result relation). In half of the stimuli, the referent indicated by pronoun gender supported the coherence frame, and in the remaining half it did not. Examples are given in (10a,b).

(10) a. Fiona complimented Craig and similarly James congratulated her/him after the match, but nobody took any notice.
b. Fiona defeated Craig and so James congratulated her/him after the match, but nobody took any notice.

For the Parallel stimuli, faster reading times were measured when the antecedent was in a parallel grammatical role than when it was not. For the Result stimuli, which were semantically biased towards a non-parallel referent, faster reading times were measured for non-parallel antecedents. Wolf *et al*. thus confirmed that preferences for pronoun interpretation can be reversed by manipulating coherence, per Kehler (2002).

Several questions remain that warrant investigation, however. First, Wolf *et al*. used gender-unambiguous pronouns, which, in the causal continuations, resulted in interpretations that were less coherent in the parallel antecedent condition than in the non-parallel condition (consider the variant of (10b) with the pronoun *him* v. *her*). As a result, the increased reading times could have been caused by this incoherence

rather than by the pronoun interpretation process. Second, the stimuli of Wolf *et al*. contained only object pronouns, and thus each possible interpretation was supported by either the grammatical subject preference or the grammatical role parallelism preference. We ask whether similar results would be found for subject pronouns with object antecedents, which are dispreferred by both preferences. Third, since their stimuli all include a prepositional phrase in the second clause but not in the first, the passages did not have fully parallel structure. Whereas this property is irrelevant to a general grammatical role parallelism preference, it makes their results potentially compatible with the EFMH's prediction that a lack of full syntactic parallelism will result in a reduced parallelism bias. Finally, it has been proposed that connectives carry their own focusing properties (Stevenson *et al*. 1994, 2000) that can affect antecedent selection, such that the use of *and similarly* and *and so* in the data of Wolf *et al*. could be claimed to redirect the current focus of attention in different ways. While we find this idea to be uncompelling in several respects (see footnote 25), we can test whether similar effects will be found for stimuli without connectives by not relying on them to disambiguate coherence.

3.3 *Experiment 1*

The present experiment addresses a variety of factors by independently varying sentence structure, pronoun position and coherence relation in an ambiguous pronoun resolution task (Kertz *et al*. 2006). Two versions of each preference are evaluated—a 'basic' version, which characterizes it as a single, all-purpose processing strategy, and a particular 'modified' version, which corresponds more closely to one of the aforementioned proposals in the literature. The basic version of the grammatical subject preference states an across-the-board preference for antecedents that occupy the subject position of the previous clause. The modified version allows for the possibility that the subject preference will not override an interpretation favoured by a strong pragmatic bias (Crawley *et al*. 1990). The basic version of the grammatical role parallelism preference states an across-the-board preference for antecedents that occupy the same grammatical role as the pronoun. The modified version adds an additional constraint requiring that the syntactic structures of the two clauses be fully parallel; otherwise the grammatical subject preference is invoked (Smyth 1994).[8]

[8] As Crawley *et al*. did for the grammatical subject preference, Smyth likewise appeals to the idea that plausibility factors could limit the applicability of the parallel grammatical role preference. We address this proviso in section 3.4.

3.3.1 *Stimuli* In a 2 × 2 × 2 design, stimulus sets were constructed with eight variants, as in (11a–d). Each stimulus contains two clauses: an introduction and a follow-on that contains an ambiguous pronoun. Both clauses contain a transitive verb in active voice.

(11) Samuel threatened Justin with a knife, and
  a. . . . Erin blindfolded him (with a scarf). [Parallel]
  b. . . . Erin stopped him (with pepper spray). [Result]
  c. . . . he blindfolded Erin (with a scarf). [Parallel]
  d. . . . he alerted security (with a shout). [Result]

Sixteen stimulus sets were constructed for a total of 128 experimental stimuli. Each set varied pronoun position (subject/object), sentence structure (fully/partially parallel) and coherence relation (Parallel/Result). With possible antecedents as the subject and object of the first clause, we are able to test the full 2 × 2 configurations of possible coreference patterns.

As in Wolf *et al*. (2004), passages participating in Result relationships semantically favoured the non-parallel referent, whereas those participating in Parallel relations incorporated no semantic bias. The modified grammatical subject preference can thus be evaluated by analysing the Parallel condition only. The distinction between full and partial syntactic parallelism between the clauses was implemented by either including or excluding a modifier phrase in the second clause to match the modifier in the first clause, which were varied between pre-verbal adverbial phrases and post-verbal prepositional phrases, balanced across sets. Varying the stimuli across this dimension allows us to determine if pronoun interpretation is affected by the existence of full v. partial syntactic parallelism, as predicted by the modified grammatical role parallelism preference, when coherence is controlled for separately.

Coherence type was assessed in a norming phase, during which three trained judges, blind to our hypothesis, were asked to categorize stimuli as instances of either Parallel coherence or Result coherence. All three judges agreed on the coherence relation for 119 out of 128 total stimuli. For the remaining nine stimuli, two of three judges agreed with an averaged confidence level above a pre-determined threshold.

3.3.2 *Participants* Thirty-two undergraduates from the University of California San Diego (UCSD) participated for extra course credit. All were self-reported monolingual speakers of English.

3.3.3 *Task* A repeated measure design was used, in which each participant was tested on two stimuli from each of the eight types, with no two variants from the same set presented to the same participant. The two replications were block randomized, and the 16 experimental stimuli were interleaved with 24 distractors (16 of which also contained ambiguous pronouns and 8 of which contained unambiguous pronouns). The resulting sixteen lists were then reversed to rule out ordering effects, yielding 32 unique stimulus lists.

Participants were presented with a paper and pencil task, for which they read a two-clause passage and answered a question immediately after, as in (12).

(12) Samuel threatened Justin with a knife, and he blindfolded Erin with a scarf.
Who blindfolded Erin?

The participant's answer was taken to indicate the antecedent selected in interpreting the ambiguous pronoun.

3.3.4 *Predictions* As we have characterized it, the 'basic' form of the grammatical subject preference predicts a strong bias towards interpreting all pronouns to refer to the subject of the previous clause. The modified form predicts the same bias, but only in Parallel relations since the Result stimuli were pragmatically biased. The 'basic' form of the grammatical role parallelism preference predicts a strong bias towards interpreting subject and object pronouns to refer to subject and object antecedents, respectively. The modified grammatical role parallelism preference makes the same predictions, but only for the stimuli in the fully parallel condition. The coherence hypothesis makes the same predictions as the basic grammatical role parallelism preference for Parallel coherence stimuli (regardless of the full/partial syntactic parallelism distinction), but predicts an interpretation bias towards grammatically non-parallel referents for Result stimuli.

3.3.5 *Results* The results followed the predictions of the coherence hypothesis, confirming the expected interaction between pronoun position and coherence type, but were not consistent with the other hypotheses. The raw number of subject v. object assignments for each of the eight conditions is shown in Table 1.

Table 2 organizes the results according to the predictions of each account. These results show that the manipulations to test the basic and modified forms of both the grammatical subject and grammatical parallelism preferences all resulted in near 50/50 splits, whereas the

Coherence	Syntax	Pronoun position	Subject ante	Object ante
Parallel	Parallel	Subject	64	0
		Object	5	59
	Non-parallel	Subject	61	3
		Object	8	56
Result	Parallel	Subject	2	62
		Object	59	5
	Non-parallel	Subject	4	60
		Object	61	3

Table 1 Results of Experiment 1 by condition

Condition	Subject ante	Object ante	*n*
The subject preference			
All pronouns	0.52	0.48	512
The qualified subject preference			
Non-biasing context (Parallel coherence)	0.54	0.46	256
The parallel structure preference			
Subject pronouns	0.51	0.49	256
Object pronouns	0.52	0.48	256
The qualified parallel structure preference			
Subject pronouns (fully parallel structure)	0.52	0.48	128
Object pronouns (fully parallel structure)	0.50	0.50	128
The coherence hypothesis			
Subject pronouns (Parallel coherence)	0.98	0.02	128
Subject pronouns (Result coherence)	0.05	0.95	128
Object pronouns (Parallel coherence)	0.10	0.90	128
Object pronouns (Result coherence)	0.94	0.06	128

Table 2 Results of Experiment 1 by analysis

predictions of the coherence analysis were all confirmed with at least a 90/10 split.

The dependent measure for our statistical analyses was the rate of assignments to the subject antecedent (subject and object assignments received scores of 1 and 0, respectively). A full factorial analysis of variance was conducted with pronoun position (subject/object), sentence structure (fully/partially parallel) and coherence relation (Parallel/Result) as factors, with separate analyses treating participants

(F_1) and items (F_2) as random variables. The analysis confirms that the interaction between coherence type and pronoun position, predicted by the coherence hypothesis, is significant [$F_1(1, 31) = 1379.23$, $P < 0.0001$; $F_2(1, 15) = 2016.158$, $P < 0.0001$]. A second smaller effect, which we did not predict, was found for coherence alone [$F_1(1, 31) = 4.429$, $P < 0.05$; $F_2(1, 15) = 7.105$, $P < 0.05$], where subject antecedents were selected more often in Parallel coherence relations than in Result relations.

Collapsing across conditions, the overall mean score was 0.516 ± 0.062. A one-sample *t*-test comparing this mean to a hypothetical mean of 0.5 demonstrates that the overall rate of subject antecedent assignment is not significantly different from chance, contra the grammatical subject preference. Whereas the main effect of coherence described above could potentially be interpreted as slight support for the modified subject preference, this effect is overwhelmed by the effect predicted by the coherence analysis. The main effect of pronoun position, predicted by the grammatical role parallelism preference, is not statistically significant; nor is the interaction between sentence structure and pronoun position predicted by the modified grammatical role parallelism preference. There was no significant effect of structure, and no significant interaction between structure and coherence. Likewise, there was no significant three-way interaction among coherence, syntactic structure and pronoun position. As such, lack of parallel structure did not impact the likelihood of a parallel pronoun assignment in the Parallel condition, for either subject or object pronouns. Finally, modifier type (pre-verbal adverbial v. post-verbal prepositional phrase) was not a significant factor alone or within any interaction.

These results support the coherence hypothesis, confirming that pronoun interpretation preferences can be triggered or suppressed by manipulating coherence relations. They also suggest that the contradictory results reported in the literature to date may stem at least in part from a failure to control for coherence in experimental stimuli.

We forgo a detailed discussion of these results with respect to the grammatical subject preference until section 7, since Experiments 2 and 3, to be presented subsequently, are relevant to that discussion as well. We discuss these results with respect to the EFMH further in the next section. We then follow up in section 3.5 with a semantic analysis that, we claim, demonstrates that the grammatical role parallelism preference is an epiphenomenon of the interaction of information structure and accent placement in Parallel relations.

3.4 *Comparison with the EFMH*

We conclude from the results of Experiment 1 that coherence is the dominant factor in determining parallel reference assignments, and not grammatical structure: (i) parallel structure did not give rise to parallel coreference in Result stimuli and (ii) reduced syntactic parallelism did not reduce the likelihood of a parallel interpretation in Parallel stimuli.

As we have discussed, these results contrast with the predictions of the EFMH, which characterizes pronoun interpretation as a feature-matching process that is sensitive in part on the degree of syntactic parallelism between clauses. With respect to result (i) above, however, it should be noted that Smyth acknowledges that pragmatic biases might be at work in some examples: 'in some cases, a conjunction can introduce a pragmatic bias which is incompatible with a PF interpretation' (p. 208). He cites example (13), in which a causal interpretation supports interpreting the non-subject pronoun *him* as coreferent with the subject Phil:

(13) Phil tickled Shanley, and (so) Liz poked him.

While it is not completely clear to us how such biases are predicted to interact with the feature-matching mechanism of the EFMH, example (13) is of the sort employed in our stimuli for the Result condition.

Following Sheldon (1974), however, Smyth also correctly notes that the parallelism effect is so strong that it can seemingly trump gender mismatches (see also Oehrle 1981):

(14) William bumped Bonnie and ?she/SHE poked Rod.

That is, example (14) is infelicitous without accent on *she* (cf. 13), even though there is only one female referent available.[9] This fact weakens the force of Smyth's appeal to pragmatic biases with respect to examples like (13), however, since one is left with a parallelism effect that is so strong that it can withstand a firm semantic constraint like a gender mismatch but yet is soft enough to be overridden by a more pliable pragmatic bias. And the fact of the matter is that pragmatic biases cannot override parallel function if the operative coherence relation is Parallel, as pointed out in Kehler (2002). Consider (15):

(15) Condi Rice admires Hillary Clinton, and George W. Bush absolutely worships her.

[9] We will offer an explanation for this infelicity in the next section.

Assuming a Parallel relation with a deaccented *her*, informants reliably report that the referent must be Clinton, despite a strong pragmatic bias towards Rice given the political persuasions of the politicians involved. It therefore needs to be explained why plausibility can save a non-parallel pronoun interpretation in (13) but not in (15). The crucial difference between the acceptable (13) and the unacceptable version of (14) is that (13) participates in a Result relation, whereas (14) participates in a Parallel relation. Plausibility only comes into play in determining which referent makes for a coherent Result relation in (13).

3.5 *A semantic analysis of the grammatical role parallelism preference*

Experiment 1 revealed a dramatic bias in Parallel (but not Result) coherence relations towards a referent in a parallel grammatical role, across both the subject pronoun and object pronoun conditions. In this section, we ask why Parallel coherence is so strongly aligned with parallel coreference, giving the appearance that a parallel grammatical role bias is at play. We argue that the bias emerges from the interaction between coherence relations and information structure, for reasons that are independent of a theory of pronoun interpretation.

3.5.1 *Parallelism effects* That there would be an association between Parallel coherence and parallel coreference may, at first blush, seem unsurprising on a coherence-driven analysis. After all, when establishing Parallel coherence, the inference mechanism attempts to establish points of similarity between a pronoun and its parallel element. It stands to reason that the way to establish maximal similarity is to assume coreference between the two. Indeed, Kehler (2002) posited an analysis of just this sort.

This observation does not fully explain the behaviour we discussed in the previous section, however. For one, as we just saw, the parallelism effect is recalcitrantly strong as compared to other types of preferences noted in the literature, able to withstand strong pragmatic biases (15) and even gender conflicts (16).

(16) Condi Rice admires Donald Rumsfeld, and George W. Bush absolutely worships her. [=Rumsfeld?]

Yet the strength of the parallelism bias cannot be attributed only to the semantics of the Parallel relation, since substituting a mention of either referent by name in place of the pronoun in either of these examples results in a perfectly coherent Parallel passage.

No other preference proposed in the literature is resilient to grammatical and world-knowledge influences in a similar way. Yet, as we have already seen, the effect simply appears to go away when the operative coherence relation is non-Resemblance. That is, the coreference pattern that was infelicitous for (15) and (16) is perfectly acceptable on a Result interpretation, per (17a,b), respectively.

(17) a. Condi Rice defeated Hillary Clinton and George Bush congratulated her.
b. Condi Rice defeated John Kerry and George Bush congratulated her.

Lest there be any doubt that these different interpretation patterns are due to the difference in coherence type, we can ask whether passages that are ambiguous between Parallel and Result construals enforce different constraints on the interpretation of unaccented pronouns across the two coherence construals. This is indeed the case; consider (18):

(18) Powell defied Cheney, and Bush punished him. (Kehler 2002)

On the Parallel construal of (18) (paraphrase *and* as *and similarly*), *him* can only refer to Cheney if unaccented (i.e. it can refer to Powell only if it receives accent). On the other hand, on the Result construal (paraphrase *and* as *and as a result*), *him* can refer to Powell if it is unaccented. All these data show a clear pattern whereby Resemblance relations (e.g. Parallel) require an unaccented pronoun to corefer with its parallel element, whereby pronouns in non-Resemblance (e.g. Result) relations are not similarly constrained.

In the remainder of this section, we argue that these facts are predictable from the manner in which different coherence relations partition utterances information structurally with respect to focus and background, and how this partition in turn determines the placement of accent on referring expressions (whether pronominal or not) within an utterance. The analysis explains the data that have been used to support a grammatical role parallelism preference without appeal to any pronoun-specific interpretation mechanisms or strategies.

3.5.2 *Coherence, coreference and accent* The idea that the aforementioned facts are unrelated to pronominalization goes against the common wisdom in the literature, which often treats accented pronouns in English as being governed by special rules or associated with specific discourse functions. For example, Kameyama (1999) proposes a Complementary Preference Hypothesis, which says that 'a focused pronoun takes the complementary preference of the unstressed counterpart', that

is to say, one first computes the preferred referent for an unaccented pronoun, and then selects an entity from the remainder of the 'currently salient' set of entities. Similarly, Beaver (2004) offers an analysis in which Kameyama's predictions result from partial blocking effects between accented and unaccented pronouns in a bidirectional-optimality-theoretic (OT) implementation of a Centering-based pronoun interpretation system. Smyth (1994) likewise posits that accented pronouns selectively block the parallel interpretation when the EFMH applies. Finally, Gundel *et al*. (1993), in their treatment of referring expressions and cognitive status, place unaccented and accented pronouns into two different categories (*in focus* and *activated*, respectively, p. 283, footnote 14).

However, it turns out that all the aforementioned facts concerning coherence and accentuation are actually constraints on *coreference* rather than merely *pronominalization* (Akmajian and Jackendoff 1970; Venditti *et al*. 2002; de Hoop 2004). This can be seen by considering variants of our previous examples in which the pronouns are replaced by proper name mentions of their referents. In all these cases, the requirements on accenting the direct object (marked using capital letters) are insensitive to whether a full name or pronoun is used:

(19) Condi Rice$_i$ admires Hillary Clinton, and George W. Bush absolutely worships $\left\{\begin{array}{c}\text{HER}_i\\ \text{RICE}\\ \#\text{her}_i\\ \#\text{Rice}\end{array}\right\}$. (cf. 15)

(20) Condi Rice$_i$ admires Donald Rumsfeld, and George W. Bush absolutely worships $\left\{\begin{array}{c}\text{HER}_i\\ \text{RICE}\\ \#\text{her}_i\\ \#\text{Rice}\end{array}\right\}$. (cf. 16)

(21) Powell$_i$ defied Cheney, and Bush punished $\left\{\begin{array}{c}\text{HIM}_i\\ \text{POWELL}\\ \#\text{him}_i\\ \#\text{Powell}\end{array}\right\}$. (cf. 18, on the Parallel reading)

Likewise, the lack of accenting on the pronoun in the Result cases remains when a proper name is used instead:

(22) Condi Rice$_i$ defeated $\left\{\begin{array}{c}\text{Hillary Clinton}\\ \text{John Kerry}\end{array}\right\}$ and George Bush congratulated $\left\{\begin{array}{c}\text{her}_i\\ \text{Rice}\end{array}\right\}$. (cf. 17a,b)

(23) Powell$_i$ defied Cheney, and Bush punished $\left\{ \begin{array}{c} \text{him}_i \\ \text{Powell} \end{array} \right\}$. (cf. 18, on the Result reading)

Therefore, the information structural constraint at work is one that relates coherence and coreference to accentuation, and is not specific to pronouns. Simply put, pronouns are *not* constrained to refer to their parallel elements in Parallel relations. Instead, the information structural constraints imposed by Parallel relations (but not Result relations) require that the pronoun, like any other referring expression, receive accent when it is not coreferential with its parallel element. The factors that determine the ability to pronominalize a mention and those that determine accentuation, while independent, interact to entail that unaccented pronouns in Parallel relations can only corefer with their parallel elements. As such, these data neither result from any special-purpose functions of accented pronouns nor can be used to support the existence of a grammatical role parallelism bias.

3.5.3 *An analysis* The facts described so far call instead for an explanation for why Parallel and Result relations differ information structurally, such that they impose different constraints on what elements of a sentence must receive accent. Kehler (2005) outlines an analysis, cast using the machinery of Schwarzschild's (1999) optimization-driven theory of focus marking and accent placement, that accounts for these differences. We only summarize the arguments here.[10] The crucial fact is that Parallel and Result relations will give rise to different F(ocus)-markings for otherwise similar (or, in the case of example 18, identical) examples, which in turn results in different distributions of accents. A brief discussion of Schwarzschild's system should suffice to understand the argument.

In Schwarzschild's analysis, *F-marking* serves as the interface between semantics and phonology. On the semantics side, felicitous utterances are entailed by the prior discourse (that is, *Given*), with the proviso that F-marking a phrase effectively turns it into a 'wildcard' (or 'F-variable') when matching against an antecedent. For instance, in a context that mentions *a red apple*, the NP *a* $[green]_F$ *apple* will be considered Given. On the phonology side, there is a constraint that *FOC*-marked nodes—F-marked nodes that are not immediately

[10] Readers who are not interested in the technical details of the argument can skip the remainder of this section without loss of continuity.

dominated by another F-marked node—must contain an accent. As such, the word *green* in *a [green]$_F$ apple* will require accent.[11]

In establishing Givenness, FOC-marked nodes are assigned discourse antecedents by a function *h*; in the example just given, *h* will map the denotation of *green* to that of *red*. In Schwarzschild's system, an OT-style optimization procedure solely determines *h*. Kehler (2005) argues against this aspect of the analysis, claiming that it cannot predict accent patterns for passages like example (18), repeated below as (24a), in which accent varies depending on the coherence relation inferred.

(24) a. Powell defied Cheney, and Bush punished him.
b. BUSH_{F_1} PUNISHED_{F_2} HIM_{F_3}. (Parallel, *HIM* = Powell)
c. BUSH_{F_1} $[\text{PUNISHED}_{F_2}$ $\text{him}]_{F_3}$. (Result, *him* = Powell)

Kehler claims that *h* assigns different mappings to the two coherence construals. In particular, the mapping established for the Parallel relation is precisely the one that results from the identification of parallel elements (i.e. the a_i and b_i). As such, if the pronoun *him* refers to Powell in (24a), the Parallel relation (and hence *h*) will enforce the following mapping between entities and predicates in the second clause (left side of the equations) and their parallel elements in the first clause (right side of the equations):

(25) a. $[\![\text{Bush}_{F_1}]\!]^{g,h} = [\![\text{Powell}]\!]^{g}$
b. $[\![\text{punished}_{F_2}]\!]^{g,h} = [\![\text{defied}]\!]^{g}$
c. $[\![\text{Powell}_{F_3}]\!]^{g,h} = [\![\text{Cheney}]\!]^{g}$

Loosely speaking, this F-marking results in the background *Who did what to whom*. Because *h* applies only to F-marked constituents, *him* must be F-marked for this mapping to hold, and by FOC must be accented despite it representing Given information, per (24b). On the other hand, if *him* refers to Cheney in (24a), *h* need not map it to a distinct entity as it would then be coreferential with its parallel element. In this case, $[\![\text{Cheney}]\!]^{g,h} = [\![\text{Cheney}]\!]^{g}$, and Cheney becomes part of the background (i.e. *Who did what to Cheney*).

Unlike the Parallel relation, however, F-marking in a Result relation is not governed by a pairwise mapping since its definition does not incorporate one. Instead, the F-marking in (24c) is favoured. In this partition, unlike that in (24b), Powell is part of the background,

[11] A variety of other rules and constraints are also at play, which we will not discuss here.

representing a shared variable in the causal relation used to establish coherence (e.g. the P in $defy(P, C) \rightarrow punish(B, P)$). As such, it is not F-marked, and thus need not receive accent.

The crucial fact to be abstracted from this brief synopsis is that Parallel relations, by way of establishing a mapping between parallel elements, give rise to a particular focus/background partition. A side effect of this partition is that a noun phrase (pronominal or not) that does not corefer with its parallel element will require accent *regardless of its Givenness status in the remainder of the discourse*. Result relations are not similarly restricted, and as such, the optimal focus/accent distribution will often result in the deaccenting of a noun phrase that denotes Given information without any parallelism restriction. Hence, we find different constraints at play in (24a) depending on the coherence relation that is construed. This analysis likewise explains the full set of interpretation patterns witnessed in (19–23), and in particular demonstrates how the resistance of the apparent parallelism bias to influences of semantic plausibility (19) and gender conflicts (20) results without recourse to any pronoun-specific principles.[12]

To summarize this section, our experiments and analysis show how a coherence-driven analysis predicts when evidence for the parallel grammatical role preference will emerge—particularly, in Resemblance relations like Parallel—and the underlying information structural reason why it does. As a result, there is no work left to be done by positing a separate parallel grammatical role bias or heuristic.

[12] The above analysis is restricted to cases in which a common relation over parallel entities comprises the background (the 'common topic'), which is the case in all of the examples we have considered. Oehrle (1981) notes that in other 'discourse frames' a pronoun can remain deaccented even when not coreferential with its parallel element, as in (i):

(i) A: Can you give me an exact description of Bill's role in the fight?
B: John hit Bill_b and he_b hit Max.

The difference between this example and the others is that the context sets up Bill's participation as the background, as opposed to the question *Who hit who?* Our analysis predicts this accent pattern given that A's question is the antecedent of both clauses in B's response, rather than the first clause of B's response serving as antecedent to the second.

Also, whereas we have focused on accented pronouns in Parallel relations since those are the cases relevant to our argument, accented pronouns can of course occur outside of Parallel relations. For the results of a corpus analysis see Wolters and Beaver (2001), who conclude that most instances of accented pronouns in their data can be seen as signalling rhetorical contrast, of which the examples discussed here would presumably constitute one type. See also Kehler (2005) for a discussion of examples that involve accented pronouns in Result relations—for example, *John_i pushed Bill and HE_i/ JOHN_i fell*—in which accentuation is similarly orthogonal to the form of referring expression used.

4 THEMATIC ROLE BIASES

A finding of Experiment 1 was that participants reliably interpret a subject pronoun to refer to a non-subject referent in Result relations if the semantics of the passage supports that interpretation. This possibility is not limited to Result relations, however. As we indicated in section 2, Stevenson *et al*. (1994) report on a series of story-completion experiments that suggest that the occupants of some thematic roles are systematically preferred to others. Of particular interest here are the patterns they found for passages with a transfer-of-possession context sentence followed by an ambiguous pronoun, as in (26):

(26) John handed a book to Bob. He_____

In such cases, the subject fills the Source role and the object of the prepositional phrase fills the Goal role. Participants were asked to provide a natural completion to the pronoun prompt provided in the second sentence, and the pronoun was then categorized as referring to the Source or the Goal. They found that Goal continuations, that is those which correspond to a Goal interpretation for the pronoun, occurred about as frequently as Source continuations (a 49–51% split).

The result seems intuitive enough: In a passage such as (27), in which the Occasion relation is operative, pronominal reference to Bob appears to be unobjectionable:

(27) John handed a book to Bob. He began reading it.

Yet this is unexpected in light of the grammatical subject and grammatical role parallelism preferences, since both point to John as the preferred referent. Whereas participants could have first assigned the pronoun using these biases and then written a continuation that accommodated that assignment, apparently this is not what happened.

Stevenson *et al*. describe two potential explanations for their result. The first is a thematic-role bias which amounts to a heuristic that ranks Goals above Sources. The second is a bias for focusing on the end state of the previously described event, under the assumption that the Goal is more salient to the end state than the Source. Stevenson *et al*. ultimately argue for the end-state bias; under this interpretation, the apparent heuristic preference for Goals is an epiphenomenon.

Our coherence analysis predicts an end-state bias, but only specifically for passages related by Occasion. Recall that in our analysis, the different biases underlying pronoun interpretation are ultimately traceable to properties of the inference processes that are used to

establish coherence. Among the coherence relations discussed in section 2, Occasion is the only one that specifically incorporates a bias towards focusing on the end state of the previous eventuality:

> **Occasion:** Infer a change of state for a system of entities from the assertion of S_2, establishing the initial state for this system from the final state of the assertion of S_1.

As such, the coherence analysis would predict that different pronoun interpretation biases will emerge for different coherence relations, and in particular, that Occasion relations will give rise to a Goal preference.

4.1 *Experiment 2*

An experiment was designed to distinguish the two possible explanations of Stevenson *et al.*, as well as to test the predictions of the coherence analysis (Rohde *et al.* 2006). Passages like (26) were paired with versions in which the imperfective form of the main verb was used (28).

(28) John was handing a book to Bob. He_____

Crucially, the thematic roles remain the same in examples (26) and (28), but the perfective verb in (26) describes a completed event which is compatible with end-state focus, whereas the imperfective verb in (28) describes an event as an ongoing process, making it incompatible with end-state focus (Moens and Steedman 1988). The thematic role preference thus predicts a similar distribution of Source and Goal interpretations between the two conditions, whereas the event-structure hypothesis predicts a greater percentage of Source interpretations in the imperfective condition than in the perfective condition. We focus the present discussion on testing these predictions, and will return to the predictions of the coherence analysis momentarily.

4.1.1 *Stimuli* Twenty-one experimental stimuli consisted of a transfer-of-possession context sentence followed by an ambiguous pronoun prompt, as in (26) and (28). Participants saw either the perfective or the imperfective form of each verb, but not both. The Source referent always appeared in subject position, and the Goal was always the object of a *to*-phrase. All verbs described physical transfer events (e.g. *hand, throw*); we excluded verbs that described abstract or conceptual transfer (e.g. *show, teach*).

We also included 29 filler passages with non-transfer verbs (transitive and intransitive) in the context sentence that varied between perfective or imperfective. The transitive verbs (Agent-Patient and

Experiencer-Stimulus) varied in active and passive voice. Adverbs, proper names or gender-unambiguous pronouns served as prompts.

4.1.2 *Participants* Forty-eight monolingual English-speaking undergraduates at UCSD participated in the study for extra credit in linguistics courses.

4.1.3 *Task* Our design followed Stevenson *et al.* closely. Participants were asked to write continuations for the 50 passages. They were instructed to imagine a natural continuation to the story, writing the first continuation that came to mind and avoiding humour. As noted by Arnold (2001), in this task participants create a mental model of the event described by the context sentence before writing a continuation; as such, the task involves both interpretation and production. While the prompt constrains the surface realization of the subject to a pronoun, we hypothesize that their continuation depends in part on their expectations about how the discourse will proceed and which individual in the event will be mentioned again.

4.1.4 *Evaluation and analysis* Two trained judges assessed the participants' intended pronoun interpretations. Judges were instructed to be cautious, erring on the side of categorizing a pronoun as ambiguous if the pronoun could be plausibly interpreted as coreferential with either referent, even if their personal interpretation biases strongly indicated a particular one. As such, not all responses could be disambiguated.[13]

4.1.5 *Results* The results, shown in Table 3, indicate that pronoun interpretation is sensitive to verbal aspect: Imperfective context sentences yielded significantly more Source interpretations (70%) than perfective sentences [51%; $F_1(1, 47) = 52.854$, $P < 0.0001$; $F_2(1, 20) = 30.079$, $P < 0.0001$].[14] As such, the event-structure hypothesis is supported over a thematic role bias, since the latter predicts no difference in the distribution of interpretations across conditions.

4.2 *Effects of coherence*

These results suggest that the Goal bias is at least in part an epiphenomenon of a bias towards focusing on the end state of the previous eventuality. We now examine the main prediction of the

[13] Our use of judges follows Arnold (2001). Stevenson *et al.* (1994) had participants circle their intended referents after completing the passages. However, they too ultimately relied on judges to remedy contradictions in the participants' circling.

[14] Table 3 excludes cases that were judged to be ambiguous.

	Perfective	Imperfective
Source	0.51	0.70
Goal	0.39	0.17

Table 3 Results of aspect manipulation

coherence analysis, specifically that the end-state bias will be primarily an epiphenomenon of establishing Occasion relations. As for other coherence relations, the predictions are as before: Resemblance relations (particularly Parallel) should favour a grammatically parallel antecedent, and Cause–Effect relations (e.g. Explanation, Result) will depend on the semantics incorporated in the passage and the referent to which causality or consequentiality is most likely to be imputed in a particular context.

To test this prediction, our judges annotated all unambiguous responses with the coherence relation that held between the context sentence and the continuation. Judges resolved disagreements through discussion, following Stevenson *et al.* (2000). Our analysis is restricted to the perfective cases since only these are compatible with end-state focus.[15] Six coherence relations were annotated: Occasion, Explanation, Result, Violated Expectation (another relation in the Cause–Effect category), Parallel and Elaboration (another relation in the Resemblance category), although Parallel, representing less than 2% of the continuations, is not analysed further.[16]

The results are shown in Table 4, which lists for each coherence relation its overall frequency and the percentage of pronoun interpretations to the Source. We found that Occasion relations were dominated by Goal continuations, whereas Elaborations and Explanations showed a Source preference [Occasion: $t(45) = 5.3537$, $P < 0.0001$; Elaboration: $t(42) = -19.66$, $P < 0.0001$; Explanation: $t(30) = -6.4983$, $P < 0.0001$ (one-sample t-tests)].[17] The restriction of the

[15] Analysis of the imperfective condition revealed a different distribution of coherence relations, but a highly similar relationship between each coherence relation and the corresponding distribution of Source and Goal interpretations. The fact that the different distributions in Figure 3 can be attributed to a different distribution in coherence relations across the perfective and imperfective conditions provides further support for the coherence analysis.

[16] This analysis is similar to one conducted by Arnold (2001), who ran a passage-completion experiment in a no-pronoun, full-stop condition, allowing participants to use either a pronoun or name to re-mention a referent at their discretion. Coding a coarser three-way split between *cause*, *end point* and *other* relations, she similarly found differences in the biases across continuation type.

[17] These t-tests use subject means. The results over item means are consistent [Occasion: $t(20) = 7.2642$, $P < 0.0001$; Elaboration: $t(19) = -69.7292$, $P < 0.0001$; Explanation: $t(19) = -9.1115$, $P < 0.0001$ (one-sample t-tests)].

Coherence relation	Percentage of corpus	Source bias
Occasion (171)	0.38	0.18
Elaboration (126)	0.28	0.98
Explanation (82)	0.18	0.80
Violated Expectation (38)	0.08	0.76
Result (25)	0.06	0.08

Table 4 Probabilities from Experiment 2 (perfectives)

Goal preference to Occasion relations reinforces the conclusion that a generic thematic role preference is insufficient as a predictor of pronoun interpretation.[18]

Whereas our results support the conclusion of Stevenson *et al*. that the Goal preference is an epiphenomenon of a bias towards focusing on end states, they further show that the end-state bias is to a large degree an epiphenomenon of the inference processes used to establish Occasion relations. The bias towards the Goal simply disappears when either of the other two common relations (Elaboration or Explanation) is operative. While the context sentences in all of our perfective stimuli describe events with salient end states, the results summarized in Table 4 strongly suggest that it is the coherence relation that dictates the extent to which that end point is relevant. Occasion relations exhibit a clear preference for the Goal, as they are precisely the relations that rely specifically on the end state of an eventuality in establishing coherence. Thus, thematic role biases constitute another case in which a coherence-driven analysis can explain the underlying reasons we see evidence for an interpretation heuristic, as well as why this evidence emerges only in particular contextual circumstances.

5 ONLINE INTERPRETATION

Summing to this point, the previous two experiments have provided support for a coherence-driven theory of pronoun interpretation.

[18] We also found a Goal bias for Result relations, but the small set of Result continuations (< 6%; $n = 25$) was very homogenous, more than half consisting of the form *X transfers Y to Z. Z thanks X*, making extrapolation difficult. Whereas our coherence analysis would predict that causal inference plays a greater role in establishing Result relations than Occasion relations, the effect described by the second eventuality in a Result sequence is often a direct result of the end state brought about by the first, and thus it would perhaps not be surprising to find an end-state bias for Result relations as well. This notwithstanding, Stewart *et al*. (1998) show that verbs are highly variable with respect to their biases in Result relations; see section 6 for further discussion.

Experiment 1 showed that the grammatical subject and parallel grammatical role preferences can be neutralized when coherence has been carefully controlled for in the stimuli. Experiment 2 supported the proposal of Stevenson *et al*. that event-structure biases are involved in pronoun interpretation (rather than thematic role biases), and furthermore localized them to those coherence relations that could be expected to encode such a bias as a side effect. In each case, we closely followed the design of the antecedent work to which we compared ourselves, which meant using offline methods for assessing interpretations.

An obvious remaining question for a coherence-driven theory is what it predicts about incremental processing. There is a wealth of online evidence that language interpretation proceeds in a highly incremental fashion, and pronoun interpretation has been a rich source of such evidence (Caramazza *et al*. 1977; Gordon and Scearce 1995; Stewart *et al*. 1998; Koornneef and van Berkum 2006, *inter alia*). The question is how coherence establishment can influence pronoun interpretation in cases in which the pronoun is encountered before the coherence relation is known. We begin addressing this question in this section, and then continue in the sections that follow with respect to two case studies: implicit causality (IC) effects and the grammatical subject preference.

Our proposal follows the lead of Arnold (2001), who hypothesized that referent accessibility is influenced by a hearer's probabilistic expectations about what referents will be subsequently mentioned in the discourse, which are in part driven by expectations about how the discourse is likely to be continued.[19] We focus our analysis on the role of *coherence-driven* expectations associated with discourse contexts in terms of two types of probabilistic information that are naturally combined: (i) expectations concerning how the discourse is likely to be continued with respect to coherence relation, and (ii) the likelihood that a certain referent will get mentioned by a pronoun *conditioned on the occurrence of that coherence relation*. These come together in the following equation (in which *ante* stands for an antecedent in a particular grammatical or thematic position, and *CR* stands for coherence relation):[20]

[19] It also follows recent work in sentence processing that contends that online measurements of interpretation difficulty can be successfully predicted by probabilistic, expectation-driven models (e.g. Hale 2001; Levy 2007). These models posit that the sentence processor implicitly makes predictions about what words are likely to come next in an utterance; degree of processing difficulty corresponds inversely with how well these expectations align with the material that is actually seen. Hale and Levy show that expectations can be estimated to good effect using generative models trained from online corpora, and that they predict a variety of reading time data that have been reported in the sentence processing literature.

[20] All terms in (29) are of course conditioned on the current context as well.

$$(29)\ P(pronoun = ante) = \sum_{CR \in CRs} P(CR) * P(pronoun = ante|CR)$$

For example, to compute the likelihood that a pronoun will corefer with the subject of the previous sentence, we simply sum, over all coherence relations, the likelihood of seeing that coherence relation multiplied by the likelihood of a subject reference given that coherence relation.[21] This equation makes explicit the idea that at any point during comprehension the hearer will have expectations about how the discourse will be continued with respect to coherence and that the difficulty in interpreting the linguistic material to follow will be conditioned in part on those expectations. These expectations will then evolve based on subsequent linguistic input that influences the probabilities represented.

Values for these terms need to be estimated in order to make predictions about online interpretation. However, we do not have direct access to the relevant probability distributions that language processors implicitly represent at a particular point in a discourse, nor is corpus analysis feasible if one desires a tight control on contextual factors. Instead, sentence-completion tasks like those used in Experiment 2 have become a standard way to estimate such biases (Caramazza *et al*. 1977; McKoon *et al*. 1993; Stewart *et al*. 1998; Koornneef and van Berkum 2006, *inter alia*). For the case of the perfective transfer-of-possession sentences used in Experiment 2, therefore, the two columns of biases shown in Table 4 provide estimates of $P(CR)$ and $P(pronoun = source|CR)$, respectively. When applied to (29), these numbers result in an average of 56.7% bias towards the Source at the time that a subject pronoun is encountered.[22]

Following a substantial previous literature that demonstrates that such biases impact reading times (see the review in the next section on implicit causality), these numbers would predict at most a modest reading time delay for Goal interpretations over Source ones. While the overall results are similar to the near 50/50 split found by Stevenson *et al*. (1994), our results show that there is nothing 50/50 about the pattern once coherence is taken into account. Each of the coherence relations encodes a considerably stronger bias one way or the other about who

[21] This formula is no doubt too simplistic as a full theory of probabilistic pronoun interpretation (one reason will be discussed in section 7); however, we can nonetheless use it for current purposes to illustrate how our analysis can make predictions about incremental processing.

[22] The bias towards the Source was reported as 51% in section 4.1.5, which was the percentage in a Source/Goal/ambiguous distinction (see footnote 14). The 56.7% bias reported here represents normalized percentages after setting aside the ambiguous cases.

will be mentioned next; it is only after the frequencies of coherence continuation are factored in that the biases have a cancelling effect.

Equation (29) further predicts that other phenomena which influence the likelihood of the upcoming coherence relation could impact pronoun interpretation biases, and as such, influence reading times. An obvious example is coherence-constraining connectives. Consider the connective *because*, which is only consistent with the Explanation relation. Because the occurrence of *because* after a Source–Goal passage would essentially drive the probability of Explanation towards one and the others towards zero, the probabilities in Table 4 would predict an average 80% bias for a subject referent. In this case, we would expect proportionately longer reading times for pronouns that referred to the Goal as compared to the Source.[23]

This expectation-driven view of incremental processing contrasts with a common view in the literature, whereby surface-level features determine the initial referent assigned to a pronoun, to be later confirmed or contradicted by plausibility factors (e.g. Gordon and Scearce 1995, *inter alia*). We believe our analysis provides a more parsimonious account that simultaneously captures documented preferences based on surface cues and a range of phenomena that are problematic for them. We elaborate in the sections that follow, considering first the phenomenon known under the rubric of implicit causality.

6 IMPLICIT CAUSALITY

Perhaps the most well-studied phenomenon relevant to the interaction between coherence and pronoun interpretation involves the so-called Implicit Causality (IC) verbs. The literature on the topic is voluminous; out of necessity our discussion will not be comprehensive (but see Rudolph and Forsterling (1997) for a comprehensive review as of their writing). Consider (30a,b), from Caramazza *et al*. (1977):

(30) a. Jane hit Mary because she had stolen a tennis racket.
b. Jane angered Mary because she had stolen a tennis racket.

Intuitively, the pronouns in (30a,b) refer to Mary and Jane, respectively. The reason for the difference points directly at the matrix verb, since the passages are otherwise identical. Caramazza *et al*. (1977) conclude

[23] This prediction ignores the difference between pronoun interpretation in syntactically coordinate contexts (as in our experiments) v. syntactically subordinate ones (as would be the case within an adjunct headed by *because*). As there is evidence that this distinction may matter (Miltsakaki, 2001), it may ultimately need to be accounted for in a richer probabilistic model of pronoun interpretation.

that IC is a feature of verb roots that selects one entity as the 'probable instigator or causal source for a series of events', which is in turn responsible for the corresponding bias in pronoun assignment.

Importantly, as with any statistical bias, IC biases can be violated without rendering the passage ungrammatical or incoherent, for example, compare (30a) with *Jane hit Mary because she reacts violently to criticism*. Nonetheless, one might ask whether these biases affect reading times, insofar as clauses in which the pronoun assignment is incongruent with the preceding verb's IC bias should take longer to read than ones in which it is congruent. Caramazza *et al*. (1977) ran a reading time experiment to test this prediction. Norming was done in a previous study (Garvey *et al*. 1976, Experiment 1) using a sentence-completion task of the sort we employed in Experiment 2, in which participants were asked to write completions for fragments such as (31):

(31) Tom scolded Bill because he_____

The percentage of interpretations to a referent was used as a measure of the verb's bias; with (31), for instance, they found that *scold* encodes a strong bias towards its direct object (henceforth, an NP2 verb) as opposed to one that encodes a bias towards its subject (henceforth, an NP1 verb). This NP2 bias predicts that example (32a), in which the preferred referent is congruent with the bias, should be read faster than (32b), in which the preferred referent is incongruent with the bias.

(32) a. Tom scolded Bill because he was annoying.
b. Tom scolded Bill because he was annoyed.

Pairs of stimuli per (32a,b) were joined with two controls that used gender-unambiguous pronouns. The results confirmed the prediction; sentences with bias-inconsistent pronoun interpretations took longer to read than sentences with bias-consistent ones in both conditions.

There is an obvious relationship between these experiments and our coherence analysis, in light of the fact that the connective *because* is an explicit indicator of an Explanation relation. The results of Experiment 2 shown in Table 4 also revealed a set of biases, in this case for transfer-of-possession passages, in terms of both the likelihood of each possible coherence relation to follow and of mentioning a particular referent conditioned on each coherence relation. Interestingly, we found what could be characterized as an overall NP1 IC bias here as well, with an average of 80% of NP1 references in Explanation relations. Caramazza *et al*. (1977) note as a major finding of their work that the 'IC feature' can be best represented as a continuum, that is, when the bias is represented as the proportion of continuations that suggest NP1 as the

referent, the values range continuously between 0 and 1. This is exactly what the final term of (29) captures, although crucially these biases are conditioned on coherence relations.

As we alluded to in the previous section, the inclusion of *because* in the stimulus prompts typically used in the IC literature might do no more with respect to pronoun interpretation than to restrict the operative coherence relation to Explanation. This analysis predicts that the IC bias found in sentence completions using a *because* prompt as in (31) should align closely with the IC bias found for completions in a similar no-pronoun, full-stop condition when only those passages that participate in an Explanation relation are considered. To our knowledge, such an experiment has not been carried out to date.[24] We therefore ran a sentence-completion experiment to test this question. A positive outcome would suggest that IC effects are a microcosm of a more general set of biases that apply in all contexts, distinguishing themselves only with respect to the strength of their bias towards a particular referent when an Explanation relation is operative.

6.1 *Experiment 3*

This experiment tested whether the biases found for IC verbs in passages containing a *because* prompt, mimicking the design of Garvey *et al*. (1976, *inter alia*), are similar to those found for the Explanation relationships identified in responses within a full-stop condition. Because we are mainly interested in the coherence-driven biases towards referents generated by these different classes of verbs, a pronoun was not included in the prompts. All subsequent first-mentioned referents were therefore catalogued, regardless of form of reference (i.e. pronoun or proper name). This choice allowed the use of contexts that lacked gender ambiguity, which facilitated the identification of the intended referents of pronouns (cf. Stewart *et al*. 1998; Arnold 2001).

6.1.1 *Stimuli* A 2 × 3 design was used that crossed verb type (IC verb v. non-IC verb) with continuation type (full stop v. *because* v.

[24] Ehrlich (1980) ran an experiment in which the connective used was varied between *because*, *but* and *and*:

(33) a. Steve blamed Frank because he spilt the coffee.
b. Steve blamed Frank and he spilt the coffee.
c. Steve blamed Frank but he spilt the coffee.

Her results were mixed, which is not surprising on our analysis because neither *but* nor *and* select for a single coherence relation: *but* is consistent with both Contrast and Violated Expectation (which each have different biases), and *and* is consistent with Occasion, Result and Parallel (again, each having different biases). As such, this manipulation does not reveal much about the predictions of a coherence-driven theory.

dialogue prompt). The dialogue-prompt condition was included for norming data for an orthogonal future experiment and will not be further analysed here. Examples of the full stop and *because* condition are shown in (34):

(34) a. Tony disappointed Courtney. _____
b. Tony disappointed Courtney because _____

Forty IC verbs and 40 non-IC verbs were taken from McKoon *et al.* (1993), with three replacements. (The verbs *cheat*, *jeer* and *dread* were felt to sound awkward in our sentence frames, and were replaced by *offend*, *mock* and *fear*, respectively.) The IC category was further broken down into 20 each of NP1 and NP2 verbs. All context sentences contained mentions of two possible referents, one male and one female. Twenty filler sentences used non-IC verbs and were followed by various interclausal connectives (monologue continuation) or a dialogue response that contained the beginning of a question (dialogue continuation), for a total of 100 stimulus items per participant.

6.1.2 *Participants* Seventy-five monolingual English-speaking undergraduates at UCSD participated in the study for extra credit in linguistics courses.

6.1.3 *Task* The task followed the design of Experiment 2. Participants were asked to write the first natural completion that comes to mind, without adding extra humour or creativity to the task.

6.1.4 *Results* The results for NP1 verbs, NP2 verbs and non-IC verbs are presented in Tables 5, 6 and 7, respectively. Entries for coherence relations are not included if they comprised less than 5% of the continuations in the full-stop condition. This was sometimes the case for Violated Expectation and Occasion, and was always the case for Parallel.

Table 5 summarizes the results for the IC-NP1 verbs. The NP1 bias of 85% for Explanation relations in the full-stop condition is essentially equivalent to the 84% bias in the *because* condition, as predicted. [Prompt type is not a significant predictor of bias: $F_1(1, 70) < 0.0221$, $P < 0.8822$; $F_2(1, 19) = 0.032$, $P < 0.86$.] The lower 60% overall bias found in the full-stop condition simply represents a watering down of the IC bias due to the existence of passages with coherence relations other than Explanation, to which the IC bias is not relevant.

Table 6 summarizes the results for the IC-NP2 verbs. Again the Explanation bias towards NP1 in the full-stop condition is essentially equivalent to the one in the *because* condition, as predicted. [Prompt type: $F_1(1, 73) = 0.4424$, $P < 0.5081$; $F_2(1, 19) = 1.2235$; $P < 0.2825$.]

Coherence relation	Full stop		*because* prompt	
	P(CR) (%)	P(Subj|CR)	P(CR) (%)	P(Subj|CR)
Explanation	58	0.84	100	0.85
Result	22	0.10	—	—
Elaboration	10	0.61	—	—

Table 5 Probabilities from Experiment 3 (IC-NP1 verbs)

Coherence Relation	Full stop		*because* prompt	
	P(CR) (%)	P(Subj|CR)	P(CR) (%)	P(Subj|CR)
Explanation	62	0.13	100	0.10
Result	15	0.03	—	—
Elaboration	14	0.46	—	—

Table 6 Probabilities from Experiment 3 (IC-NP2 verbs)

Coherence Relation	Full stop		*because* prompt	
	P(CR) (%)	P(Subj|CR)	P(CR) (%)	P(Subj|CR)
Explanation	24	0.57	100	0.56
Elaboration	29	0.58	—	—
Result	22	0.24	—	—
Violated Expectation	13	0.40	—	—
Occasion	9	0.53	—	—

Table 7 Probabilities from Experiment 3 (non-IC verbs)

Finally, Table 7 summarizes the results for the non-IC verbs. We see that even for non-IC verbs, the average bias towards NP1 is consistent between the *because* condition and the Explanation relations in the full-stop condition. [Prompt type: $F_1(1, 61) < 1$, $P < 0.982$; $F_2(1, 36) = 1.4598$, $P < 0.2348$.] This provides further evidence that there is nothing special about IC verbs coupled with the connective *because*; *because* simply marks an Explanation relation, and the referent bias gets adjusted accordingly for IC and non-IC verbs alike.

The hypothesis is therefore confirmed: The IC biases seen in the *because* condition are highly consistent with those found for Explanation relations in the full-stop condition across all three verb types. As in Experiment 2, the summary statistics across coherence

relations hide the considerably stronger biases that are often a play when coherence relations are conditioned on.[25]

Tables 5–7 also bring to light that there is not one but two noteworthy biases that are associated with IC verbs. Besides the biases towards particular referents that have been our focus thus far, IC verbs are also shown to be significantly more likely to evoke Explanation continuations (60% for NP1 and NP2 continuations combined) than non-IC verbs (24%), regardless of which referent gets mentioned first. This suggests that the lexical semantics of IC verbs create a stronger-than-usual expectation for an explanation. This bias may have gone unnoticed in the literature because previous studies typically have used only *because* prompts or have otherwise not categorized the coherence relations operative in their passage completions.

All these results suggest that contexts trigger rich probabilistic information that is brought to bear during interpretation. In this sense, the IC biases that have been documented in the literature represent just one of up to 10 biases that are exhibited in the no-bias conditions of Tables 5–7 (i.e. a bias for each of five relations coming next and a bias for a particular referent given each relation). In fact, another one of these biases—towards a referent given a Result relation—was previously identified and termed *Implicit Consequentiality* by Stewart *et al.* (1998). (See also the discussion in Crinean and Garnham 2006.) Using response completions to passages such as *Because John annoyed Bill, he*, they identify verbs that have both NP1 and NP2 consequentiality biases, and demonstrate that these biases impact reading times. A prediction of our analysis would therefore extend the online findings found for both IC and implicit consequentiality to the biases found for a broader range of contexts, across all coherence relations per the probabilities assigned by (29).

6.2 *Immediate focusing v. clausal integration*

A more recent controversy has centered around when IC information is used, that is, whether the information is utilized early enough so as to

[25] These results are surprising for the analysis of Stevenson *et al.* (2000), who argue for a *semantic focusing* account over a (Hobbsian, coherence-based) *relational* account. Whereas we argue that connectives influence coherence establishment and coherence establishment in turn influences pronoun interpretation, in their analysis connectives constrain pronoun interpretation more directly by modifying the salience of entities, in a second role that they consider distinct from their role in constraining coherence establishment. The assumptions that they place on coherence-driven analyses are problematic, however, and do not adequately represent either our analysis or Hobbs's original proposal. Whereas we will not go into further detail on these matters, we do note that the alignment of biases between the full-stop condition for Explanation and the *because* condition can only be viewed as a coincidence in their theory.

essentially constitute a focusing mechanism (the *immediate focusing* account, e.g. McKoon *et al*. 1993), or instead is used only as part of a sentence-final clause integration process (the *clausal integration* account, e.g. Stewart *et al*. 2000). The clausal integration account predicts that IC effects will arise later during sentence interpretation than the immediate focusing account does, at least when a pronoun occurs early in the clause. Our analysis predicts aspects of both of these models: The biases we have documented should be available at the time the pronoun is encountered and hence should influence reading times at or soon after the pronoun, but so will subsequent words that affect the likely coherence relation, and as a result, the likely referent for a pronoun given that coherence relation.

These predictions are supported by the recent study of Koornneef and van Berkum (2006, KvB). Characterizing IC biases as 'probabilistic asymmetries' that 'reflect something more subtle about the way we use various sources of information in everyday language comprehension', KvB looked for mid-sentence reading delays caused by pronouns that are inconsistent with the bias of a preceding IC verb in two experiments with gender-unambiguous pronouns. In a word-by-word self-paced reading task, they found that words in the pre-critical region were read equally fast across the bias-consistency conditions, but readers slowed down right at a bias-inconsistent pronoun, with a significant main effect emerging at the first two words thereafter. In an eye tracking study that measured mean regression path durations, again no differences were measured in the pre-critical region, but pronouns that were inconsistent with the IC bias reliably perturbed the reading process at or shortly after the pronoun. The results of both experiments therefore suggest that IC information becomes available rapidly enough to appear mid-sentence, even in passages in which the gender of the pronoun singles out a unique referent.

While these results support the immediate focusing account over the clausal integration account, KvB do not discount the latter entirely:

> However, our findings are not necessarily inconsistent with an *incremental* clausal integration account, in which the information made available by the subordinate clause is 'retroactively' related to the interpretation of the main clause *on a word-by-word basis*. (p. 459)

This view, which they similarly cast in terms of probabilistically driven expectations, is precisely the type of account that we advocate; it is evident from KvB's discussion that there is a close relationship between our respective views of expectation-driven discourse interpretation and

how it influences pronoun interpretation. We therefore consider KvB's online results as initial evidence for our approach, and would predict similar results from violating other coherence-driven biases as well. Online investigations of these additional hypotheses must await further research.

To summarize this section, our results suggest that IC biases are simply microcosms of a more general system of coherence-driven biases that drive pronoun interpretation in all context types. These results also show that IC verbs are exceptional with respect to two biases they engender: In addition to previously known biases towards a particular referent in an Explanation context, they also generate stronger-than-usual expectations for an upcoming Explanation relation.

7 THE GRAMMATICAL SUBJECT PREFERENCE

Finally, we evaluate the evidence for a grammatical subject preference in light of our analysis and results. Crawley *et al*. (1990) report on two studies which they argue support the idea that hearers use a subject assignment strategy, contrasting it specifically with the predictions of a parallel function strategy. They characterize such strategies as 'relatively mechanical rules of thumb which tell us to whom or what to assign a pronoun', which are nonetheless only invoked when 'there are no other strong constraints (such as linguistic or pragmatic constraints) on assignment'. In a self-paced reading task, participants read a three-sentence passage that ended with a clause that contained a pronoun in object position. (We restrict discussion to their ambiguous pronoun condition.) Although they acknowledge the difficulty in completely eliminating the influence of general knowledge in their stimuli, three judges checked each stimulus to ensure that either assignment of the pronoun resulted in a plausible interpretation. Participants answered a question that revealed their pronoun assignments. A bias was found towards the grammatical subject over the object, with an average of 23.7 subject interpretations for the 40 passages (a 59.25% bias). It also took slightly longer to read sentences with object referents. The results of a direct assignment task using the same stimuli were very similar, with an average of 24 subject assignments (60% subject bias).

Despite their conclusions, however, nothing in their experiment rules out the possibility that their results arose from discourse-driven expectations generated by their stimuli rather than from a distinct pronoun-specific interpretation strategy. Furthermore, as Smyth (1994) points out, properties of their reading time data suggest that some

stimuli may have been consistently interpreted with subject assignment and others consistently with object assignment. As such, perhaps a different set of stimuli would have yielded a different result.

To shed light on this issue, we analysed the biases found for the non-IC verbs in the full-stop condition in our Experiment 3. A representative sample of 10 of the 40 verb frames used in that experiment, all of which were from the stimuli of McKoon *et al.* (1993), are shown in Table 8.[26]

As can be seen, the NP1 biases show an even representation across the spectrum from 0 to 1. (The entire set of 40 verbs showed the same even distribution as well, with exactly half of the verbs having a bias above 0.5 and half falling at 0.5 or below.) Granted there are several differences between these biases and those found by Crawley *et al.*: These were collected from a sentence-completion study without a pronoun prompt, and represent first-mentioned referents rather than only subjects. Nonetheless, these results highlight the degree of freedom afforded in the selection of stimuli. Selecting verb frames from the top half of the table would presumably tilt the results towards a subject assignment strategy, whereas verb frames from the bottom half would presumably tilt the evidence away from it.

Whereas researchers who posit the existence of pronoun interpretation preferences and heuristics have consistently exempted 'pragmatically biased' examples, viewing stimuli in terms of the statistical distributions they engender makes it clear that there really is no such thing as a passage that is devoid of pragmatic bias. The numbers may be stronger or weaker, but any context will give rise to a set of biases over continuation types with respect to coherence and a set of biases for likelihood of mention given a continuation type. It is therefore incumbent on researchers to explain exactly what counts as a plausibility factor when using it to exempt examples that fail to conform. In the case of the subject assignment strategy, such conditions would not only have to include the intuitively high-bias NP2 verbs in the IC literature, but also the seemingly mundane perfective transfer-of-possession passages from Experiment 2 and the bottom five verb frames in Table 8, all of which do not follow the predictions of the strategy.

This argument should not be misconstrued to suggest that a bias towards subjects would not emerge if one could compute statistical expectations over all possible contexts. For instance, one might expect that many verb frames are frequently continued with Occasion

[26] Although 25 participants saw each of these in the full-stop condition, the actual number of qualifying entries was generally less, since some did not mention either referent, or mentioned both at once with the pronoun *they*. All but one of the entries included in the table had at least 20 qualifying continuations (*saw* had 18).

Verb frame	p(NP1)
borrowed-a-bike-from	0.857
saw	0.722
waited-to-see	0.636
counted-the-money-from	0.545
played-the-piano-for	0.500
edited-an-essay-for	0.400
repaired-a-bike-for	0.350
watched	0.261
went-to-visit	0.200
read-a-funny-story-to	0.130

Table 8 Biases for 10 selected verbs from Experiment 3 (non-IC verbs)

relations, and that many Occasions will display continuity in the agent role—and as a result, oftentimes continuity in the subject position as well—a combination that would tilt the statistical biases towards the subject position. The point is that this bias then emerges from general mechanisms without any need to posit a separate heuristic. Indeed, a coherence-driven theory is in principle capable of explaining such overall biases while still capturing the differing behaviour of certain other verb frames, using the same types of predictive interpretation mechanisms that we find evidence for in sentence processing.

A final observation is in order, however, as experiments by Stevenson *et al*. (1994) provide a type of evidence for a subject bias that we have yet to address. Recall that in addition to their pronoun-prompt condition, Stevenson *et al*. had a no-pronoun, full-stop condition (as we used in our Experiment 3), in which participants chose their own forms of referring expressions. Across their stimulus types, they found that this choice was heavily biased towards a pronoun when the referent was the previous subject, and likewise towards a name when the referent was a non-subject. (Arnold (2001) found similarly strong biases.) At first blush this result seems paradoxical: If participants have a clear production preference to refer to non-subjects with full names, why do they so readily assign a pronoun to a non-subject in the pronoun-prompt condition (e.g. 49.0% to the Goal in Source–Goal contexts)?

Stevenson *et al*. suggest that, in addition to a general thematic role preferences, 'heuristic search processes triggered by the presence of a pronoun' provide an additional bias to the first-mentioned entity, that is, that there is an overlaid subject assignment strategy. This suggestion would explain, for instance, why they found far more references to the Goal in their Goal-Source pronoun-prompt condition (84.6% by our

calculation), in which the Goal is also the subject, than in the Source–Goal condition (again, 49.0%).

However, there are other possible explanations for these results that do not require appeal to any specific interpretation heuristics. As an illustration, we consider the relationship between pronoun production and interpretation that emerges when cast in Bayesian terms:

$$P(referent|pronoun) = \frac{P(pronoun|referent)P(referent)}{P(pronoun)}$$

Whereas up to this point we have considered pronoun interpretation biases ($P(referent|pronoun)$) as conditioned by coherence-driven expectations, this formulation splits the bias into two: an expectation towards a subsequent mention of a referent ($P(referent)$), and an expectation about the form of referring expression that the speaker would use to mention that referent ($P(pronoun|referent)$).[27] Under this formulation, there is nothing inconsistent about an interpretation bias towards a non-subject referent despite a strong bias against pronominalizing non-subjects, assuming a suitably large subsequent mention bias towards the non-subject.[28] Our results and those of Stevenson *et al.* and Arnold are therefore all consistent with a scenario in which grammatical or information structural factors (subjecthood, topichood) play a greater role in conditioning $P(pronoun|referent)$ and coherence-driven expectations play a greater role in conditioning $P(referent)$. If this is the case, we would expect to find a pronominal bias towards the subject position beyond what is predicted from coherence-driven expectations alone (exempting Parallel relations, per the arguments in section 3), although importantly, without the need to posit that hearers utilize pronoun-specific *interpretation* strategies or heuristics.[29]

To conclude this section, our analysis of seemingly unremarkable verb frames as exemplified in Table 8 reveals great variance in their prior contextual biases towards particular referents. It would therefore seem essential that such biases be controlled for before the existence of an overlaid subject assignment preference can be established. Further, while we suspect that there are sources of pronoun-specific subject biases in pronoun

[27] We will ignore the term $P(pronoun)$, which is a constant factor over all possible referents in the context.

[28] Such a situation occurred in Arnold's (2001) Source–Goal condition. She found that 76.0% of the references to the subject were pronominalized, whereas only 20.1% of references to the object of the preposition were. However, the next mention bias towards the Goal was an overwhelming 85.6%.

[29] Having made this point, we want to stress that it is not our goal to argue for this Bayesian analysis, as it raises a large number of questions that we are not prepared to address. We only wish to offer it as a proof-of-concept of how a subject bias in interpretation could emerge beyond what is predicted by coherence-driven expectations alone. A fuller exploration of the model is the subject of ongoing work.

interpretation, they do not necessarily entail the existence of special-purpose, heuristic interpretation 'strategies', but instead may ultimately prove to be better captured within a more parsimonious, expectation-driven account.

8 CONCLUSION

We have presented new experimental evidence in support of a coherence-driven analysis of pronoun interpretation, and described how it can accommodate previous findings suggestive of conflicting preferences and biases. The results of our first experiment demonstrated that the grammatical subject and grammatical parallelism preferences can be neutralized when coherence is carefully controlled for. We furthermore provided a linguistic analysis that establishes that the grammatical role parallelism preference is an epiphenomenon of an independent interaction between information structure and accent placement in Parallel coherence relations that applies to referring expressions of all types.

The results of the second experiment distinguished the thematic role and event-structure biases proposed by Stevenson *et al*. (1994), supporting the event-structure bias. The experiment further showed that the bias is limited primarily to those coherence relations which implicate event structure in their formulation, and that the approximately 50/50 distribution of references found in Source–Goal passage completions represents but an average of a set of considerably stronger biases evident when coherence relations are conditioned on.

Whereas evidence for incremental interpretation has historically been seen as problematic for coherence-based analyses, we have described a model that captures how a hearer's coherence-driven expectations about how the discourse is likely to proceed could predict online measurements of pronoun interpretation difficulty. The results of Experiment 3 confirmed a prediction of this analysis, specifically that IC biases evident in passage completions with *because* prompts are essentially equivalent to those in a full-stop condition when only Explanation relations are analysed. The results of this experiment also demonstrate that IC biases represent but one instance of a more comprehensive set of biases that drive predictive discourse interpretation, which include biases for what type of continuation will ensue in addition to biases towards mentioning particular referents conditioned on continuation type. Although online tests of the predictions of the analysis await future work, biases estimated from passage-completion experiments have been repeatedly shown in the literature to influence pronoun processing difficulty.

Finally, we described how coherence-driven expectations about who will be mentioned next have the potential to dramatically affect

evidence for a grammatical subject preference. We also speculated that there are subject biases in pronoun interpretation that go beyond what can be predicted by coherence-driven expectations alone, and how these might be explainable without recourse to any heuristic interpretation 'strategies'. A suitably comprehensive evaluation of the tenability of this approach must await future work, however.

In sum, the coherence analysis is capable of explaining a wide variety of often contradictory results in the previous literature in a theoretically parsimonious manner. It offers an explanation of what the underlying sources of previously proposed biases are, and predicts in what contexts evidence for each will surface. The theory finds no need to include caveats for examples with 'pragmatic bias', since the theory directly captures the fact that all passages contain pragmatic bias. A ramification for future psycholinguistics work is the need to control for the pronoun-independent, coherence-driven expectations that are embodied in experimental stimuli, as our results argue that this is required before evidence for overlaid biases or preferences can be successfully established.

Acknowledgements

We would like to thank Jennifer Arnold, Julie Sedivy, and Ron Smyth for extensive comments that led to substantial improvements over earlier drafts, and Roger Levy, Bob Slevc, and audiences at the University of Pennsylvania, MIT, UCLA, Ohio State, the University of Minnesota, San Diego State, the University of Texas at Austin, and the UCSD Center for Research in Language for useful comments and discussions. We also thank Erica Gold for her extensive annotation efforts. This research was supported in part by National Institutes of Health Training Grant 5-T32-DC0041 to the Center for Research in Language at UCSD, and a grant from the UCSD Academic Senate.

ANDREW KEHLER
Department of Linguistics
University of California
San Diego 9500 Gilman Drive
La Jolla CA 92093-0108
USA
e-mail: akehler@ucsd.edu

REFERENCES

Akmajian, A. & R. S. Jackendoff (1970), 'Coreferentiality and stress'. *Linguistic Inquiry* **1**:124–6.

Arnold, J. (2001), 'The effect of thematic roles on pronoun use and frequency of reference continuation'. *Discourse Processes* **21**:137–62.

Beaver, D. I. (2004), 'The optimization of discourse anaphora'. *Linguistics and Philosophy* **27**:3–56.

Caramazza, A., E. Grober, C. Garvey, & J. Yates (1977), 'Comprehension of anaphoric pronouns'. *Journal of Verbal Learning and Verbal Behaviour* **16**: 601–9.

Chambers, C. C. & R. Smyth (1998), 'Structural parallelism and discourse coherence: a test of centering theory'. *Journal of Memory and Language* **39**: 593–608.

Crawley, R. A., R. J. Stevenson, & D. Kleinman (1990), 'The use of heuristic strategies in the interpretation of pronouns'. *Journal of Psycholinguistic Research* **19**:245–64.

Crinean, M. & A. Garnham (2006), 'Implicit causality, implicit consequentiality, and semantic roles'. *Language and Cognitive Processes* **21**:636–48.

de Hoop, H. (2004), 'On the interpretation of stressed pronouns'. In R. Blutner and H. Zeevat (eds.), *Optimality Theory and Pragmatics*. Palgrave Macmillan. New York. 25–41.

Ehrlich, K. (1980), 'Comprehension of pronouns'. *Quarterly Journal of Experimental Psychology* **32**:247–55.

Garvey, C., A. Caramazza, & J. Yates (1976), 'Factors underlying assignment of pronoun antecedents'. *Cognition* **3**: 227–43.

Gordon, P. C. & K. A. Scearce (1995), 'Pronominalization and discourse coherence, discourse structure and pronoun interpretation'. *Memory and Cognition* **23**:313–23.

Gundel, J. K., N. Hedberg, & R. Zacharski (1993), 'Cognitive status and the form of referring expressions in discourse'. *Language* **69**:274–307.

Hale, J. (2001), 'A probabilistic Earley parser as a psycholinguistic model'. In *Proceedings of the Annual Meeting of the North American Chapter of the Association for Computational Linguistics*. The Association for Computational Linguistics. Morristown, NJ. 159–66.

Hobbs, J. R. (1979), 'Coherence and coreference'. *Cognitive Science* **3**:67–90.

Hobbs, J. R. (1990), *Literature and Cognition*. CSLI Lecture Notes 21. Stanford, CA.

Hume, D. 1748), *An Inquiry Concerning Human Understanding*. The Liberal Arts Press. New York. 1955 edition.

Kameyama, M. (1999), 'Stressed and unstressed pronouns: complementary preferences'. In P. Bosch and R. van der Sandt (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge University Press. Cambridge. 306–21.

Kehler, A. (2002), *Coherence, Reference, and the Theory of Grammar*. CSLI Publications. Stanford, CA.

Kehler, A. (2005), 'Coherence-driven constraints on the placement of accent'. In E. Georgala and J. Howell (eds.), *Proceedings of the 15th Conference on Semantics and Linguistic Theory (SALT-15)*. CLC Publications. Cornell University. 98–115.

Kertz, L., A. Kehler, & J. L. Elman (2006), 'Grammatical and coherence-based factors in pronoun interpretation'. In R. Sun (ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates. Mahwah, NJ. 1605–10.

Koornneef, A. W. & J. J. A. van Berkum (2006), 'On the use of verb-based implicit causality in sentence comprehension: evidence from self-paced reading and eye-tracking'. *Journal of Memory and Language* **54**:445–65.

Levy, R. (2007), 'Expectation-based syntactic comprehension'. *Cognition*, forthcoming.

McKoon, G., S. B. Greene, & R. Ratcliff (1993), 'Discourse models, pronoun resolution, and the implicit causality of verbs'. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition **18**:266–83.

Miltsakaki, E. (2001), 'Toward an aposynthesis of topic continuity and intrasentential anaphora'. *Computational Linguistics* **28**:319–55.

Moens, M. & M. Steedman (1988), 'Temporal ontology and temporal reference'. *Computational Linguistics* **14**:15–28.

Oehrle, R. T. (1981), 'Common problems in the theory of anaphora and the theory of discourse'. In H. Parret, M. Sbisà, and J. Verschueren (eds.), *Possibilities and Limitations of Pragmatics*. John Benjamins. Amsterdam. 509–30. Studies in Language Companion Series, Volume 7.

Rohde, H., A. Kehler, & J. L. Elman (2006), 'Event structure and discourse coherence biases in pronoun interpretation'. In R. Sun (ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates. Mahwah, NJ. 617–22.

Rudolph, U. & F. Forsterling (1997), 'The psychological causality implicit in verbs: a review'. *Psychological Bulletin* **121**:192–218.

Schwarzschild, R. (1999), 'Givenness, AvoidF, and other constraints on the placement of accent'. *Natural Language Semantics* **7**:141–77.

Sheldon, A. (1974), 'The role of parallel function in the acquisition of relative clauses in English'. *Journal of Verbal Learning and Verbal Behavior* **13**: 272–81.

Smyth, R. (1994), 'Grammatical determinants of ambiguous pronoun resolution'. *Journal of Psycholinguistic Research* **23**:197–229.

Stevenson, R. J., R. A. Crawley, & D. Kleinman (1994), 'Thematic roles, focus, and the representation of events'. *Language and Cognitive Processes* **9**:519–48.

Stevenson, R. J., A. Knott, J. Oberlander, & S. McDonald (2000), 'Interpreting pronouns and connectives: interactions among focusing, thematic roles, and coherence relations'. *Language and Cognitive Processes* **15**:225–62.

Stewart, A. J., M. J. Pickering, & A. J. Sanford (1998), 'Implicit consequentiality'. In M. A. Gernsbacher and S. J. Derry (eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates. Mahwah. NJ. 1031–6.

Stewart, A. J., M. J. Pickering, & A. J. Sanford (2000), 'The role of implicit causality in language comprehension: focus versus integration accounts'. *Journal of Memory and Language* **42**: 423–43.

Venditti, J. J., M. Stone, P. Nanda, & P. Tepper (2002), 'Discourse constraints on the interpretation of nuclear-accented pronouns'. In *Proceedings of the 2002 International Conference on Speech Prosody*. ISCA Archive. Aix-en-Provence, France. 675–8.

Winograd, T. (1972), *Understanding Natural Language*. Academic Press. New York.

Wolf, F., E. Gibson, & T. Desmet (2004), 'Discourse coherence and pronoun interpretation'. *Language and Cognitive Processes* **19**:665–75.

Wolters, M. & D. Beaver (2001), 'What does *he* mean?' In J. D. Moore and K. Stenning (eds.), *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*. Lawrence Erlbaum Associated. Mahwah, NJ. 1176–80.

First version received: 01.08.2006
Second version received: 20.06.2007
Third version received: 05.10.2007
Accepted: 09.10.2007