To appear in: In N. Pearlmutter & E. Gibson (eds). *The Processing and Acquisition of Reference*. MIT Press: Cambridge, MA.

The Effect of Speaker-Specific Information on Pragmatic Inferences

Daniel Grodner and Julie Sedivy
Brown University

Address correspondence to:
Daniel Grodner
Box 1978
Brown University
Providence, RI 02912

email:dgrodner@brown.edu
Office: (401) 863-3989
Fax: (401) 863-2255

(in press) The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (eds). *The Processing and Acquisition of Reference*. MIT Press: Cambridge, MA.

Introduction

It is commonplace to observe that utterances can convey more information than they explicitly encode. Indeed, the field of pragmatics has grown from the insight that speakers exploit communicative conventions in order to say more with less. Less attention has been paid to the burden this places on perceivers. Because speaker meaning is underspecified, perceivers must infer implicit content in order to interpret, and successfully situate, utterances within the discourse. These inferences[1] appeal to many types of knowledge including entailment relations, world knowledge, and the discourse context. The last of these presents a particular challenge to investigators because contexts are dynamic and utterance meaning can be context sensitive. For instance, the utterance in (2) implies something like (2a) if it is a response to (1a) and something like (2b) if it is a response to (1b).[2]

(1) a. What time is it?

b. How good is the party?

(2) Some guests are already leaving

a. It must be late.

b. The party is not much fun.

The ease and prevalence of such inferences facilitate efficient communication, but at the same time create difficulty for models of language understanding. A central puzzle is how the extrasentential context is combined with intrasentential information to ultimately yield an interpretation of a sentence.

The present paper addresses this issue by exploring a particular dependency between the referential environment and linguistic form. Suppose a speaker wishes to refer to one member of a set of entities belonging to the same nominal category in the current discourse. They must use a modified expression in order to refer successfully. For instance, if one cup is the intended referent in a context containing two cups, the speaker must use a restrictive modifier such as "the cup on the left" or "the red cup." Significantly, this dependence appears to be bidirectional. Upon encountering a restrictively modified noun phrase (NP), such as *the tall cup*, two sets are invoked in the immediate discourse: (1) a target set corresponding to the literal denotation of the

[1] Here and throughout we use the term "inference" to refer to information that is communicated to the perceiver via the utterance of an expression, but which is not part of its asserted content. This would include both accommodated presuppositions and implicatures. We adopt this term because it is neutral with respect to whether the inferred content arises from the conventional or implicated meaning of the critical expression.

[2] This example is adapted from Levinson (2000).

expression (e.g., a tall cup), and (2) a contrast set containing an entity of the same type as the noun, but differing along the dimension picked out by the adjective (e.g., a short cup)

Indirect evidence for the contrastive inference comes from studies of structural ambiguity resolution. In general, the sentence processing mechanism prefers syntactic alternatives that contain simple unmodified NPs when given the option. Crain & Steedman (1985) proposed that this preference stems from the fact that modified structures occasion costly changes to the discourse model. To illustrate, consider the string, *The horse raced past the barn ...*, which is ambiguous between a main clause containing a simple NP subject and reduced relative clause (RC) modifying the subject. The simple NP reading presupposes the existence of a single referent set, namely a horse, in the current discourse model. The complex NP reading involves the projection of an additional contrasting set of entities that share the properties denoted by the head noun, but differ by virtue of the property expressed in the modifier. For the present example this corresponds to a non-empty set of horses that were not raced past the barn. In a null context, the modified NP reading requires the greatest number of additions to the current discourse model in order to be felicitous. Thus, the simple NP is preferred. There is evidence that by establishing appropriate referent and contrast sets, it is possible to alter parsing preferences (e.g., Altmann, Garnham, & Dennis, 1992; Altmann & Steedman, 1988; Sedivy, 2002; Spivey-Knowlton & Tanenhaus, 1994).

More direct evidence of contrastive inference comes from studies that monitor perceiver's eye movements as they rearrange objects according spoken instructions (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Sedivy, 2003). Sedivy et al. had individuals respond to instructions containing a prenominally modified NP such as, "Pick up the tall cup." Object arrays consisted of four entities 1) a target object (e.g., a tall cup), 2) a competitor from a different nominal category that shared the modifier property of the target (e.g., a tall pitcher), 3) an irrelevant distracter object, and 4) either a contrasting object of the same category as the target but possessing a distinct modifier property (e.g., a short cup), or a second distracter. Figure 1 depicts a sample display. Note that an indeterminacy is introduced when the modifier is uttered. The input to this point is compatible with reference to either the target or the competitor. If perceivers are driven by the literal meaning of the modifier, then their attention could be directed to either referent. However, if they interpret the modifier contrastively then they should attend to the target object in the presence of a contrast. Consistent with this, individuals made earlier fixations on the target and fewer fixations on the competitor in the

presence of a contextual contrast. The influence of referential contrast on eye movements has been observed extremely early in the speech stream --within 200 ms of the onset of the head noun (Sedivy, in press). In light of estimates that saccadic eye-movements are planned 150-200 ms prior to being launched (e.g., Matin, Shao, & Boff, 1993), restrictive modifiers can generate the expectation for referential contrast well before the nominal head can be identified. Because the speech input does not uniquely isolate the target prior to utterance of the noun, perceivers' fixations must be guided by inferential content.

Two types of accounts have been forwarded to explain the effect of discourse contrast: form-based and pragmatic. Form-based accounts attribute the contrastive inference to the conventional content of restrictively modified NPs. This type of explanation has the advantage that it only requires consideration of information contained within the sentence. Perceivers would not have to deliberate over various types of extrasentential information. This comports well with the speed and automaticity of contrast effects.

One way for the conventional meaning to engender a contrastive inference is if definite modified NP structures conventionally presuppose contrasting entities in much the same way that a definite description might be said to conventionally presuppose uniqueness of reference. This idea was suggested by Steedman & Altmann (1989). However, this account is difficult to maintain in light of examples like (3), which indicate that the contrastive inference is cancelable in a way that conventionally-expressed content is not (Grice, 1975; Sadock, 1978).

(3) Use the tall cup because there are no other cups.

Perhaps a more plausible way to play out a form-based presuppositional account is to propose that modificational content is by default placed in focus. Focused material is typically analyzed as foregrounded against a background of contrasting alternatives (e.g., Rooth, 1985, 1992). So for instance focusing the modifier for the string above as in *The horse RACED PAST THE BARN...*, would make prominent a single horse that was raced past the barn against a presupposed background of a set of horses that were not raced past the barn. Note that a focus-mediated presuppositional account is consistent with the occurrence of modifiers in the absence of referential contrast. This is because modifiers need not always be placed in focus, and unfocused material is not predicted to generate a set of alternatives.

Reading time results from studies examining syntactic ambiguities lend some support to the focus-mediated interpretation of referential effects. Namely, the presence of the overt focus operator "only" can induce a preference for ambiguous material to be analyzed as a nominal modifier (Ni, Crain, & Schankweiler, 1996; Sedivy, 2002; but see Clifton, Bock, & Rado, 2000). However, it is difficult to extend this explanation to spoken stimuli involving prenominal modification (e.g. "Pick up the tall cup."). For these utterances, focus marking would result in greatest prominence on the adjective. However, referential contrast effects were found with spoken stimuli in which greatest prominence occurred on the head noun, according to rules of nuclear stress placement. Hence, the referential contrast effect does not seem to depend on focus-marking.

Another form-based possibility is that the effect of referential contrast is lexically driven. For instance, scalar adjectives, such as *tall*, are inherently relational and must be evaluated relative to a contextually salient comparison class. Clearly what counts as tall for a child is very different from what counts as tall for an elephant (Seigel, 1980; Bierwisch, 1987). It might be a lexical idiosyncrasy of scalar modifiers that they cause perceivers to attend inordinately to the discourse context and, thereby, receive contrastive interpretations. This provides an account for the referential contrast effect with scalar adjectives, but does not extend to referential effects with other modifiers, such as the postnominal modification involved in syntactic ambiguity resolution, or even referential effects found for non-scalar modifiers, such as modifiers denoting material properties (Sedivy, 2001). Thus, at the very least, any form-based approach seems to require a combination of at least two somewhat distinct mechanisms to account for the range of effects observed to date.

Perhaps the most striking evidence against any of the above accounts, either singly or in combination, comes from evidence that whether a modifier invokes a contrast can depend on what it modifies. Sedivy (2003) examined the processing of descriptions that contained prenominal color modifiers in the presence or absence of a contrasting object. In one condition, the adjective denoted a property that was highly predictable from the noun category, as in "the yellow banana." In another, the color property was not strongly associated with the noun, as in "the yellow cup." The presence of a contextual contrast aided individuals in identifying objects with predictable colors. For instance, individuals were faster to identify a yellow banana and directed fewer eye-movements to other yellow objects in the display, in the presence of a green banana. However, there was no benefit of contextual contrast conferred when identifying objects

with unpredictable color properties. Individuals' eye-movements did not converge more quickly on a yellow cup in the presence of a green cup. The contrastive inference is therefore not an inherent part of the meaning of the color modifier but rather is related to its informativity with respect to the entity being modified.

Such informativity effects are predicted to arise within a pragmatic explanation of referential contrast (Clifton & Ferreira, 1989; Sedivy, 2003). The pragmatic account claims that contrastive inferences arise because the use of a restrictive modifier is embedded in a collaborative communicative exchange. There is an understanding between discourse participants that speakers are only as informative as they need to be. This follows from Grice's second maxim of quantity: "Don't make your contribution more informative than is required for the purposes of the present exchange." (1975, p.26). A simple NP would suffice to pick out the intended referent in a context with only a single entity. The use of a more prolix or unusual form indicates that a different state of affairs prevails (see Levinson, 2000 for an extended treatment of this idea). The most natural way to make the inclusion of the modifier informative is to ascribe it a distinguishing function. To apply this logic, perceivers must reason counterfactually about what the speaker could have said, but chose not to.

In support of the pragmatic view, contrastive inferences only arise for adjective types that are not used to label objects in isolation. Scalar and predictable color modifiers are rarely encoded in default descriptions, whereas unpredictable color modifiers are frequently encoded in default descriptions (Sedivy, 2000; 2002). The connection between inferential patterns in comprehension and distributions in production suggests that perceivers are actively generating expectations as to what referential form a speaker will opt for in selecting an entity. Deviations from those expectations result in pragmatic inferences.

The appeal of this Gricean mechanism is that it is capable of adapting appropriately to a wide range of stimulus conditions. It also provides a unified explanation for all the referential context effects found to date. On the other hand, it requires a set of as yet unspecified computations that appeal to heterogeneous and potentially unconstrained knowledge bases. Many different factors might be weighed in deciding whether a modified form should be interpreted contrastively. These include the intrinsic properties of a referent, the discourse context, the reliability of the speaker, the intentions of the speaker, the common background of the interlocutors, the goals of the communicative situation, expectations about alternative forms, and so on. It is unlikely that

perceivers could consider all of these in the limited time frame where contextual contrast effects have been observed. In fact, Clifton & Ferreira argued that such conversational implicatures would be too cumbersome for individuals to engage in on-line, and, hence, an expeditious parser would not consider such information in initial interpretive processes.

The studies described above establish only that listeners are sensitive to the use of default descriptions when inferring a contrast. This need not imply that individuals are engaged in full-blown Gricean reasoning. It may be that certain inferential steps are short-circuited. This resembles the position of the so-called neo-Griceans (e.g., Gazdar, 1979; Horn, 1984; Levinson, 2000). These researchers propose a principled division between conversational implicatures that are generalized across situations and those that are situation specific. For instance, the implicatures in (2ab) are dependent on the particular context in which (2) is uttered. However, the scalar implicature in (2c) is present across of the vast majority of discourses.

(2c) Not all of the guests are leaving.

Levinson (2000) hypothesizes that generalized implicatures like this one form part of the meaning for certain utterance *types*, whereas particularized implicatures are associated with utterance *tokens*. For quantity-based implicatures, the form type selected by a speaker is located on an informativity scale with respect to some anticipated reference type (Horn, 1971, 1989). This comparison process can generate an implicature when the selected form is less informative than an alternative. For instance, the selection of "some" in (2) triggers a comparison with the more informative scalar alternative "all." The failure to use the more constraining alternative results in the implicature (2c). Analogously, contrastive inferences arise when the form selected is more informative when compared to an anticipated alternative.

The generality and simplicity of this comparison mechanism suggest that quantity-based inferences could arise via automatic processes rather than being computed on the fly. If the routines for projecting a contrast are precompiled in this way, this could explain the immediacy with which the visual context influences reference resolution. It also predicts that generalized inferences might be harder to block or suspend than particularized ones such as (2ab). Motivated by computational constraints, Jurafsky (to appear) outlines a similar idea to explain how perceivers interpret what speech act a speaker intended to perform. In this model, perceivers use surface cues probabilistically to realize the illocutionary force of an utterance. Because the implicit content arises from statistical associations, cues that are highly reliable indicators of implicit content should be harder to ignore. In the present example, this reasoning might predict

an insensitivity to the situation in which the modifier is used. Because default descriptions are statically linked with individual referents, contrastive inferences do not require that perceivers consider the particular circumstances of the immediate discourse. Perceivers might reflexively infer a contrast whenever the speaker deviates from a stored default form.

An alternative is proposed by Carston (1998). Building on the work of Sperber & Wilson (1986), he claims that all conversational inferences are actively constructed and therefore particularized to their circumstances. Such nonce inference is clearly a necessary component of language comprehension in order to explain the robustness of the inferences made in (2ab). What is at issue is whether this nonce mechanism is involved in the earliest stages of interpretive processing.

The experiment described below makes an effort to distinguish the short-circuited and nonce inference positions by manipulating characteristics idiosyncratic to a particular speaker -- specifically, the degree to which the speaker could be considered to be adhering to normative conversational principles. If contrastive inferences are generalized implicatures triggered by deviations from a default form, then the effect of contrast should be impervious to this manipulation, at least with respect to initial referential commitments. On the other hand, if contrastive inferences are the product of pragmatic reasoning about particular discourse situations, then there should be a noticeable effect when the circumstances are altered in this critical way.

Experiment

There is a great deal of evidence that speakers make all sorts of accommodations which demonstrate that they are sensitive to the specific needs of a listener. For example, experts will adjust referential forms appropriately to the knowledge state of the addressee (e.g., Isaacs & Clark, 1987) and speakers will also hyper-articulate to make speech clearer under conditions where intelligibility is likely to be reduced for a particular listener (e.g., Bradlow 2002). Knowledge specific to a conversational partner is also a potentially powerful constraint in comprehension. For instance, suppose (5) is uttered discourse initially.

(5) He's such a jerk.

Resolving the antecedent of the pronoun depends critically on realizing what potential referent is most likely to be salient for the individual speaker. In extreme cases, an utterance such as (5)

may be effortlessly resolved even when uttered out of the blue as the first interaction in days or even weeks between interlocutors, and with no particular supporting visual context if the interlocutors share knowledge of some male person extraordinarily inclined to behave badly.

Few studies have investigated the moment-to-moment application of speaker related knowledge in comprehension. Those that have, have employed a head-mounted eye-tracking methodology to monitor visual attention as listeners perform a target identification task. One important finding that emerges from this work is that listeners are sensitive to the speaker's perspective of the referential discourse. Listeners exhibit preferential consideration for entities in the shared referential environment over items to which they have privileged access (Hanna, Tanenhaus, & Trueswell, 2003; Nadig & Sedivy, 2002; but see Barr & Keysar, 2002; Barr, 2003). Hanna et al. (2003) provide a particularly powerful illustration of this ability. They show that listeners are sensitive to the speaker's discourse model even when it is at odds with the perceptually available referential environment. These findings illustrate that speaker-based information can be recruited very quickly to resolve potential referential indeterminacies. It is interesting to note, however, that these studies exploit the conventional properties of the referring expression – in these cases, the uniqueness requirement of definite descriptions – to signal the need to constrain reference. It has been argued that circumscribing a referential domain is central to communication (e.g., Chambers et al., in press; Hanna et al., 2003; Tanenhaus, 2003). The requirement that definite NPs refer to a uniquely identifiable entity, combined with an experimental scenario in which satisfying this requirement necessitates accessing speaker-based information together create optimal conditions for observing powerful and rapid speaker-based effects.

An interesting extension of such research is the question of whether perceivers access speaker-based information even when the conventional properties of the referring expression can be satisfied without recourse to speaker-based information. Metzing & Brennan (2003) report the use of speaker-based knowledge to interpret referring expressions in just such a situation. They showed that perceivers are sensitive to the way a particular speaker has chosen to refer to an object in the past, and will use this information inferentially. When the same speaker used a different, but equally plausible, label for an item, looks to that object were delayed, and there were more fixations to other objects in the display, before converging on the target referent. Participants expected the new label to refer to a new object and interrogated the scene in an attempt to find it. This could reflect a pragmatic inference arrived at when individuals try to find a relevant function for the departure from an entrained label (cf. Grice's maxim of manner).

Intriguingly, no such penalty was observed when a new speaker used the novel label. Participants were equally quick to respond to the new or old label when uttered by a new speaker. However, caution must be used in concluding that these findings demonstrate immediate use of speaker identity in inferencing. First fixations to the target object were launched approximately one second after the onset of the referring expression on average. This is slower than the latencies seen in comparable tasks and raises the possibility that participants adopted specialized strategies or that eye-movements were being driven by non-primary processes. Still the finding demonstrates that individuals are capable of applying the pragmatic expectation that a speaker will use a specific entrained description at some point in reference resolution.

The present experiment extends the Metzing & Brennan result by looking at whether perceivers generate expectations about the *type* of referring expression a speaker will use, and then use this as the basis for generating inferences of referential contrast. In this study we constructed a scenario in which a particular speaker does not obey the standard communicative conventions. Specifically, a number of cues were given to indicate that the speaker's use of restrictive modification was not a reliable signal to the presence of a contextual contrast. If the contrastive inference is predicated on the perceiver's belief that the speaker rationally chose one linguistic form over another, then the contrast effect should be defeated in this situation. If instead the inference is a reflex of low-level informativity expectations, then this manipulation should not affect the projection of a contrast and the presence of a contextual contrast should aid listeners in identifying the target.

A further manipulation in the present study was to look a the differential properties of material and scalar modifiers. Material properties are not context dependent in the same way as scalar properties. Nevertheless, just as for scalar and redundant color adjectives, the referents for NPs containing material modifiers are easier to identify in the presence of a contextual contrast (Sedivy, 2001). One possibility is that the different modifiers engender a contrastive inference in disparate ways. Adjectives that denote properties which are inherently relational, such as scalar terms, might engender a contrast effect as a result of their lexical semantics. Modifiers that have no special relation to a the context of their use, such as material and color terms, might instead convey contrast pragmatically. If so, then the expectation for contextual contrast given by scalar terms might be less susceptible to being blocked by the present pragmatic manipulation than that of material items.

Participants

Thirty-one members of the Brown University community were paid for their participation in the present experiment. Each was a native speaker of English and had normal or corrected-to-normal vision. All participants were naive with respect to the goals of the experiment.

Materials and Design

The methodology and design were adapted from Sedivy et al. (1999, 2001). Participants were asked to manipulate arrays of four objects according to a set of prerecorded instructions. Twenty stimulus sets like that given in Figure 1 were constructed. For each display array there were two to four instructions. Critical instructions, containing prenominally modified NPs (e.g., "Pick up the tall cup."), always occurred first in the series. In ten of the twenty critical NPs, the prenominal modifier referred to a scalar property. For the other ten, the modifier described a material property (e.g., "The glass mug."). Appendix A lists the contrast, competitor, and target objects for the stimuli. For the scalar items, the competitor object was selected so that it was a better exemplar of the modifier property than the target (e.g., an unusually tall pitcher). Thus if there is an early bias toward the literal interpretation of the adjective, it should result in elevated erroneous looks to the competitor rather than the target. This could not be done for the material items. Half of the experimental items in each session included a contrasting object, and half did not. Each participant saw only the contrast or no contrast variant of any stimulus set. An additional ten trials contained displays like those in the contrast condition, but where a modified NP was used to refer to the competitor item. These counterbalancing trials were intended to ensure that individuals would not be cued to expect a modified NP to refer to a member of the contrasting pair. Stimuli and counterbalancing trials were pseudorandomly ordered with 26 filler trials.

All participants were told that the experimental instructions were generated by "an individual who was asked to direct a listener through a sequence of object configurations" and that the experiment was designed to test how effectively individual speakers were able to convey instructions by observing a perceiver's responses. Fifteen participants were assigned to the reliable speaker condition. The remaining sixteen participants were assigned to the unreliable speaker condition. The impression of unreliability was conveyed in three ways. First, participants were told that the speaker who had recorded the instructions had an "impairment that

caused language and social problems." Second, the speaker described objects and locations erroneously. Five times over the course of the experiment an object was mislabeled. For instance, a toothbrush was called a "hairbrush." On three occasions the instructions directed the perceiver to move an object to a location that did not exist. For instance, a destination might be described as above object A and below object B, when object B was below object A. Both of these error types occurred in a minority of the nearly 200 instructions. Third, the speaker consistently used overly informative descriptions. There were 234 non pronominal references where an unmodified form would have been sufficient to indicate the object of interest. Of these, 197 contained a superfluous modifier. The remaining 37 were unmodified nominal descriptions.

Note that the presence of the modifier was not a reliable cue to the presence of a contextual contrast for the reliable speakers either. Overall, participants in the reliable speaker condition heard thirty prenominally modified descriptions: the twenty stimuli and ten counterbalancing trials. For each participant only ten of these (the number of stimuli presented in the contrast condition) were uttered in the presence of a contextual contrast for the target. Although the reliable speaker generates far fewer modified forms than the unreliable speaker (30 versus 207), there was no contingency between modification and contextual contrast over the course of the experiment for either speaker.

In order to avoid placing prominence on the prenominal modifier, nuclear stress was placed on the head noun of the NP in critical instructions. Durations of the adjective and noun are given in Table 1. Note that critical regions were comparable across the unreliable and reliable speaker conditions.

Table 1: Duration of modifier and noun in target instructions across conditions in ms (Standard Errors in parentheses)

	Scalar		Material	
	Unreliable	Reliable	Unreliable	Reliable
adj	288 (18)	289 (8)	351 (42)	372 (39)
noun	378 (37)	373 (33)	332 (36)	355 (42)

Procedure

Display changes took approximately 5 seconds, and participants were permitted to watch the display as it was being changed. Every display contained a centrally located fixation cross. Each trial began with a request for the subject to look at the cross, and participants were instructed to rest their eyes on the central cross between instructions. This was done so that eye movements to the target objects could be measured from a default position that was equidistant to all of the objects in the display. Participants were told to follow the instructions as quickly and accurately as possible.

While the participant followed instructions to move objects in the workspace, eye movement data were recorded using a lightweight ISCAN head-mounted video-based tracking system. The camera provided an infrared image of the eye, and determined monocular eye position by monitoring the locations of the center of the pupil and the cornea reflection. A scene camera was mounted on the side of the helmet, providing an image of the subject's field of view. Calibration was carefully monitored throughout each trial, and minor adjustments were occasionally made between trials. A VCR record with a time code stamp was made at 30 Hz for each experimental trial, consisting of the instructions recorded via microphone, as well as the participant's moment-by-moment gaze fixation superimposed over the scene camera image. Because the scene camera was mounted onto the helmet itself, and moved with the participant's head, the VCR record took into account any head movements made by the participant, allowing for unrestricted head and body movements throughout the experiment. The entire experiment, including introduction to the equipment and task, practice and experimental trials, and debriefing took approximately 35 minutes.

Eye movements were analyzed by playing the audio-video record back for each 33ms frame on a SONY DSR-30 digital VCR. For the experimental trials, critical points in the speech stream were identified, corresponding to the onsets of the adjective and head noun, as well as the offset of the head noun. Continuous eye movements occurring from the beginning of the instruction were noted until the participant reached for the target object. Thus, the joint identification of critical points in the speech stream and the eye movement data allowed for the alignment of eye movements to speech, as presented in the Results section below. For purposes of analysis, the work surface was divided into a 3X3 grid, with visibly demarcated boundaries. Eye movements to an object were coded from the first frame at which a saccade was launched to the square containing that object. Occasionally, poor calibration or eye blinks resulted in a temporary loss of track. In the event that the tracking was restored less than 5 video frames later, and reappeared

on the same location, the track was treated as continuous; otherwise, the eye movement record noted the loss of track, and for that time period, treated the fixation as falling on none of the objects in the display.

Results

Trials where individuals reached for the incorrect object or where the participant was fixated on the target at the beginning of the adjective were omitted from analyses. This latter condition was intended to exclude fixations on the target that were not initiated on the basis of speech input. Further, data points more than 2SD away from the mean for each condition and time frame were replaced with the mean for that condition. This affected 3.2% of the data.

Figure 2 depicts the proportion of trials including fixations to each of the objects in the display over time from the onset of the adjective. To correct for variability in the auditory duration of the modifier across stimuli, each trial was aligned to the offset of the adjective. The average offsets for the noun and adjective in each condition are indicated on the graph. Critical comparisons were conducted over the 500 ms window beginning 200 ms after adjective offset. This corresponds to where manipulations of discourse contrast have been observed in previous work.

Analyses were performed for each modifier type separately. For scalar items, target advantage scores were computed by subtracting fixations to the competitor from fixations to the target over the critical interval. This provided a composite measure of the relative proportion of fixations to the entities that should be affected by the presence of a referential contrast. Figure 3 depicts target advantage scores for the scalar items. A 2 X 2 ANOVA crossing speaker type and the presence or absence of a contrasting object in the display resulted in an interaction significant in the participants analysis ($F1(1,29) = 5.05$, $MSE=.043$, $p < .05$), but not in the items analysis ($F2(1,9) = 2.24$, $MSE = .032$, $p = .17$). Planned comparisons for each type of speaker revealed that perceivers responding to the reliable speaker benefited from the presence of a contextual contrast ($F1(1,14) = 17.98$, $MSE = .024$, $p < .001$; $F2(1,9) = 4.6$, $MSE = .041$, $p < .05$) but those responding to the unreliable speaker did not ($Fs < 1$). To establish the relative contributions to this pattern of looks to the target and competitor, additional analyses were conducted for fixations to each of these objects separately. Analysis of the proportion of fixations to the target patterned similarly to target advantage scores. There was a trend toward and interaction of speaker and contrast marginal by participants ($F(1,29) = 3.36$, $MSE = .01$, $p = .08$), though not by items

(F2(1,9) = 1.41, MSE = .01, p = .26). Independent comparisons established that reliable speakers elicited significantly more looks to the target in the presence of a contrast (F1(1,14) = 11.9, MSE = .007, p < .01; F2(1,9) = 6.4, MSE = .01, p < .05) and unreliable speakers did not (Fs < 1). Proportions of fixations to the competitor also appear to have contributed to the interaction of target advantage scores. There was a marginal interaction by participants (F1(1,29) = 4.07, MSE = .019, p = .05), but not by items (F2(1,9) = 1.44, MSE = .021, p = .26). For the reliable speaker, there were fewer spurious looks to the competitor in the presence of a contrast (F1(1,14) = 15.2, MSE = .008, p < .01; F2(1,9) = 2.2, MSE = .026, p = .09). The manipulation of contrast did not affect fixations to the competitor for the unreliable speaker (Fs <.03).

Figure 4 portrays looks to objects in the display in response to instructions containing material modifiers. Target advantage scores were computed and submitted to a 2 X 2 ANOVA crossing speaker reliability and contextual contrast. This revealed a significant interaction by items (F2(1,9) = 9.03, MSE = .011, p < .05), but not by participants (F1(1,29) = 1.92, MSE = .048, p = .18). Planned comparisons demonstrated a marginally reliable trend for higher target advantage scores in the presence of a contrast for reliable speakers (F1(1,14) = 2.63, MSE = .036, p = .06; F2(1,9) = 2.3, MSE = .024, p = .08). In contrast, target advantage scores were numerically lower in the unreliable speaker condition, though this trend was not reliable (Fs < 2.8). Just as for the scalar conditions, looks to the target and competitor were analyzed separately. There were no clear effects or interactions for fixations to the target (Fs < .5). The absence of these effects and the relatively high target advantage scores likely reflect the tendency for individuals to identify the target extremely rapidly for the material modifiers. Looks to the competitor therefore provide a more sensitive indicator of the effect of contrast. Speaker type and contrast interacted reliably (F1(1,29) = 5.6, MSE = .011, p < .05; F2(1,9) = 18.4, MSE = .003, p < .01). In response to reliable speakers, there were fewer looks to the competitor when a contrasting object was in the display (F(1,14) = 7.64, MSE = .013, = p < .05; F2(1,9) = 18.1, MSE = .004, p < .01). This was not the case for individuals in the unreliable speaker condition (Fs < .75).[3]

[3] Inspection of Figure 4 suggests that there were frequent looks to the second distracter (labeled as "contrast") prior to adjective offset in the unreliable speaker no contrast condition. Though proportion of fixations appears to be elevated relative to looks to the other distracter, this difference is not significant (Fs < 1.7). There are also no such elevated fixations for the reliable speaker no contrast condition. Further the elevation occurs extremely early (200 ms prior to adjective offset) which makes it unlikely that it was affected by the critical description.

It was possible that perceivers in the unreliable speaker condition might have delayed interpretive processes in light of the irregular instructions. In order to ensure that this was not the case analyses were performed to establish when combined fixations to the competitor and target, each of which match the modifier property, diverged from looks to other objects in the display. In response to the reliable speaker, participant fixations isolated the target and competitor with marginal reliability between 67 and 100 ms after adjective offset ($F(1,14) = 2.36$, $MSE = .009$ $p = .07$). For instructions containing scalar modifiers, this divergence first occurred in the window between 200 and 233 ms after adjective onset ($F(1,14) = 2.43$, $MSE = .009$, $p = .07$). For material modifiers, the difference was first observed between 33 and 66 ms ($F(1,14) = 2.04$, $MSE = .011$, $p = .09$). Adjective-linked eye movements in response to the unreliable speaker were initially observed between 67 and 100 ms over all stimulus items ($F(1,15) = 2.87$, $MSE = .005$, $p = .06$), between 100 and 133 ms for scalar conditions ($F(1,15) = 2.04$, $MSE = .012$, $p = .09$), and between 67 and 100 ms for the material conditions ($F(1,15) = 3.53$, $MSE = .033$, $p < .05$). For every comparison given here, fixations to the target and competitor were significantly higher than to other objects in the display at each subsequent 33 ms analysis frame ($Fs > 3.5$, $ps < .05$). Given estimates that programming and launching a saccade takes approximately 200 ms (e.g., Matin, Shao, & Boff, 1993), these comparisons indicate that participants in both conditions are using the speech input to incrementally fix reference prior to the offset of the modifier. Importantly, participants in the unreliable speaker condition were at least as rapid to respond to the literal denotation of the modifier (and, in the case of the scalar adjectives, perhaps quicker) as those in the reliable speaker condition. The unnaturalness of the speaker did not cause interpretation to be any less incremental. Note too that eye movements were not retarded by the slightly faster material adjectives in the unreliable speaker condition.

Discussion

The results demonstrate that speaker-specific attributes influence whether a restrictive modifier will be interpreted contrastively. In line with previous work, individuals in the reliable speaker condition were aided in determining the referent for a restrictively modified nominal description by the presence of a contextual contrast. With scalar modification, there were more early looks to the target, and fewer looks to a competing object that matched the scalar property. For the unreliable speaker, neither effect was observed. When the reliable speaker uttered NPs containing material modifiers, fewer looks were elicited to the competing object in the presence of a contrasting object. This too was not observed with the unreliable speaker. Thus, individuals

were not aided by a contextual contrast when they had reason to believe a speaker did not use modification cooperatively. This is in accord with the view that participants in the unreliable speaker condition did not take the presence of a modifier to imply the existence of a contrast set. Intriguingly, these participants were just as responsive to the literal meaning of the adjective as for the reliable speaker. The speaker manipulation selectively eliminated the generation of the contrastive implicature licensed by the adjective.

The fact that contrastive inferences can be eliminated by manipulating speaker reliability provides strong support for a pragmatic interpretation of referential contrast effects, and furthermore implies that early inferencing admits episodic information. The speed of contrastive inferences therefore cannot be explained as an automatic reflex of deviating from an immutable default form. This is at odds with the most straightforward reading of the short-circuited implicature proposal introduced above. However, this result does not imply that individuals are engaging in an overt counterfactual deductive reasoning process in order to use the contextual contrast appropriately. We return to this point in the general discussion.

Each modifier condition was similarly affected by speaker-reliability and contextual contrast. The presence of a contrast improved target identification for both the scalar and material conditions in response to reliable speakers. In both cases, the effect of referential contrast vanished with unreliable speakers. Thus the contrastive inference is pragmatic in origin for relational (scalar) and non-relational (material) modifiers alike. This rules out the possibility that scalar terms convey the expectation for contextual contrast as part of their literal meaning. If they did, then the contrast effect should not have been cancelled by manipulating the pragmatic context of the modifier's use. This is not to say that scalar and material adjectives were interpreted identically. Intriguingly, individuals were somewhat slower to map scalar meanings to the subset of items which matched the modifier denotation. It is possible that the relational component of scalar meaning complicates the task of converging on a target. Parallel findings from a recent production study conducted in our laboratory found that more disfluencies occur prior to uttering prenominal scalar modifiers than material or color modifiers (Gregory, Grodner, Joshi, & Sedivy, 2003). This is particularly striking in light of the fact that scalar terms are more frequent in both spoken and written corpora. Taken together, these observations buttress the thesis that scalar denotations are more conceptually complex than non-comparative adjective denotations. Investigating how the lexical semantics of various modifiers mediates reference resolution is an interesting direction for future research.

General Discussion

The results reported herein strengthen the case for a pragmatic explanation of referential contrast effects by demonstrating their defeasibility. In addition, they indicate that there is a limit to the generality of quantity-based inferences and that characteristics particular to a speaker are taken into account in the generation of referential contrast inferences. A number of open questions remain. For example, the results reviewed above do not address how perceiver expectations are updated or represented in a way that they can influence early inferencing. It is intriguing to speculate which cue to unreliability, or combination of cues, caused the attenuation of the contrast effect. One possibility is that perceivers' inferences hinge directly on their overt beliefs about the degree to which the speaker is conforming to principles of rational, orderly communication. Hence, an awareness of the mistakes made by the speaker and the explicit identification of the speaker as non-normal defuse the contrastive inference. Perceivers may believe that the speaker's impairment causes them to be uncooperative or unreliable communicative partners and thus suspend any inferences made on that basis.

A second possibility is that pragmatic inferencing reflects a more implicit assessment of the communicative proclivities of the speaker. For instance, perceivers may attend to the statistics of modifier use for a particular speaker. Over the course of the experiment they would note that the presence of the adjective is not a reliable cue to contrast. As a result of the overuse of the modifier, they might recalibrate the anticipated default description to include a modifier. This view predicts that the effect of contrast might get weaker with repeated exposure to overly descriptive NPs (though it is unclear how much local experience might be necessary to override the prepotent contingency between modification and contextual modification). Anecdotal evidence from post-experimental debriefing suggests that perceivers do not register conscious awareness of the extent to which speakers exhibit optimally informative communicative behavior. When queried whether they noticed anything unusual about the experiment, participants frequently mentioned trials where objects were mislabeled or where object destinations made reference to impossible configurations. They rarely if ever mentioned the overly explicit object labels. These observations hint that the speaker's overspecified descriptions did not have a large impact on the overt assessment of the reliability of the speaker. Still, the speaker-specific modulations of the contrast effect could be attributable to high-level beliefs about speaker characteristics, lower-level statistical properties of the speakers output, or some combination of

both. If only high-level cues are needed to indicate speaker unreliability, then perceivers need not learn about the particular way in which the present speaker is uninformative, and should suspend the computation of a broad range of implicit meanings. Further there should be no evidence of greater reduction in the contrast effect as the perceiver accumulates evidence of over-informativity. On the other hand, sensitivity to lower-level cues should result in incremental changes to how perceivers respond to the descriptive patterns used by a speaker as more tokens of speaker descriptions are encountered.

To evaluate these alternatives, exploratory analyses were conducted comparing performance on items in the first and the second halves of the experiment. The hypotheses differ with respect to their predictions for the reliable speaker conditions. To see why this is so, note that reliable speakers uttered modified forms to refer to competitors as often as they did to pick out a member of a contrasting pair (For ten items across the 56 experimental trials in each case). Hence, there was no reliable contingency between modification and contextual contrast for the reliable speaker condition. Thus, over the course of the experiment, perceivers may have come to adjust their expectations about the information conveyed by modification as a result of encountering a significant number of modified forms in the absence of contrast. The unreliable speaker on the other hand, consistently used overly specific forms on a much large number of trials, with 15 modified forms occurring in the absence of contrast even before the first experimental item. Therefore, statistics corresponding to the overuse of the modifier may have already been adjusted on the expected referential form by the point at which inferential processes could be assessed in the first half of the experiment.

The most interesting measure for the reliable speaker was looks to the competitor for scalar items depicted in (Figure 6). The interaction of contrast and block order was marginal ($F(1,13)=3.61$, $MSE=.021$, $p = .08$). Consistent with the statistical tuning hypothesis, looks to the competitor were reduced in the presence of a contrast in the first block of the experiment ($F(1,13)=7.41$, $MSE = .031$, $p <.05$), but not in the second. ($F < 1$). There were no effects of contrast or interactions with block order for the unreliable speaker ($Fs < 1.5$). Two caveats are in order: The present experiment was not specifically designed to test these hypotheses and no other measure differed reliably across blocks. Still this trend is suggestive that overuse of the modifier contributes to the elimination of the contrast effect.

A follow up study delves further into the source of the pragmatic effect observed here. A reliable speaker condition was compared to two unreliable speaker-conditions. In one of these, unreliable conditions, the speaker overused prenominal modifiers consistently just as in the present experiment. In the other, the speaker encoded the same redundant content as the prenominal condition, but did so with a postnominal modifier (e.g., "Pick up the cup that's tall.). Like the previous experiment, critical trials tested instructions containing prenominal modifiers for all three conditions. But unlike the first experiment, none of the explicit cues to speaker irregularity were given. If perceivers are sensitive to the redundancy of content, then both unreliable speaker-types should result in a reduction of the contrast effect. If perceivers are sensitive especially to a redundancy of a particular form, then the excessive use of the prenominal modifier should result in a greater reduction of the contrast effect than the postnominal condition. It is also possible that high-level knowledge of the speaker's impairment would be necessary to draw attention to low-level redundancies in the present study. If so, then there might not be a marked reduction in the effect of contrast.

Figure 7 depicts target advantage scores for instructions containing scalar terms over the same temporal region analyzed in the present experiment. For all three conditions there was an apparent benefit of the presence of a contrast in identifying the target. This was significant by participants and items for the reliable speaker ($F1(1,18) = 5.16$, $MSE = .05$, $p < .05$; $F2(1,9) = 5.48$, $MSE = .02$, $p < .05$), and by participants for the condition with excessive prenominal modification ($F1(1,18) = 3.87$, $MSE = .03$, $p < .05$; $F2(1,9) = 1.28$, $MSE = .03$, $p = .14$). While a similar trend was observable for the condition with excessive postnominal modification, it was marginally reliable only in the items analysis ($F1(1,16) = 1.59$, $MSE = .09$, $p = .11$; $F2(1,9) = 2.44$, $MSE = .04$, $p = .08$). There was no hint of an interaction between speaker-type and contrast ($Fs < .5$). This work demonstrates that overusing restrictive modification is not sufficient by itself to eradicate the contrastive interpretation of scalar modifiers. This does not immediately imply that the high level cues to speaker reliability were solely responsible for defusing the expectation of contrast in the earlier experiment. It is possible that high-level cues were necessary to draw the listener's attention to the low-level redundancy. It is also possible that the pattern observed in the follow-up experiment was unique to scalar modifiers. On this view excessive modification might have been sufficient to block the contrastive interpretation of material modifiers, which are not inherently comparative.

Another open question is what kind of pragmatic mechanism could be responsible for the above findings. A Gricean explanation for contrastive interpretations of restrictive modifiers claims that perceivers make the following inferential steps when a speaker utters a modified NP:

(i) If the speaker means to use this utterance to pick out an intended referent in a context with only one entity of that type, then a default description would be the most natural means to do so.
(ii) The speaker chose a more specific description than the default by using a modified NP.
(iii) If the speaker is behaving cooperatively, they should not be more informative than is necessary. Overspecification must have some purpose.
(iv) Because of (ii) and (iii) the conditions for uttering a default form in (i) must not hold. Namely, it must not be the case that perceiver intends the utterance of the modified form to pick out an entity in isolation.

One possible conclusion of (i-iv) is (v)

(v) There are multiple entities in the context of the same type.

There are a number of challenges inherent to implementing this sort of reasoning in a real time processing system. For one, it is unclear what kind of reasoning system would permit arriving at (v) as rapidly as perceivers do. That is, given step (iii), what ranges of "purposes" for the inclusion of modification are considered, and what weight is each of them given? It is certainly not the case that distinguishing between like entities is the only, or even the most frequent function of modification (e.g. see Fox & Thompson (1990) for a taxonomy of functions of relative clause modifiers). Second, a critical step is embedded in (ii). How does the system decide the appropriate alternative to compute and compare to a given referential expression? For instance, "the plastic cup" is a more specific description than "the plastic entity," yet intuition tells us that the contrastive entity invoked is another cup and not another plastic object. The comparison of alternative forms is made even more impressive when we consider that it is occurring incrementally, even before the noun is encountered. A partial answer to both of these questions might be that statistical regularities constrain the consideration of alternative expressions and their functions, and indeed, that inferential steps may be statistically linked. This would permit the processor to bypass the complexity inherent in counterfactually reasoning about alternatives.

Let us sketch one way this approach might work to account for the currently known facts about contrastive inferences. First, determining the referential forms to be compared could be a function of patterns of co-occurrence among properties in referential descriptions. Modifiers are more promiscuous than nouns. "Plastic" will be used to modify a wide variety of artifact labels, whereas "cup" will be used to describe a narrow set of object referents. The most accessible default for "the plastic cup" is "the cup" rather than "the plastic thing" because the descriptions in which "cup" participates form a more coherent set of properties than the descriptions in which "plastic" participates. This would explain our intuitions about the dimension along which a contrast set differs from the referent set. Whether or not an expression that is actually used deviates from the expected default could be computed based on the likelihood of using a modifier encoding some specific property together with a particular noun. For instance, the ratio of expressions encoding yellowness as a property given the total occurrences of a noun such as "banana" would presumably be lower than the ratio of encoding "yellow" given "notebook". Thus, "yellow banana" would represent more of a deviation from the expected default than "yellow notebook", triggering a search for some function to the modifier. The preferred function (i.e. as referentially contrasting) could be arrived at via a statistical link between steps (ii) and (v) (though note that this statistical link remains to be empirically established, given the observations in Fox & Thompson). Clearly this is an extremely partial sketch of how the aforementioned challenges might be met. Our present purpose is merely to point to potential directions for realizing a reasoning system that is flexible enough to consider situation based information, but sufficiently constrained to operate in real time.

There are at least two points in the process outlined above where speaker-specific effects could have had their influence in the present experiment. One is to directly recalibrate the default form at step (ii) for a particular speaker based on statistical regularities in the recent episodic record. An alternative that seems more computationally cumbersome is to generate expectations of defaults more generally in step (ii) as a function of global experience across many speakers, and then to invoke the criterion in (iii) as a prerequisite to identifying the most likely function for the modifier. In the former case, speaker-particular regularities for the speakers in the experiment reported above might be used to calibrate expectations about default expressions. That is, modified NPs would count as deviations from the expected default for reliable speakers, but not for unreliable speakers. This type of rapid calibration to characteristics of a speaker's speech could be similar to the low-level acoustic calibration that is automatically achieved to take into account speaker characteristics such as age, gender, etc., allowing perceiver's to cope efficiently

with speaker variability. The second alternative posits that the use of a modified NP would count as a deviation from the default in both speaker conditions; however, step (iii) would then be invoked, and the determination of a speaker's unreliability with respect to communicative norms would suspend a search for an appropriate function for the modifier. This latter explanation is more computationally complex, and is not able to exploit direct statistical links between steps (ii) and (v).

The results of our follow-up experiment represent an attempt to distinguish between these two explanations. Thus far, the results are suggestive that a determination of speaker reliability (i.e. step (iii)) may not be entirely dispensable in accounting for contrast set inferences. Further research is needed to determine whether overt signaling of speaker unreliability or uncooperativeness is both necessary and sufficient to suspend contrast inferences, as well as what sorts of evidence can be used to make this determination.

Another way of exploring the explanatory power of statistical regularities in generating pragmatic inferences is to examine whether perceivers' generation of a contrastive implicature depends upon the identification of plausible alternative functions of modifiers. To test this theory, it would be interesting to see if a pragmatic manipulation that did not alter the cue validity of a modifier could defuse the contrastive implicature. Consider the sentence in (7) uttered within a discourse where a woman is established who is trying and failing to reach a cup on a shelf.

(7) The short woman could not reach the cup on the top shelf.

The modifier provides a causal explanation for the a prominent event. Intuitively no contrasting woman is conjured by this example. It is possible that the prepotent identifying function of the modifier initially results in a contrastive interpretation, which is later retracted on the basis of secondary deliberative processes. Alternatively, one might see no evidence of contrast inferences in initial referential commitments, suggesting that the identification of alternative functions comes into play prior to the conclusion reached in step (v). This could be implemented either at step (ii), in allowing relevance-based considerations to constrain expected default expressions, or at step (iii), once a deviation from the expected expressions triggers a search for plausible functions for that deviation. Careful temporally-sensitive experimentation along these lines has the potential to clarify key issues in the computation of pragmatic inferences that have hitherto been addressed primarily by theoretical linguists. Representational distinctions have been posited between encoded content implicit content, and between generalized and particularized implicature. These formal distinctions provide useful starting points for formulating processing

hypotheses. Further investigation along these lines will allow us to gain ground in understanding which parts of the inferential process are generated automatically, and which parts computed ad hoc.

Acknowledgements

Appendix A

Table A1: Experimental items employed in present study

Modifier Type	Target	Contrast	Competitor	Distracter 1	Distracter 2
scalar	small pad	large pad	small doll	razor	shampoo
scalar	thin marker	thick marker	thin brush	potato masher	crayons
scalar	thick notebook	thin notebook	thick book	rag	horseshoe
scalar	tall cup	short cup	tall pitcher	eraser	diskette
scalar	narrow post-its	wide post-its	narrow ribbon	peanut butter	pink bow
scalar	long envelope	short envelope	long spatula	toy shovel	lotion
scalar	tall doll	short doll	tall mug	black pen	yellow folder
scalar	long spoon	short spoon	short pencil	peach	quarter
scalar	wide tape	narrow tape	wide belt	banana	red pencil
scalar	fat crayon	thin crayon	fat marker	egg	tupperware
material	brass frame	wood frame	brass candle holder	fork	thread
material	leather glove	wool glove	leather wallet	blue bow	tape measure
material	paper plate	plastic plate	paper bag	salt shaker	shot glass
material	porcelain bowl	plastic bowl	porcelain saucer	candle	sunglasses
material	plastic spoon	metal spoon	plastic comb	tie	battery
material	styrofoam ball	rubber ball	styrofoam cup	comb	lego
material	metal ladle	plastic ladle	metal pan	orange	toy octopus
material	wool sock	cotton sock	wool cap	duct tape	plastic ruler
material	wood cutting board	plastic cutting board	wood ruler	ribbon spool	mirror
material	glass mug	ceramic mug	glass vase	pencil	floss

References

Altmann, Garnham, & Dennis (1992). Avoiding the garden-path: Eye movements in context. *Journal of Memory & Language*, *31*, 685-712

Altmann & Steedman (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191-238.

Barr, D. J. (2003) Listeners are Mentally Contaminated. Poster presented at the 44th annual meeting of the Psychonomic Society, Vancouver, B.C.

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language, 46* , 391-418.

Bierwisch, M. (1987). The Semantics of Gradation . In: M. Bierwisch , E. Lang (eds.): *Dimensional Adjectives*. Berlin et al.: SpringerVerlag. 71-261

Bradlow, A. R. (2002). Confluent talker- and listener-related forces in clear speech production. In Gussenhoven, C. & Warner, N. (Eds.) Laboratory Phonology 7. Berlin & New York: Mouton de Gruyter. Pp. 241-273.

Carston, Robyn (1998). Pragmatics and the Explicit-Implicit Distinction. University College London PhD thesis. To be published by Blackwell, Oxford

Chambers, C., Tanenhaus, M., & Magnuson, J (In Press). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*

Clifton, Bock, & Rado, (2000). Effects of the focus particle 'only' and intrinsic contrast on comprehension of reduced relative clauses. Kennedy, Radach, Heller, Pynte (eds.) *Reading as a Perceptual Process*. Amsterdam: Elsevier Press.

Clifton, C., & Ferreira, F. (1989). Ambiguity in context. *Language and Cognitive Processes, 4 (SI)*, 77-103.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunnen, & A. Zwicky Eds.). *Natural Language Parsing*. Cambridge, UK: Cambridge University Press.

Fox & Thompson (1990) A discourse explanation of the grammar of relative clauses in English conversation. *Language*, *66,* 2.

Gregory, M., Grodner, D., Joshi, A. & Sedivy, J. (2003) Adjectives and processing effort in production: So, uh, what are we doing during disfluencies? Paper presented at the 16 th annual CUNY conference on human sentence processing. Boston, MA.

Grice, H.P. 1975. Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics*. New York: Academic Press.

Hanna, J.E., Tanenhaus, M.K. & Trueswell, J.C. (2003). The effects of common ground and perspective on domains of referential interpretation.. *Journal of Memory and Language, 4*9, 43-61.

Isaacs, E.A., & H. H. Clark. (1987). References in conversations between experts and novices. *Journal of Experimental Psychology: General,* 116, 26-37.

Jurafsky, Daniel. 2002 (to appear) Pragmatics and Computational Linguistics. To appear in Laurence R. Horn & Gregory Ward (eds.) *Handbook of Pragmatics* . Oxford: Blackwell.

Levinson, S. (2000) *Presumptive Meanings*. Cambridge, MA: MIT Press.

Metzing, C. & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects in the comprehension of referring expressions . *Journal of Memory and Language, 49* , 201-213.

Nadig, A.S. & Sedivy, J.C. (2002) Evidence of perspective taking constraints in children's on-line reference resolution. *Psychological Science, 13*, 329-336

Ni & Crain (1989) How to resolve structural ambiguities. Proceedings from the 20th North East Linguistic Society Conference.

Ni, W., Crain, S. & Shankweiler, D. (1996a). Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes, 11* (3), 283-334

Rooth, M. (1985) Association with Focus. Ph.D. Dissertation, University of Massachusetts at Amherst. Amherst: GLSA.

Sadock, J.(1978) On Testing for Conversational Implicature" In Cole, P., ed. *Syntax and semantics. IX. Pragmatics* . NY: Academic Press.

Schober, M.F. (1998). Different kinds of conversational perspective-taking. In S.R. Fussell & R.J. Kreuz (eds.), *Social and cognitive psychological approaches to interpersonal communication* (pp. 145-174). Mahwah, NJ: Lawrence Erlbaum.

Steedman, M., & Altmann, G. (1989). Ambiguity in context: a reply. *Language and Cognitive Processes, 4 (SI)*, 211-234.

Sedivy, J. C. (In press). Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in on-line language interpretation. To appear in J. Trueswell & M. Tanenhaus (Eds.). *World Situated Language Use: Psycholinguistic, Linguistic and Computational Perspectives on Bridging the Product and Action Traditions.* MIT Press.

Sedivy, J. C. (2001) Evidence of Gricean expectations in on-line referential processing. Paper presented at 14th CUNY Conference on sentence processing.
Sedivy, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language, 46, 2,* 341-370Sedivy 2002 CUNY
Sedivy, J.C. (2003) Pragmatic Versus Form-Based Accounts of Referential Contrast: Evidence for Effects of Informativity Expectations. *Journal of Psycholinguistic Research*, 32 (1): 3-23.
Sedivy, J. C., Chambers, C., Tanenhaus, M., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 109-147.
Siegel, M. (1980). *Capturing the Adjective*. New York: Garland.
Sperber, D. & Wilson, D. (1986) *Relevance*. Cambridge, MA: Harvard University Press.
Spivey-Knowlton, M., & Tanenhaus, M. (1994). Referential context and syntactic ambiguity resolution. In C. Clifton, L. Frazier & K. Rayner (Eds.). *Perspectives on Sentence Processing*. Hillsdale, N.J.: Erlbaum.
Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632-1634.
Tanenhaus, M. (2003) Referential Domains in Language Processing. Paper given at the 2003 Architectures and Mechanisms of Language Processing conference. Glasgow, U.K.

List of Figures

Figure 1

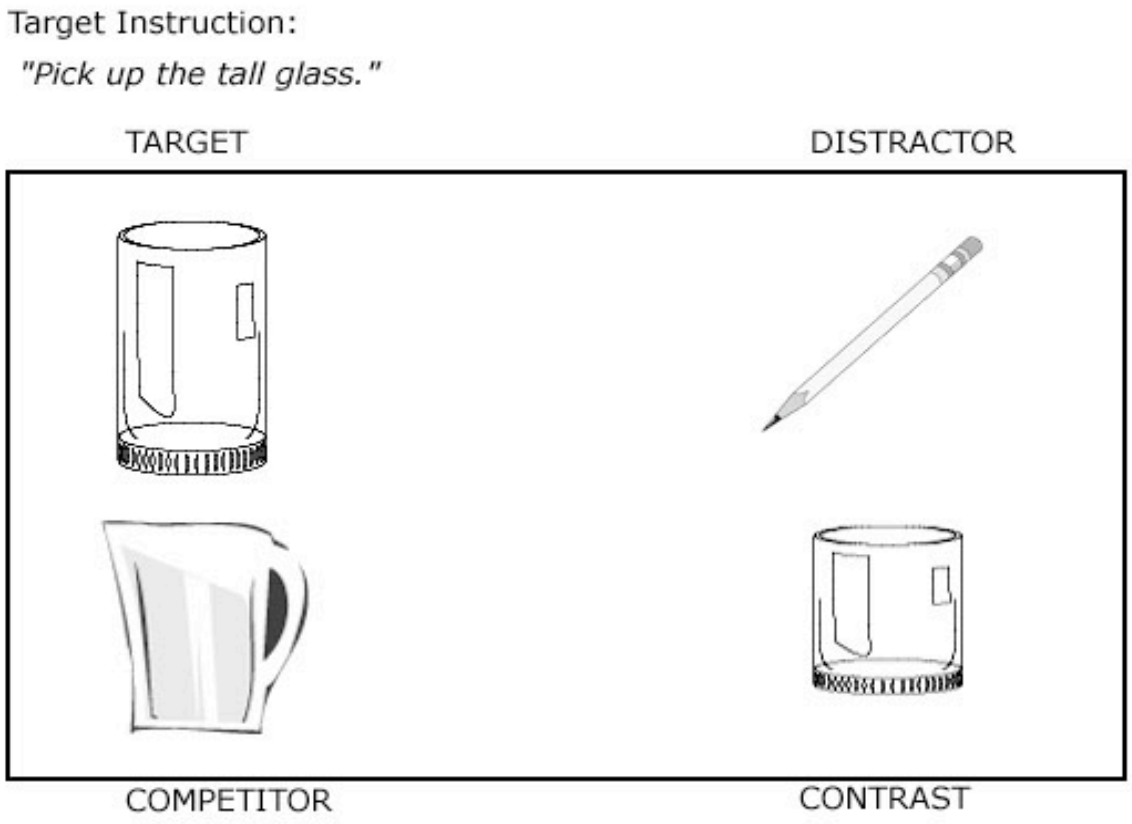

Target Instruction:
"Pick up the tall glass."
TARGET
DISTRACTOR
COMPETITOR
CONTRAST

Figure 2

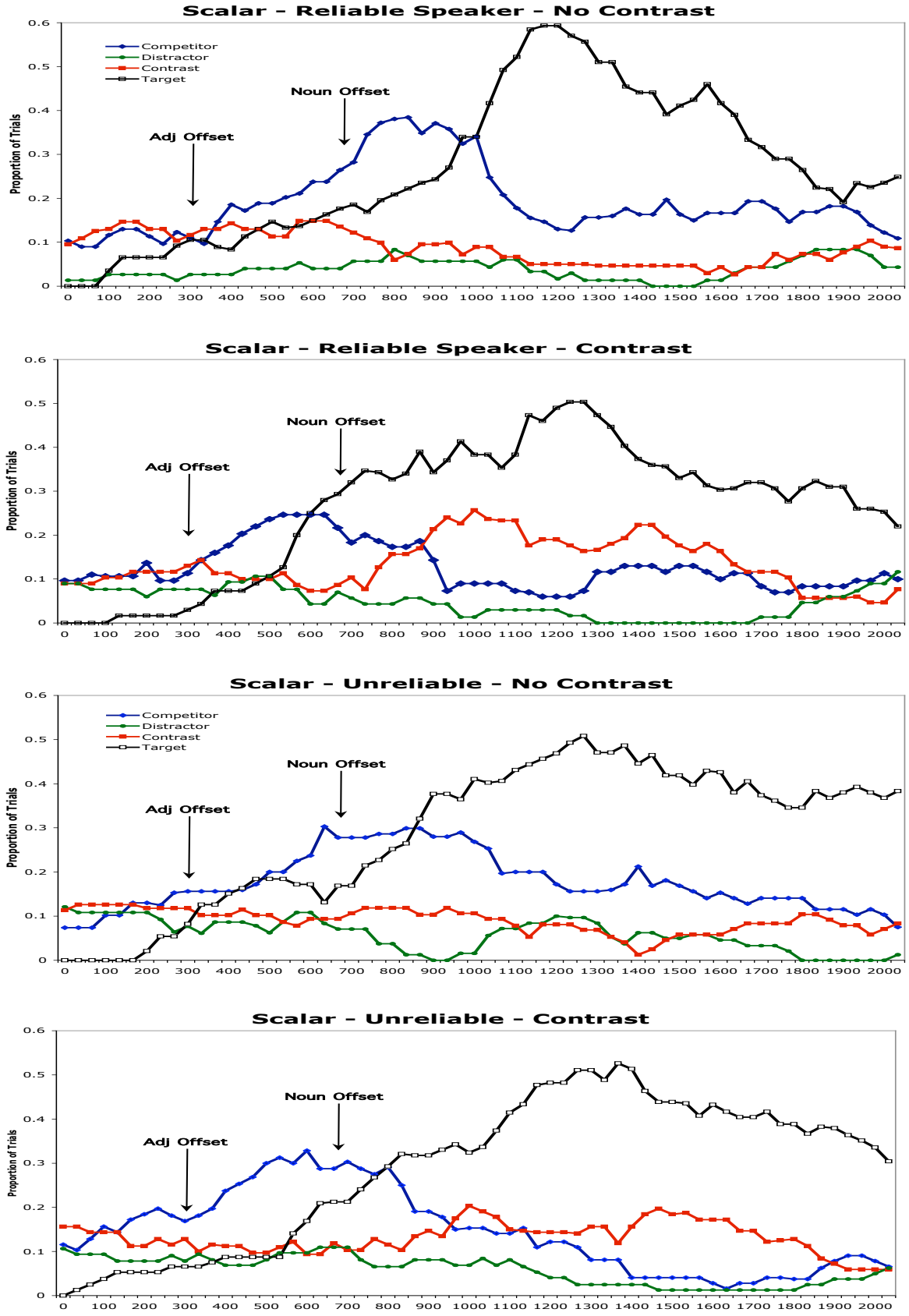

Scalar – Reliable Speaker – No Contrast
Competitor
Distractor
Contrast
Target
Noun Offset
Adj Offset
Proportion of Trials
Scalar – Reliable Speaker – Contrast
Noun Offset
Adj Offset
Proportion of Trials
Scalar – Unreliable – No Contrast
Competitor
Distractor
Contrast
Target
Noun Offset
Adj Offset
Proportion of Trials
Scalar – Unreliable – Contrast
Noun Offset
Adj Offset
Proportion of Trials

Figure 3

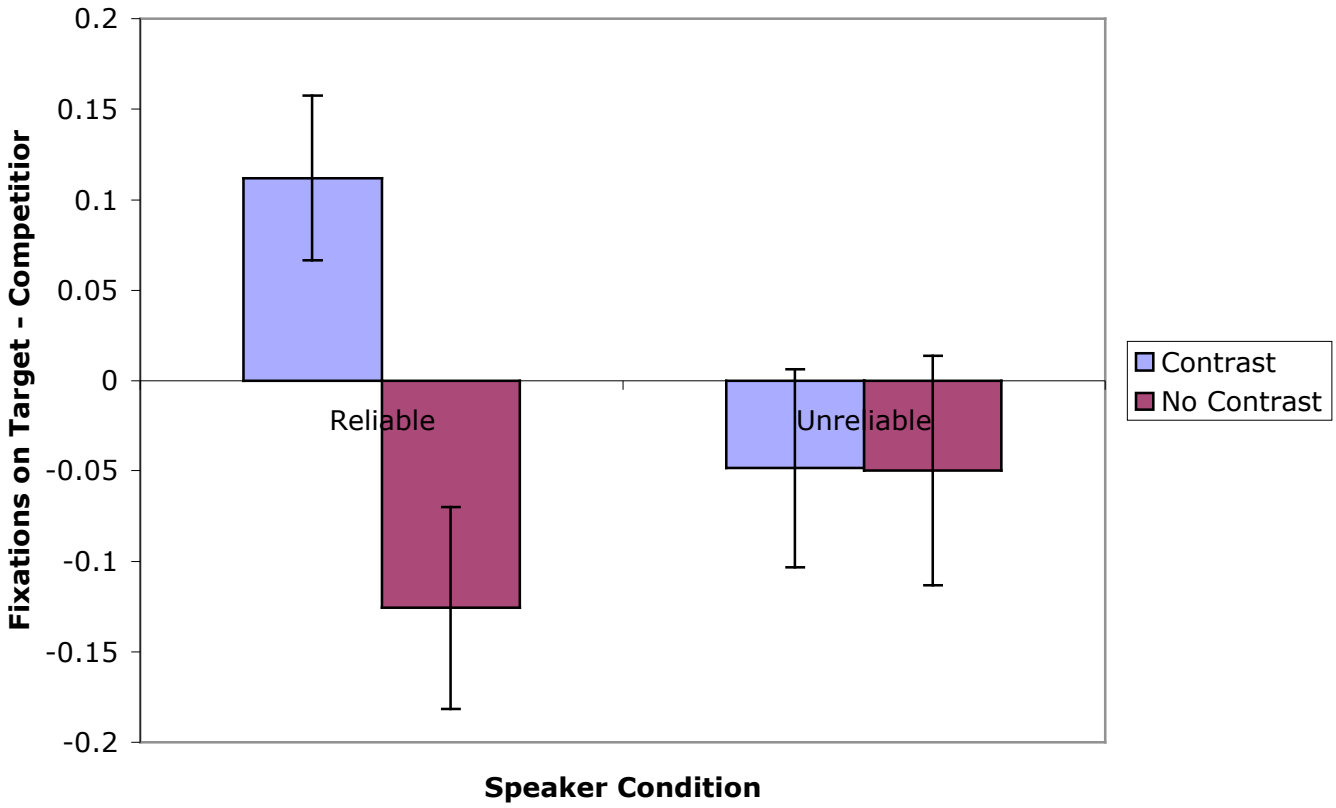

0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
Fixations on Target - Competitior
Reliable
Unreliable
Speaker Condition
Contrast
No Contrast

Figure 4

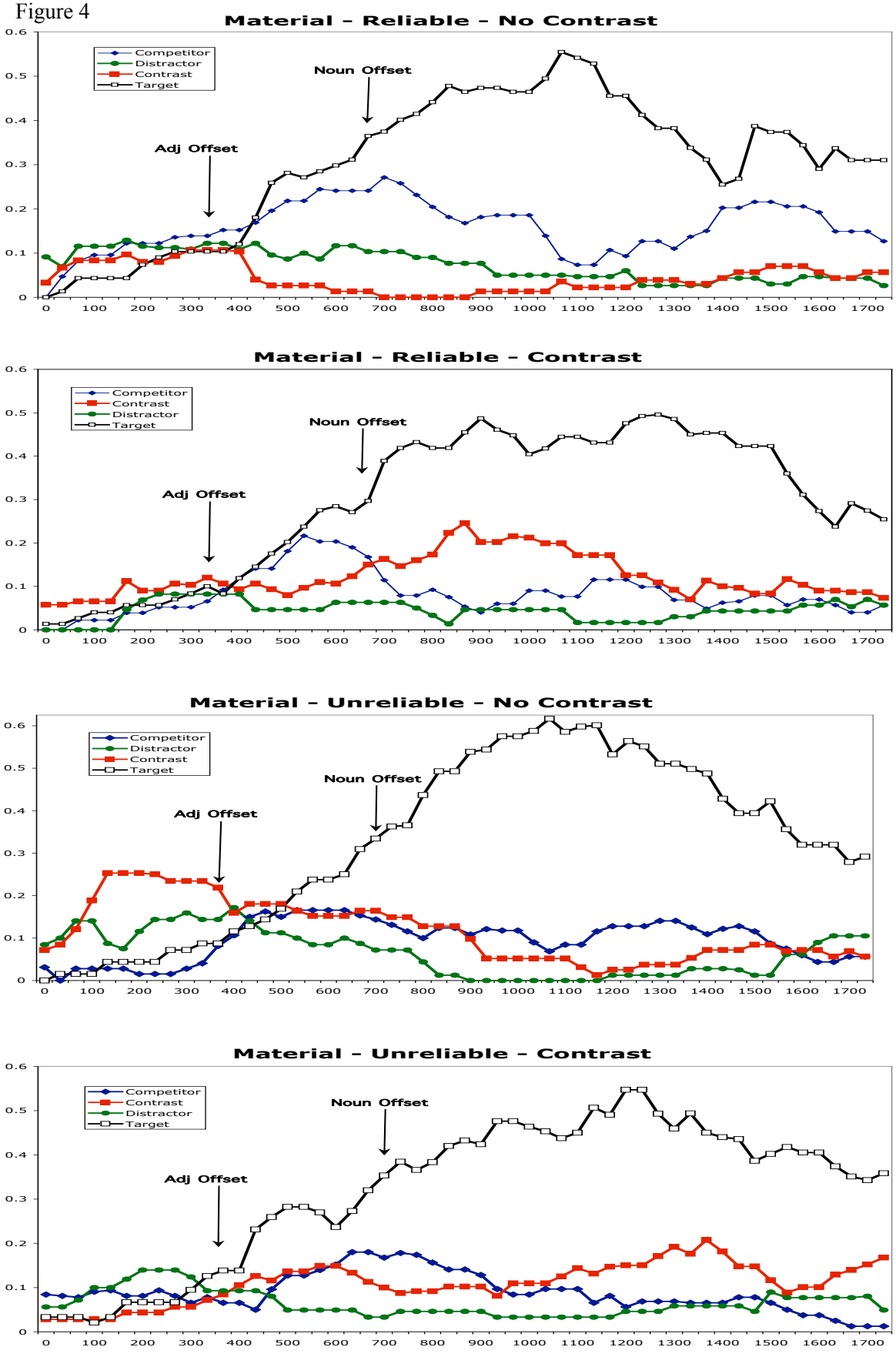

Material – Reliable – No Contrast
Competitor
Distractor
Contrast
Target
Noun Offset
Adj Offset
Material – Reliable – Contrast
Competitor
Contrast
Distractor
Target
Noun Offset
Adj Offset
Material – Unreliable – No Contrast
Competitor
Distractor
Contrast
Target
Noun Offset
Adj Offset
Material – Unreliable – Contrast
Competitor
Contrast
Distractor
Target
Noun Offset
Adj Offset

Figure 5

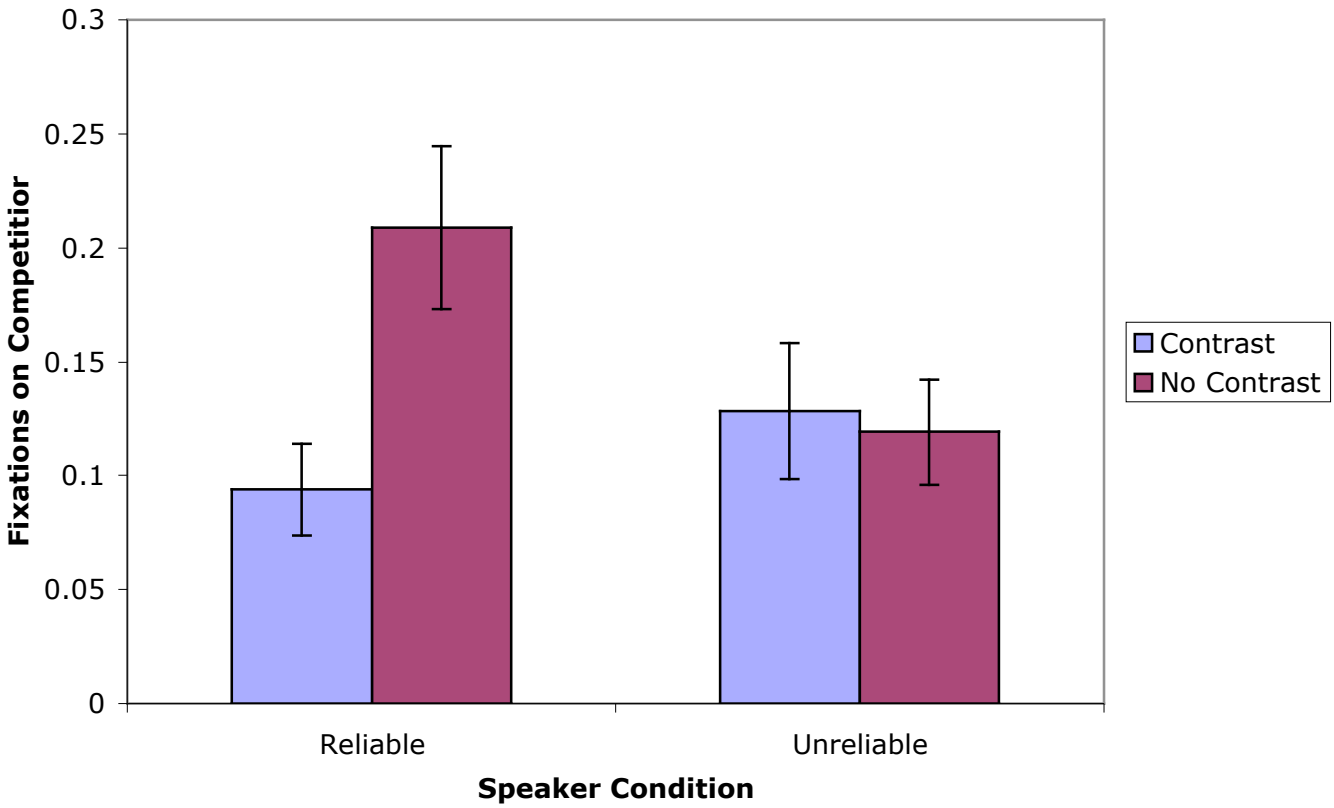
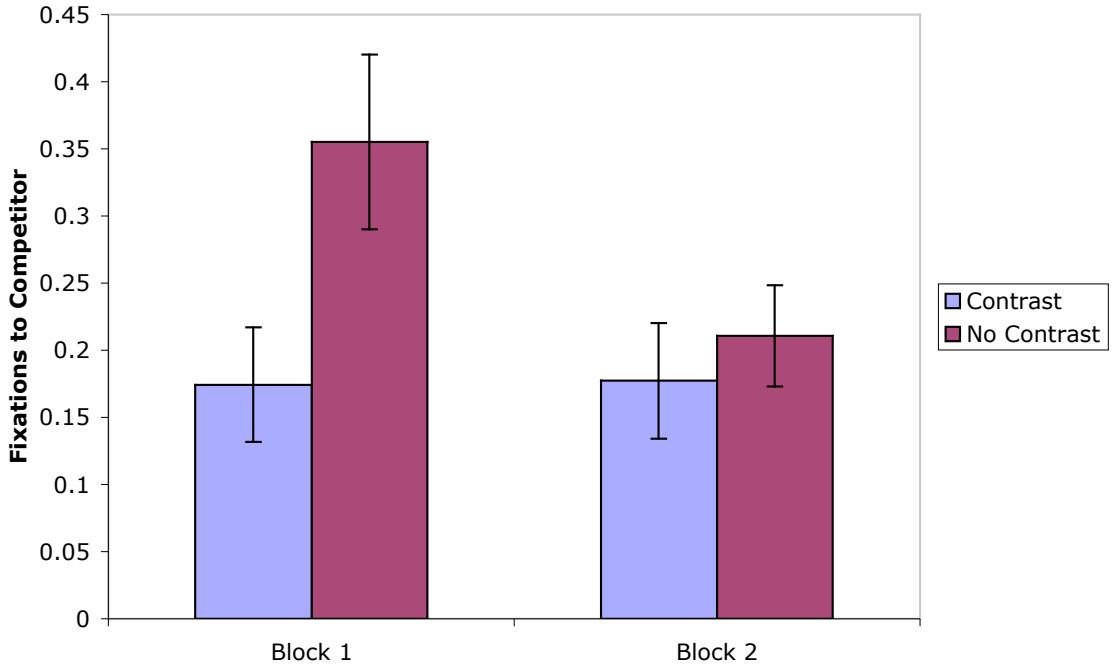

0.3
0.25
0.2
0.15
0.1
0.05
0
Fixations on Competitior
Reliable
Unreliable
Speaker Condition
Contrast
No Contrast

Figure 6


0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Fixations to Competitor
Block 1
Block 2
Contrast
No Contrast

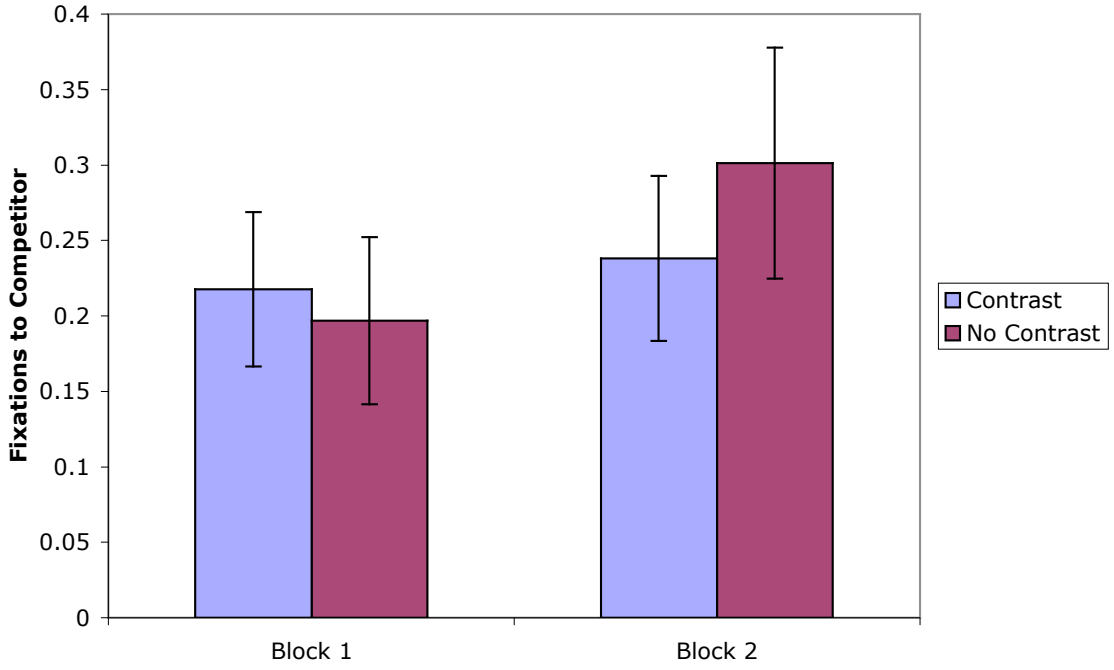
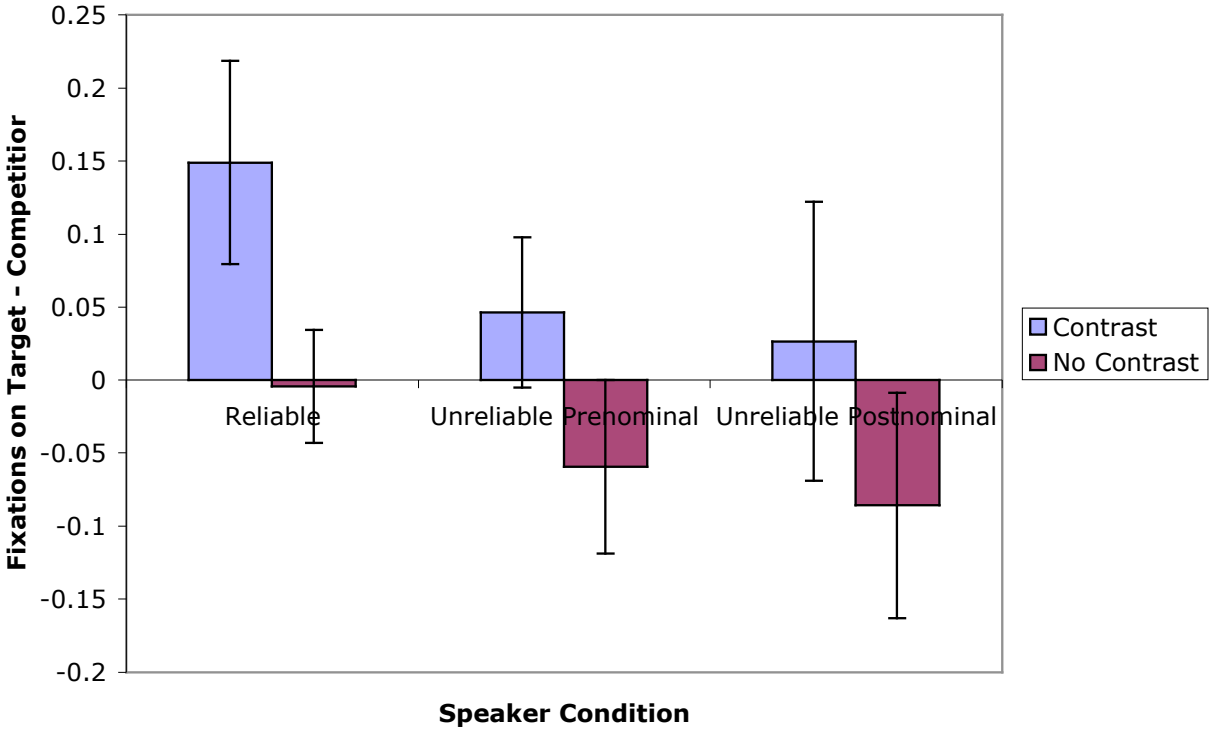

0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Fixations to Competitor
Block 1
Block 2
Contrast
No Contrast

Figure 7


0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
Fixations on Target - Competitior
Reliable
Unreliable Prenominal
Unreliable Postnominal
Speaker Condition
Contrast
No Contrast