

Published in final edited form as:

*J Mem Lang.* 2009 August 1; 61(2): 171–190. doi:10.1016/j.jml.2009.04.003.

## Partner-specific interpretation of maintained referential precedents during interactive dialog

Sarah Brown-Schmidt

Beckman Institute and Department of Psychology University of Illinois at Urbana-Champaign

### Abstract

In dialog settings, conversational partners converge on similar names for referents. These *lexically entrained* terms (Garrod & Anderson, 1987) are part of the common ground between the particular individuals who established the entrained term (Brennan & Clark, 1996), and are thought to be encoded in memory with a partner-specific cue. Thus far, analyses of the time-course of interpretation suggest that partner-specific information may not constrain the initial interpretation of referring expressions (Kronmüller & Barr, 2007; Barr & Keysar, 2002). However, these studies used non-interactive paradigms, which may limit the use of partner-specific representations. This article presents the results of three eye-tracking experiments. Experiment 1a used an interactive conversation methodology in which the experimenter and participant jointly established entrained terms for various images. On critical trials, the same experimenter, or a new experimenter described a critical image using an entrained term, or a new term. The results demonstrated an early, on-line partner-specific effect for interpretation of entrained terms, as well as preliminary evidence for an early, partner-specific effect for new terms. Experiment 1b used a non-interactive paradigm in which participants completed the same task by listening to image descriptions recorded during Experiment 1a; the results showed that partner-specific effects were eliminated. Experiment 2 replicated the partner-specific findings of Experiment 1a with an interactive paradigm and scenes that contained previously unmentioned images. The results suggest that partner-specific interpretation is most likely to occur in interactive dialog settings; the number of critical trials and stimulus characteristics may also play a role. The results are consistent with a large body of work demonstrating that the language processing system uses a rich source of contextual and pragmatic representations to guide on-line processing decisions.

### Keywords

conceptual pact; entrainment; precedent; perspective; reference; common ground; conversation; eye-tracking

---

In recent years, a central focus of research in language processing has been the contribution that contextual information makes to on-line language interpretation. One problem that has garnered significant interest is whether addressees use information about the identity of the speaker to constrain interpretation of referring expressions. For example, imagine a situation in which two friends look up at the clouds and agree to call a particular cumulus cloud *the fuzzy*

---

© 2009 Elsevier Inc. All rights reserved.

Please direct correspondence to: Sarah Brown-Schmidt, 405 N Mathews Ave Urbana, IL 61801, brownsch@illinois.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

*bunny*. In this situation, they would jointly know that if one of them used the term *fuzzy bunny*, that it would refer to that particular cloud, and that if they use a different term, say, *the running man*, that it most likely refers to a different entity. To understand whether addressees use information about the speaker's identity, we can ask whether the same assumptions would be made of a new speaker using these terms who was not present when the friends agreed to call that cloud *the fuzzy bunny*. The present research examines the question of whether interpretation of expressions like *the fuzzy bunny* is sensitive to the speaker's identity, and explores whether the use of speaker information is sensitive to contextual factors including properties of the potential referents and the interactivity of the conversational situation.

The process of developing shared names has been termed *lexical entrainment* (Garrod & Anderson, 1987). Entrained terms are common in conversation (Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989), and are associated with increases in the efficiency and accuracy of communication that occur over the course of a conversation (Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992; Issacs & Clark, 1987). Entrained terms are created through interactive dialog processes in which speakers tailor expressions for particular addressees (Bromme, Jucks, & Wagner, 2005; Lockridge & Brennan, 2002; Haywood, Pickering & Branigan, 2005; Issacs & Clark, 1987), and addressees accept, refuse, or refine these terms (Clark & Brennan, 1991; Traum, 1999).

Brennan and Clark (1996) argued that entrained terms represent *partner-specific* conceptual pacts between the dialog partners, which are flexible agreements for how to conceptualize and refer to discourse referents. On this view, when partners entrain on an expression, the mapping between referent and entrained expression is associated with the particular individuals who were involved in the entrainment, and is part of their common ground (Clark & Marshall, 1978, 1981; Clark, 1992; 1996). Individuals not privy to the conversation would not be expected to share this information, since common ground is defined with respect to specific individuals.

Partner-specificity of entrained terms has several implications for the representation and processing of referring expressions in dialog. Regarding representation, the association between an entrained term and its referent are thought to be stored in episodic memory representations along with contextual information that includes, among other things, information about the discourse partner (Metzing & Brennan, 2003; also see Clark & Marshall, 1978; 1981). Indeed, conversational partners retain entrained terms in memory, and are likely to re-use them after delays of 2–5 days (Markman & Makin, 1998). Further, re-use of entrained terms is sensitive to the knowledge state of the addressee: whether a speaker uses a previously entrained term, or offers a more specific one, depends on whether the speaker believes the addressee to be familiar with the entrained term (Wilkes-Gibbs & Clark, 1992; Brennan & Clark, 1996; Horton & Gerrig, 2005a).

Representational claims that entrained terms are stored together with contextual information naturally lead to the question of how this information is used in subsequent linguistic exchanges. For language comprehension, which is the focus of this paper, the central question that has emerged is whether, and when, partner-specific information is available to the language processing system. Two broad views have been proposed which differ primarily in whether partner-specific information is thought to be available to the language processing system early, or only after a delay. Thus far, empirical findings point to delayed use of partner-specific information (exceptions being Metzing & Brennan, 2003, and Brennan & Hanna, 2009), however, a variety of methodological concerns complicate interpretation of existing results.

The primary aim of the present research is to re-examine the role of partner-specific representations in the processing of referring expressions and to propose a revised account of

the role of partner-specific information in language processing. In service of this aim, I present the results of three experiments that examine the use of partner-specific information during on-line processing in both interactive (Experiments 1a, 2), and non-interactive settings (Experiment 1b). I go on to critique the use of non-interactive methodologies in investigations of interaction-based processes and suggest that both automatic and strategic language processes are sensitive to the context of language use.

## Partner-specific language interpretation? Two views

According to one view, partner-specific and other contextual information guides language processing decisions from the earliest moments of comprehension (Metzing & Brennan, 2003). This view is consistent with constraint-based theories of language processing, which propose that a large number of partial constraints guide language processing decisions (MacDonald, Pearlmutter & Seidenberg 1994; Tanenhaus & Trueswell, 1995). It is also consistent with a variety of results demonstrating that many other forms of contextual information, including common ground representations (Hanna, Tanenhaus & Trueswell, 2003; Brown-Schmidt, Gunlogson & Tanenhaus, 2008; Heller, Grodner & Tanenhaus, 2008), object affordances (Chambers, et al., 2002), and the reliability and certainty of the speaker (Grodner & Sedivy, in press; Arnold, Tanenhaus, Altmann & Fagnano, 2004), all constrain on-line processing. Thus, on this view, any information in memory associated with an entrained term, such as where and with whom the term was entrained with, and so forth, could potentially constrain future interpretations of the term. However, in practice, Metzing and Brennan (2003) argue that partner-specific information may be a relatively weak cue which can be overridden by other, conflicting cues.

An alternative view distinguishes early, on-line processes, which are partner-nonspecific, from late recovery processes which do have access to partner-specific information (e.g. Kronmüller & Barr, 2007; Barr & Keysar, 2002). This view is motivated by the notion that taking rich contextual information into account during on-line processing is too taxing to be part of routine processing strategies. It is based on dual-process accounts of the role of perspective in language processing, which propose that initial processing of language is egocentric and that perspective only plays a role in delayed, recovery processes (Keysar, et al., 1998; Keysar, Lin & Barr, 2003). This view is consistent with results that show that listeners sometimes generate perspective-inappropriate interpretations of their partner's utterances (Keysar, et al., 2000; Keysar, Lin & Barr, 2003). According to this view, stored associations between naming precedents and referents guide initial on-line processing, however, stored associations between these naming precedents and specific *partners* only affect processing decisions at a delay.

Work supporting the dual-process view includes findings by Barr and Keysar (2002), who hypothesized that if partner-specific representations are used on-line, entrained terms spoken by the speaker who established the term should be easier to understand than entrained terms spoken by a new speaker. In their second experiment, a live speaker produced expressions that had previously been entrained by either that live speaker, or by a different, pre-recorded voice. Listeners identified the referent of the entrained term equally quickly, regardless of which speaker established the term, suggesting that addressees did not take advantage of a partner-specific cue. However, aspects of the experimental design may have led to the null result. Partner-specific conceptual pacts are thought to enter common ground when they have been interactively established, but in this experiment, participants did not have the opportunity to refer to or collaboratively establish names for the objects. Thus, participants may not have associated the terms with a particular speaker. It is also possible that the terms were associated with both speakers. After all, the live speaker was in the room at the time of the entrainment by the pre-recorded speaker, and work by Horton (2007) demonstrates that simply having someone in the room can serve as a memory cue for representations associated with that person.

Barr and Keysar's (2002) experiment 3 tested whether participants would exhibit partner-specific interpretation of basic level terms such as *car* in contexts containing a car and a flower. Prior to the test trials, either the test voice, or a different voice (both pre-recorded) described these pictures in contexts that required using the specific terms *sportscar* and *carnation*. They reasoned that if participants use partner-specific information as they interpret *car*, lexical competition from the carnation (flower) should be stronger when the voice at test had previously established the precedent of *carnation*. They also manipulated whether the participant was led to believe (through an elaborate ruse) that they were talking with a live person, or if they knew they were listening to pre-recorded speech. The results showed equal competition, regardless of which voice established the precedent, suggesting that partner-specific representations are not used on-line. Whether the participant was told they were listening to a live person or not had no effect. However, even in the condition with the elaborate ruse, the experiment was essentially non-interactive, thus there was no opportunity for collaborative establishment of entrained terms. Additionally, as Metzing and Brennan (2003) point out, a robust lexical competition effect may have swamped out whatever small partner-specific effect there might have been.

In one experiment that did use an interactive task with live speakers, Metzing and Brennan (2003) reported that maintained precedents were interpreted equally quickly, regardless of speaker. However, they did find evidence for the use of partner-specific cues when precedents were broken: After the participant and a partner collaboratively established terms for various images, that partner left the room and either returned, or a new partner entered. The original or new partner then referred to a test image either using the entrained term or a new term. When a new term was used, eye-tracking data showed that participants were significantly slower to fixate the referent when it was the original partner speaking, compared to a new partner (for similar findings with 3-year-olds and an interactive task, see Matthews, Lieven, & Tomasello, 2008). This partner-specific penalty for broken precedents demonstrates that addressees formed partner-specific expectations for how the various images would be referred to. With a new speaker, there were no expectations, however with a familiar speaker, addressees expected them to continue to use the established terms.

Kronmüller and Barr (2007) questioned whether this partner-specific effect represents initial, on-line interpretation processes or only a late recovery process. Like Metzing and Brennan (2003), they monitored participants' eye movements as they interpreted maintained and broken precedents when either the original or a new speaker produced the expression. However, they used a non-interactive setting with two different pre-recorded voices. Like Metzing and Brennan (2003), they reported no partner effect for maintained precedents<sup>1</sup>. Instead, they observed an early effect of precedent, with less competition when speakers continued to use established precedents, compared to new terms. A partner-specific penalty was observed for original speakers using a new term compared to new speakers, but only after a delay. This delay was used to argue for a two-stage model in which initial interpretation processes are "perspective-free", and thus do not take advantage of partner-specific information. The authors concluded that when listeners hear a new term, the existing precedent preempts mapping of the new term to the target referent, yielding the early precedent effect. The later speaker effect emerges when listeners use speaker-specific information to inhibit precedents not known to the current speaker. However, one concern with the interpretation of these results is that the use of a non-interactive paradigm may have yielded precedents that were only weakly associated with one speaker over the other, resulting in a relatively late and weak speaker-specific effect.

---

<sup>1</sup>A post-hoc analysis of these data (Barr, 2008a) found that for maintained precedents, looks to the target increased at a higher rate for original vs. new speakers, however, condition differences at baseline complicate interpretation of this result.

This article presents the results of three experiments that re-examine the question of whether partner-specific information is used by early, on-line interpretation processes. A primary intuition guiding this work is that insights into the role of interactively established representations, such as collaborative referential pacts, will require examinations of language use in natural, interactive settings. After all, partner-specific conceptual pacts are thought to emerge from an interactive process during which interlocutors interactively refine proposed conceptualizations of a referent (Brennan & Clark, 1996). In non-interactive discourse settings, an addressee does not have the opportunity to interactively establish precedents, so it is likely that partner-specific information would not be stored when referring precedents are learned. Thus, previous work that failed to observe the use of partner-specific representations during on-line interpretation (e.g. Barr & Keysar, 2002), or found only very late partner-specific effects (e.g. Kronmüller & Barr, 2007) only reaffirms the collaborative nature of the representations, rather than establishes their non-use in on-line processing.

The motivation for studying entrained terms using interactive methodologies comes from experiments that show that actively engaging in a conversation and interactively establishing entrained terms affects how those terms are used and understood: In direction-giving tasks where partners entrain on terms for abstract pictures, overhearers who are privy to the entirety of the conversation do not comprehend entrained terms as well as active participants. If the overhearer subsequently completes the same task with one of the original partners, the resulting conversation tends to be less efficient, taking more time and more words to complete the task (Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992). Further, the feedback that dialog partners give each other plays a particularly important role in understanding what is said. In story-telling tasks, providing feedback to a speaker improves the listener's understanding of the story. Eavesdroppers on the conversation do not show the same advantage, suggesting that active engagement allows listeners to coordinate their specific needs with what the speaker says (Kraut, Lewis & Swezey, 1982). Similarly, recorded instructions are harder to follow than those produced in a live, interactive setting (Clark & Krych, 2004). Interactive dialog also results in better interpersonal perception of the dialog partner compared to simply observing the speech of the partner (Powell & O'Neal, 1976). These results demonstrate that in dialog tasks, interactivity plays a direct role in the addressee's understanding and ability to complete the task. Thus, determining whether partner-specific information is used during the interpretation of entrained terms will likely require experiments that allow these terms to be interactively established.

The experiments presented here examine the role of partner-specific information in the on-line interpretation of broken and maintained precedents. Experiments 1a and 2 both use an interactive conversation methodology with live, co-present interlocutors. Experiment 2 differed from Experiment 1a in that it provided more experimental control over the critical referring expressions (at the expense of spontaneity), and used scenes with unmentioned images to control for an alternative explanation of Experiment 1a's results (see Experiment 2). Experiment 1b was designed to be as similar as possible to Experiment 1a, but without interaction. Participants in Experiment 1b listened to pre-recorded experimental instructions, taken from the experimental recordings of Experiment 1a. The contrast of Experiments 1a and 1b provides insights into the role of interactivity in the use of partner-specific representations.

The design of these experiments is similar to that used by Metzing and Brennan (2003), however, several design changes were employed to maximize the likelihood of observing early effects. First, the critical expressions on which the factors of partner and expression were manipulated were designed to be temporarily ambiguous with respect to the referential context (only some of the stimuli used by Metzing & Brennan were temporarily ambiguous). For example, given a scene that contains some multicolored blocks and some multicolored triangles, the underlined portion of the phrase *the multicolored blocks* is ambiguous between



the blocks and the triangles. The ambiguity is only temporary because it resolves at *blocks*. Temporary ambiguity increases the likelihood of observing early effects because if addressees expect the ambiguous portion of the expression to continue with the entrained term, this would resolve the ambiguity, allowing the addressee to more quickly identify the intended referent.

Recall that Metzing and Brennan (2003) observed a partner-specific penalty for interpretation of broken precedents, but no partner effect for the interpretation of maintained precedents (although condition means showed an advantage for the original partner). One reason for the lack of a partner-specific effect for maintained precedents could be a lack of power; their design included only 2 critical trials per condition. In order to increase the statistical power of the design, the experiments presented here used more experimental items and more participants. These changes increase the likelihood of observing partner-specific effects for maintained precedents, and also increase the likelihood of observing partner-specific effects early in the interpretation process.

## Experiment 1a

Participants in Experiment 1a followed an experimenter's instructions to rearrange abstract images on a computer screen as their eye movements were monitored. Over several trials, participants established entrained terms for these images with an experimenter. For example, they might call what would eventually become the target image *the multicolored blocks*, or alternatively, *the stacked cubes*. On critical trials, either the same experimenter or a new experimenter would then refer to the target image as *the multicolored blocks*. Whether this was the entrained term (maintained precedent trial), or a new term (broken precedent trial), was determined by whether or not the experimenter had previously established this name for the target referent (e.g. *the multicolored blocks*), or a different name (e.g. *the stacked cubes*).

## Predictions

If partner-specific information guides interpretation of a referring expression from the earliest moments of understanding, the initial interpretation of expressions should be sensitive to the speaker, and the addressee's experience with that speaker. Specifically, this view predicts two different types of results, depending on whether the critical expression *maintained* the referring precedent or *broke* the precedent.

An early partner-specific effect for maintained precedents would be indicated by an early facilitation in the interpretation of the expression when the original partner is speaking, compared to a new partner. In terms of the eye fixation response measure, in the window of time beginning just after the onset of the critical expression, this would result in more target fixations and fewer competitor fixations with the original speaker, compared to a new speaker. This effect would occur if, when hearing an entrained expression spoken by the partner who entrained it, addressees interpret the ambiguous portion of the expression (e.g. *the multicolored..*) as mapping to the target, rather than the competitors. In contrast, when hearing an entrained expression spoken by the new partner, addressees should not eliminate the competitors from consideration, and thus fixate the target and competitors with equal probability until the disambiguating word *blocks*. Note that partner-specific interpretation of maintained precedents has not been observed previously. This may be because the partner-specific cue is easily overwhelmed by other sources of information (see argumentation in Metzing & Brennan, 2003), or because partner-specific information does not routinely guide interpretation processes (Kronmüller & Barr, 2007). Thus, a null effect for maintained precedents is not conclusive. By contrast, an early partner-specific effect for maintained precedents would be clear evidence in favor of the view that entrained terms are stored in memory with partner-specific information, and that the processing system takes advantage of this information to facilitate on-line interpretation.

The second type of result predicted by the view that partner-specific information guides initial interpretation processes is that addressees, upon hearing a broken precedent, should be *less* likely to consider the target when the original partner is speaking, compared to a new partner. In terms of fixations, this would result in fewer target and more competitor fixations immediately after the onset of the broken precedent when the original partner is speaking, compared to a new partner. This would occur if addressees assume that if the original speaker had intended to refer to the target, that they would use the entrained term, thus use of a different term is interpreted as a non-target reference (e.g., *the principle of contrast*, E. Clark, 1987). In contrast, with a new speaker, use of a broken precedent would be interpreted as temporarily ambiguous between the target and competitors, thus target fixations should be slightly higher with a new speaker. If the initial interpretation of broken precedents is impaired when the original partner is speaking, this would demonstrate that addressee's on-line interpretation of an expression takes into consideration speaker-specific expectations for how particular entities would be referred to even when the speech input does not match that expectation.

Alternatively, partner-specific information may not guide the initial, on-line interpretation of maintained or broken precedents. If so, maintained precedents should be easier to interpret than broken precedents (due to learned associations between referents and expressions), and this should not depend on whether the current partner established the precedent. Finally, if, as Kronmüller and Barr (2007) argue, partner-specific effects for broken precedents are part of a late recovery process, a partner effect for broken precedents should emerge later in processing, after an early effect of precedent. In terms of fixations, this would predict that in early analysis regions (e.g. just after referring expression onset), there should be more target fixations and fewer competitor fixations for maintained, compared to broken precedents, with no interaction with speaker. Following this precedent effect, there should be more target fixations and fewer competitor fixations for new partners and broken precedents, compared to original partners and broken precedents.

## Method

**Participants**—Forty-eight native English-speaking participants from the student community at the University of Illinois at Urbana-Champaign participated in exchange for \$16 or partial course credit. Data from an additional 8 participants were collected but not analyzed due to not following directions (n=1), a poor calibration (n=1), or experimenter error (n=6).

**Procedure**—Participants were seated in front of a computer screen. Each participant interacted with two different experimenters (one male, one female). One experimenter sat at a different computer in the same room, positioned so that the participant and experimenter could not view each other's screen. A second experimenter sat in the hallway, behind a closed door. Participants understood that the experimenters worked in the lab, and it was clear that the experimenter in the hallway could not hear into the room when the door was closed. Participants were told that they would receive instructions from two different experimenters, and that they should follow the instructions as best they could. Across participants, a total of eight different individuals played the role of experimenter, however each participant only interacted with two individual experimenters.

The participant's eye movements were monitored throughout the experiment. Fourteen of the participants were run on a head-mounted Eyelink II eye-tracking system, and the remaining participants were run on an Eyelink 1000 desktop mounted eye-tracking system<sup>2</sup>. The experimenter wore a headset microphone. An audio record of the conversation was recorded to disk.

<sup>2</sup>Statistical analyses confirmed that none of the variables of interest interacted with the type of eye-tracker used.

The task was a referential communication task (Krauss & Weinheimer, 1966; Clark & Wilkes-Gibbs, 1986) in which participants followed an experimenter's spoken instructions to rearrange a set of hard-to-name images on a computer screen. On each trial, the participant and experimenter saw the same 10 images on their respective computer screens, but in different arrangements (Figures 1a–b). The experimenter described each image in turn (left to right, top to bottom) and the participant's task was to rearrange his or her images into the experimenter's order. On the experimenter's screen only, there were verbal prompts which indicated how the experimenter should begin his or her descriptions of two of the ten images. These prompts were employed in order to have control over the name that was given to the critical images (see *Auditory stimuli*, below). The task was interactive, and participants were encouraged to ask for clarification if they did not understand an instruction. The task typically lasted 60–80 minutes.

Participants rearranged 16 sets of 10 images according to the experimenter's instructions. Each set of images was rearranged 4 times in a row, for a total of 64 rounds of trials (Figure 2). Each participant was assigned a primary experimenter (the gender of the primary experimenter was counterbalanced across subjects). For a given set of images, the first three rounds of rearranging were always done with the primary experimenter. These rounds were used to collaboratively establish referring expressions for the 10 images in that set. On the fourth round, the experimenter always left the room for a minute and either the primary experimenter returned (same-partner trials) or the alternate experimenter came into the room (different-partner trials). In either case, the experimenter who was not participating in the fourth round remained in the hallway with the door closed. The fourth round began with a drift correction on the eye-tracker, and then the experimenter (either the primary or alternate) gave the participant instructions to rearrange the images for a fourth time.

For each set of images, the first instruction on the fourth round was the critical trial during which the variables of interest were manipulated (within subjects): partner (same or different), and referring expression precedent (maintained or broken). The first instruction always referred to the target image for that set of images. Interpretation of the target referring expression was monitored by examining participants' gaze as they interpreted this expression.

## Materials

**Visual stimuli:** The visual stimuli were 160 abstract images. Several of the images were created for a previous experiment by Arnold, Hudson Kam and Tanenhaus (2007). Abstract images were used so that participants would not expect both speakers to use the same canonical label for the image (e.g., *dog*). In order to test expressions with a variety of modifiers, half of the images were black and white, half were colorized. Thirty-two of these images were the target or competitor images on critical trials; 16 were colored, 16 were black and white. The 16 colorized critical images were grouped into 8 pairs of images that had the same color theme (e.g. two multicolored images), and could be described using the same adjective, as in *the multicolored...* The 16 black and white critical images were grouped into 8 pairs that could be described using the same adjective, such as *the crushed...* The adjectives for the black and white images were selected based on an earlier norming study in which speakers who did not participate in the present experiments named each image.

The 160 images were grouped into 16 sets of 10 images each, with a single critical pair of images per set. In addition to the target and competitor (critical) images, several of the sets of images contained additional potential competitors, such as a greenish image on a trial where the target could be described as *the green blob*. Experimenters were instructed not to begin their descriptions of the non-target images using the critical adjective. For each image set, the 10 images were arranged randomly on the participant's screen into two rows of 5 pictures. The order of the images on the experimenter's screen (i.e. the order that the pictures would be



rearranged into) was also randomized, with the exception that on the fourth round, the target image for that trial was always the first image mentioned (i.e. the top left image).

**Auditory stimuli:** On the experimenter's computer screen, the target and competitor images each had a descriptive word written below it (Figure 1a). The experimenter began his or her description of the image with this word, as in *the multicolored pile of blocks*, or *the pointed row of teepees*. As is typical with naturally produced picture descriptions in referential communication tasks (e.g. Clark & Wilkes-Gibbs, 1986), the description was usually longest on the first round of matching, as in *the multicolored pile of like kiddy blocks that kind of look like they're going to fall over*. This description would then be shortened to *the multicolored blocks* by the third round. For the remaining 8 images, the experimenter developed names for them on the fly, in collaboration with the participant<sup>3</sup>.

In order to examine the effect of a maintained vs. a broken precedent, whether the name of the target image in rounds 1–3 was the same as it was on the critical trial (first reference in the fourth round) was manipulated (Table 1). For a given set of images, the critical descriptive word in the fourth round was the same across the maintained precedent and broken precedent conditions. What was manipulated was the descriptive word used in rounds 1–3. Thus, for broken precedent trials, a different descriptive word was used for rounds 1–3 than in the critical fourth round. In contrast, for critical trials with a maintained precedent, the description of the target image was the same across the 4 rounds of matching.

**List rotations:** This experiment manipulated 2 variables, partner (same or different) and referring precedent (maintained or broken), yielding four experimental conditions. The 16 sets of images were rotated through the 4 conditions across 8 lists, which also counterbalanced the target and competitor (i.e. on half of the lists, in the example image set, the cubes would be the target, and in the other half of the lists, the teepees would be the target). The 16 sets of images were presented in one of two random orders. Since each set of images yielded a single critical trial, each subject saw a total of 4 critical trials in each of the 4 conditions.

## Analysis and Results

**Characteristics of Critical Instructions**—Because the stimuli in this experiment were naturally produced, it is necessary to examine the experimenters' productions to see if and how they varied across conditions. First, the experimenters' descriptions' of each image was transcribed. Trials on which an experimenter made an error (e.g. said the wrong name), or made a disfluent repair (e.g. *the bloc- uh the stacked blocks*) were excluded from all following analyses. This eliminated 6% of critical trials; trial rejection rate did not differ across the conditions.

Next, prosodic analyses of the pitch (minimum, maximum and average), intensity, and duration of the determiner and critical adjective were analyzed (it was not possible to analyze other words in the expressions due to variations in expression length). Because speakers differ in absolute pitch and speech rate, prosodic analyses were conducted for each experimenter separately. Of the eight different experimenters, only four ran enough subjects and had enough codeable trials to be included in the analysis. The analysis of a randomly selected subset (50%) of critical trials revealed no prosodic differences across the conditions.

Finally, in order to determine when the critical instructions uniquely identified the target referent, the time from the onset of the critical ambiguous adjective (*multicolored*) to the onset

<sup>3</sup>Because expressions were partially scripted and experimenters completed the task multiple times (between 1 and 22 times, mean=12), there was some consistency of expressions across participants, however individual participants were not aware of this consistency.

of the head noun (*blocks*) was measured. Since critical instructions were spontaneously produced (with the exception of the critical adjective), expressions were quite variable. However, the noun provided a reasonable estimate of the point at which the target was uniquely identified, also known as the point-of-disambiguation (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). The point-of-disambiguation did vary significantly across conditions (see Appendix A): The point-of-disambiguation was earlier for original partners (maintained precedents = 354ms; broken precedents = 546ms), compared to new partners (maintained precedents = 625ms; broken precedents = 835ms), though this difference did not reach significance by participants. Additionally, the point-of-disambiguation was earlier for maintained expressions. There was no interaction. Since this was naturally produced speech, these effects are not unexpected, since repeated expressions in dialog tend to be shorter and more fluent (see Bard et al., 2000), and original partners had more experience producing expressions to a given participant (though not more experience overall, since experimenters swapped primary and secondary roles across subjects). And, as Metzing and Brennan (2003) point out, controlling this feature of natural language production would likely require use of pre-recorded, cross-spliced speech, which would necessarily eliminate features of the interactive situation which are central to the phenomena of interest. Critically, however, the earliest disambiguating point (354ms) is still late enough to allow for an examination of fixations made in response to the critical adjective, but before the point-of-disambiguation would be expected to affect looking patterns.

**Eye-tracking Analyses**—The participants' interpretation of the critical referring expression was analyzed in two ways. First, we examined the effect of partner (original or new) and expression (maintained or broken) on the latency to fixate the target. This was the primary gaze analysis used by Metzing and Brennan (2003). Second, we examined the fixations that participants made over time, as they interpreted the referring expression. Time course analyses were based on the "target advantage" score (Arnold, et al. 2000), which we defined, like Kronmüller and Barr (2007), as the proportion of fixations to the target minus the average proportion of fixations to all other images in the scene. For both of these measures, there was no main effect of the type of critical image (black and white vs. colorized), nor did image type significantly interact with either of our variables of interest, thus image type is not included in our analyses presented below.

**Latency to fixate the target:** The latency at which participants fixated the target following critical adjective onset (Figure 3), was analyzed in an ANOVA with partner and expression as factors (Table 2). Stimulus driven fixations are not expected until approximately 200 ms following stimulus onset, given the time needed to plan and execute an eye movement (Hallett, 1986), thus we only examined fixations beginning 200 ms after adjective onset. For trials on which the participant was already fixating the target at 200 ms (7% of the data), the fixation time was set to 200 ms.

An effect of expression was significant by participants only, due to earlier target fixations when the precedent was maintained. A partner by expression interaction was significant by items. Recall Metzing and Brennan's (2003) finding that target fixations were significantly delayed when the original partner used a broken precedent compared to a new partner, but no difference for maintained expressions. In order to compare our findings with theirs, planned comparisons examined the interaction. Unlike Metzing and Brennan's findings, paired comparisons by expression type were not conclusive: for maintained precedents, there was no effect of partner ( $t_s < 1$ ), and for broken precedents the effect of partner was significant by items only,  $t(47) = 1.62, p = .11$ ,  $t(15) = 2.83, p < .05$ . Instead, comparisons by partner revealed the partner-specific effect: When a new partner was speaking, there was no effect of expression type,  $t_s < .5$ , however, with the original partner, participants were significantly slower to fixate the target

with a broken precedent compared to a maintained precedent,  $t1(47)=2.22, p<.05, t2(15)=2.66, p<.05$ .

While the pattern of results was not identical to the results reported by Metzing and Brennan (2003), the findings are consistent: the biggest delay in target fixations was when the original partner broke the precedent. This suggests that addressees were particularly confused when the person who created the precedent broke it. This confusion was not present for new partners, suggesting that, consistent with partner-specificity, addressees had different expectations for how new and old partners would refer to the game-pieces.

**Target advantage scores:** While the results for target fixation latencies are consistent with partner-specific interpretation of the critical referring expressions, latency measures do not reveal whether the effects occurred early in the interpretation process, or only after a delay, as suggested by Kronmüller and Barr (2007). Target advantage scores (Figure 4) were time-locked to the onset of the critical adjective and analyzed in five consecutive 400 ms time regions. The first region was used as a baseline, and encompassed fixations between 200 ms before to 200 ms after critical adjective onset. Region 2 included fixations between 200 to 600 ms after adjective onset. This region captured fixations that were planned in response to the critical adjective, and for the most part before the disambiguating noun, which occurred between 354–835ms following adjective onset (given 200ms to program and launch an eye movement, fixations in response to the noun are not expected until 554–1035ms). Region 3 included fixations from 600 to 1000 ms post-adjective, and so forth. The target advantage scores were analyzed in a series of planned ANOVAs, one at each time region, with partner and precedent as factors (Table 3).

The first region to show significant effects of condition was region 2, which immediately followed adjective onset (Figure 5). A significant partner by expression interaction at region 2 was due to a significant effect of partner when the precedent was maintained, with higher target advantage scores for the original partner, compared to the new partner,  $t1(47)=2.29, p<.05, t2(15)=2.72, p<.05$ . There was no effect of partner for broken precedents,  $ts<1^4$ .

This early partner-specific effect on the interpretation of an entrained expression demonstrates that initial interpretation of the expression took into account the partner-specific referring precedent: When addressees heard a familiar partner use a contextually ambiguous adjective such as *multicolored*, given scenes like Figure 1b, addressees interpreted this expression as referring to the target, consistent with their established conceptual pact. In contrast, when a new partner used the same adjective, addressees did not assume that the conceptual pact held.

At this time region, we did not observe partner-specific interpretation of broken precedents; when a broken precedent was used, addressees were just as confused, regardless of who was speaking. However, post-hoc analyses at a slightly earlier time region (180–300ms, which was the first time region analyzed by Kronmüller & Barr, 2007), revealed a very early partner-specific effect for broken precedents: When the precedent was broken, target advantage scores were significantly *lower* for original compared to new partners ( $-.042$  [SE=.01] vs.  $.021$  [SE=.02], respectively),  $t1(47)=2.46, p<.05, t2(15)=2.55, p<.05$ . This early effect is inconsistent with Kronmüller and Barr's claim that partner-specific effects for broken precedents is due to a late recovery process. With maintained precedents, there was a marginally significant advantage for original partners compared to new partners ( $.002$  [SE=.02] vs.  $-.041$  [SE=.02], respectively), consistent with the effect observed in region 2,  $t1(47)=1.73, p=.09, t2(15)=1.55, p=.14$ .

<sup>4</sup>An identical pattern of results obtained when region 2 was truncated to only include fixations which occurred before the earliest average point-of-disambiguation (554ms post-adjective).

The main effect of expression at regions 3–5 was due to higher target advantage scores for maintained expressions, demonstrating that a familiar expression-to-picture mapping was easier to interpret than a new expression-to-picture mapping. An effect of partner at regions 4 (marginal) and 5 was due to higher target advantage scores for new partners. This late effect may be due to participants looking longer at target referents before clicking on them when a new partner was speaking, possibly reflecting uncertainty or unfamiliarity with the new partner. Thus, in this case, late and lingering target looks may reflect the slower linking of expression and referent when the speaker was new.

## Conclusions

The results of experiment 1a demonstrate that addressees represented partner-specific information about entrained terms, and that they used this information as they interpreted the utterances on-line. The results of the time-course analysis showed a clear, early advantage for the original over the new partner for maintained expressions. This suggests that when interpreting an expression beginning with a term like *multicolored*, which was temporarily ambiguous between multiple potential referents in the scene, addressees used the established conceptual pact to interpret the expression as referring to the referent of the entrained term, but only when the addressee knew that the speaker was aware of the precedent. When the addressee did not know if the speaker knew about the precedent, they did not assume the speaker was talking about the referent of the entrained term, and instead considered multiple potential referents.

The partner effect for maintained precedents is inconsistent with previous work using non-interactive paradigms (Barr & Keysar, 2002; Kronmüller & Barr, 2007), which found no advantage for original partners and maintained precedents. The one experiment which did use an interactive setting (Metzing & Brennan, 2003) consistently showed non-significant effects in the same direction reported here, suggesting the lack of an effect in previous work may have been due to a lack of power. Another factor which may have contributed to the observation of this effect is that the scenes contained multiple competitor images, including at least one which temporarily matched the critical expression. These competitors may have given the task enough complexity to lead addressees to temporarily consider plausible alternative referents when the speaker was not expected to be aware of the precedent (i.e. in new partner trials).

When addressees interpreted expressions that broke a referring precedent, comprehension was impaired if it was the original partner who was speaking. Like Metzing and Brennan (2003), in the analysis of latency to fixate the target, the condition with the largest delay was the condition where the original partner used a broken precedent. The fact that the original partner advantage for maintained precedents did not extend to broken precedents shows that the maintained precedent effect was not simply due to familiarity with the original partner. The penalty that occurred when original partners used broken precedents suggests that addressees assumed that if the original partner wanted to refer to the target, they would use the entrained expression; this assumption was not applied to new partners. However, the timing of the broken-precedent effect was different from what has previously been reported. The partner effect for broken precedents was small and very early (180–300 ms post-stimulus). In comparison, the same effect emerged at about 900 ms post-stimulus in Metzing and Brennan's experiment (Brennan & Hanna, 2009), and at 1500 ms (Exp 1.) and 900 ms (Exp 2., no-load condition) in Kronmüller and Barr's experiments. One explanation for the earlier effects in the current study is the use of an interactive paradigm paired with a more powerful experimental design than was used by Metzing and Brennan (2003). Another possibility is that subtle differences in stimulus characteristics, such as the number of alternative potential referents, or the inclusion of competitors which temporarily matched the critical expression, modulate whether addressees consider competitor referents before eventually fixating the target.

These results add to the evidence that addressees use common ground representations to guide on-line interpretation of utterances (Hanna, Tanenhaus & Trueswell, 2003; Brown-Schmidt, Gunlogson & Tanenhaus, 2008; Nadig & Sedivy, 2002; Heller, Grodner & Tanenhaus, 2008). More generally, the results are consistent with a view of sentence processing in which many partial constraints, including pragmatic and contextual information, guide on-line language processing decisions (MacDonald, Pearlmutter & Seidenberg, 1994; Tanenhaus & Trueswell, 1995). The locus of partner-specific facilitation for maintained precedents may be learned associations between particular speakers and referring precedents, such that precedents are more easily retrieved during the interpretation process when retrieval is cued by the original partner (Metzing & Brennan, 2003; see Horton & Gerrig, 2005b for a related argument).

Finally, the results of Experiment 1a suggest, but do not demonstrate that an interactive setting is a critical component of addressees' ability to use partner-specific information during on-line processing. Experiment 1b tests the interactivity hypothesis by presenting the same expressions that were heard by participants in Experiment 1a to a new group of participants in a non-interactive setting. If interactivity is critical to partner-specific interpretation, this makes the prediction that in this non-interactive context, the early partner-specific effects should be eliminated.

## Experiment 1b

### Method

The design of Experiment 1b was similar to Experiment 1a, thus only changes to the design are noted.

**Participants**—48 native English speaking participants from the student community at the University of Illinois at Urbana-Champaign participated in this experiment in exchange for \$16 or partial course credit. An additional 3 participants were excluded from analysis due to experimenter error ( $n=2$ ), and a poor calibration ( $n=1$ ). None of the participants had participated in Experiment 1a.

**Procedure**—Participants' eye movements were recorded with an EyeLink 1000 desktop-mounted eye-tracking system. Participants were told that they would be following the instructions of two different (pre-recorded) speakers to re-arrange a series of images. An experimenter who did not appear in the recordings sat with the participant in the room and operated the eye-tracker. The experiment lasted approximately 60 minutes.

On each trial, 10 images were presented on the screen, followed by a description of the first image to be re-arranged. The participant pressed the right mouse button to hear each subsequent description. This was similar to Experiment 1a in that it allowed participants to move at their own pace. Like in Experiment 1a, critical (4<sup>th</sup> round) trials began with a drift correction. The critical description began approximately 1024 ms following presentation of the critical images<sup>5</sup>.

**Materials**—The visual stimuli and trial rotations were identical to Experiment 1a. Participants in Experiment 1b were run on the same experimental lists as participants in experiment 1a, and listened to recordings of image descriptions which were produced for a given Experiment 1a participant who ran on the same list. Thus each of the 48 sets of image descriptions (one for each participant in Experiment 1a) was heard by a single participant in Experiment 1b. The

<sup>5</sup>This description latency was used because it advanced the task at a comfortable pace, although it was ultimately shorter than the (highly variable) description latency in Experiment 1a of ~3500ms. Since eye-tracking analyses are measured from description onset, rather than display onset, there is no reason to expect this difference to affect the critical partner  $\times$  expression interaction of interest.



image descriptions that each Experiment 1b participant heard, and the images that they saw, were identical to what participants in Experiment 1a heard and saw.

The auditory recordings of each image description were saved as individual .wav files and were edited to remove microphone noises, coughs, and any discussion between the participant and experimenter, such as procedural talk about the eye-tracker. On occasion, an image description was unusable due to computer error or microphone problems. In these cases, the description was replaced with a description made by the same speaker on another trial.

In summary, the visual stimuli were identical across Experiments 1a and 1b, and the image descriptions were near-identical across the two experiments. Also recall that the primary and secondary speakers were always of different gender, so their voices should be just as easy to distinguish in Experiment 1b as in Experiment 1a. The key difference, then, is that participants in Experiment 1a interpreted expressions which were specifically designed for them in the context of an interactive dialog setting.

## Results

**Latency to fixate the target**—The latency to fixate the target was calculated in the same way as Experiment 1a (Figure 6). An ANOVA was used to analyze fixation latencies (Table 4). A significant effect of expression was due to faster target fixations for maintained expressions, demonstrating that participants learned the expression-image mappings. Unlike Experiment 1a, the partner  $\times$  expression interaction did not approach significance, suggesting that participants did not distinguish between the speakers when interpreting their expressions<sup>6</sup>.

**Target advantage scores**—Target advantage scores were analyzed in the same way as Experiment 1a (Figure 7). Planned ANOVAs at each time region revealed a significant effect of expression at regions 3–5 (Table 5). The interaction of partner and expression was not significant at any time region<sup>7</sup>, with the exception of region 5 which showed a partner  $\times$  expression interaction that was significant in the items analysis only. At this region, pairwise comparisons showed that the effect of expression was somewhat stronger for original partners,  $t(47)=3.41, p<.01$ ;  $t(15)=4.21, p<.01$ , compared to new partners,  $t(47)=1.97, p=.06$ ;  $t(15)=1.84, p=.09$ , possibly due to delayed target fixations with new speakers and new expressions.

Finally, an analysis of the very early time region (180–300 ms) which showed a significant partner-specific effect for broken precedents in Experiment 1a revealed no significant effects,  $F_s<.5$ .

## Conclusions

The results of Experiment 1b provide compelling evidence that a critical component of the partner-specific interpretation pattern observed in Experiment 1a was the interactivity of the situation. The fact that Experiment 1b participants showed a significant effect of expression in both the target fixation latency and the timecourse analyses demonstrates that they succeeded in learning the expression-image mappings. The fact that the expression effect emerged slightly later in this experiment (region 3) may be due to participants not learning the precedents as well as participants in the interactive experiments. What participants did not appear to

<sup>6</sup>A supplementary analysis of target fixation latencies that included experiment (1a vs. 1b) as a factor revealed a 3-way interaction (experiment  $\times$  partner  $\times$  expression) that was significant by items,  $F(1,94)=2.75, p=.10$ ,  $F(1,30)=6.35, p<.05$ ,  $\min F'(1,122)=1.92, p=.17$ .

<sup>7</sup>A supplementary analysis at region 2 (the region that contained the critical partner  $\times$  expression interaction in Experiment 1a) revealed a 3-way experiment  $\times$  partner  $\times$  expression interaction that was significant by items,  $F(1,94)=3.79, p=.055$ ,  $F(1,30)=4.34, p<.05$ ,  $\min F'(1,97)=2.02, p=.16$ .

appreciate was the association between particular speakers and particular referring precedents. In Experiment 1a, face-to-face interaction with a live speaker provided participants with the opportunity to collaboratively establish referring precedents and associate these precedents with individuals who worked with them on the task. Participants in Experiment 1b did not have an individual face to associate precedents with, thus they may have formed weaker partner-to-expression associations, if they were formed at all. Weak associations may explain the very late and weak partner  $\times$  expression interaction in the final analysis region (1400–1800ms).

Experiment 2 returns to the use of an interactive setting to explore an alternative interpretation of the partner-specific interpretation pattern observed in Experiment 1a. A secondary goal of Experiment 2 is to replicate experiment 1a's previously unobserved partner-specific effect for maintained expressions using a paradigm which provided more control over the experimental stimuli.

## Experiment 2

One feature of the design of Experiment 1a was that each picture in the scene had been described prior to the critical trial, which was comparable to the design used by Metzing and Brennan (2003). However, Kronmüller and Barr (2007) argued that on broken-precedent trials, this may cue participants to use partner-specific information, since it is immediately obvious that the original partner has broken a precedent. In their experiment, which did not find early partner-specific effects, the critical scenes included an unmentioned image which provided a plausible alternative referent on broken-precedent trials. So, when the original speaker used a broken precedent, it was plausible that the speaker was referring to the new image, thus the new term did not immediately signal that a precedent had been broken.

Given this argument, the partner-specific interpretation pattern observed in Experiment 1a may not be representative of typical language processing, because natural settings typically contain unmentioned entities which would prevent this cuing of the use of partner-specific information. The design of Experiment 2 controls for this possibility by including images in each scene which were not mentioned prior to the critical trial.

A second goal of Experiment 2 is to replicate the robust partner effect for maintained precedents in Experiment 1a. Unlike the broken-precedent effect, this effect has not been observed in any of the previous experiments that manipulated partner and expression. The lack of maintained precedent effects in previous work may be due, as I have argued, to the use of non-interactive settings (e.g., Barr & Keysar, 2002; Kronmüller & Barr, 2007), or a lack of power (e.g., Metzing & Brennan, 2003). However, another possibility is that the observed partner-specific effect for maintained precedents was due to early disambiguation in some conditions. The fact that Experiment 1b failed to replicate the partner-specific advantage for maintained precedents despite the fact that it used the same expressions as Experiment 1a suggests that this is not the case, however a replication of this effect using stimuli that are disambiguated relatively late would provide stronger evidence.

To have more control over the form of the critical expressions in Experiment 2, experimenters were prompted with the entire target noun phrase; these expressions and the associated scenes were designed to have a relatively late point-of-disambiguation. These design changes made it possible to test whether partner-specific information is used to constrain interpretation of longer-lasting temporary referential ambiguities.

## Predictions

Based on the results of Experiment 1a, it was predicted that addressees would identify the target referent faster when the original partner was speaking, compared to a new partner when

interpreting maintained precedents. If this partner-specific processing is part of the initial interpretation process, it should affect fixations made immediately after the onset of the critical expression, well before the point-of-disambiguation. Alternatively, if the partner-specific effect in Experiment 1a was due to early disambiguating information in some trials, the partner-specific effect for maintained precedents should be absent due to the extended temporary ambiguity.

If the early partner-specific broken-precedent effect in Experiment 1a occurred because participants were cued to take partner-specific information into account by scenes that did not contain unmentioned images, then the partner-specific effect for broken precedents should occur at a delay; well after an earlier precedent effect. Alternatively, if we do observe an early partner-specific effect for broken precedents, this would add to the evidence that addressees have speaker-specific expectations about how potential discourse referents might be referred to. With the original partner, addressees should not initially consider the target referent upon hearing a new expression, thus there should be few target fixations and many competitor fixations. In contrast, with a new partner and a new expression, addressees should consider the target and competitors equally, resulting in slightly more target fixations for new partners.

## Method

The method was similar to that used in Experiment 1a; only changes are noted here.

**Participants**—Thirty-two native English-speaking participants from the student community at the University of Illinois at Urbana-Champaign participated in this experiment in exchange for \$16 or partial course credit. An additional 2 participants were excluded due to experimenter error (1) and a poor calibration (1)<sup>8</sup>.

**Procedure**—The participant's eye movements were recorded with an Eyelink 1000 desktop-mounted eye-tracking system. On each trial, the participant and the experimenter saw the same 4 images on their respective computer screens, but in different arrangements (Figures 8a–b). The participant's task was to rearrange his or her images into the experimenter's order. The participant and experimenter viewed the same 4 images for 5 rounds in a row, and then moved onto the next set of images. As in Experiments 1a–b, the final round of matching (round 5 in this experiment) was the critical round during which the participant's interpretation of the critical expression was monitored. The task typically lasted 45 min–1 hour.

Participants rearranged 32 sets of 4 images according to the experimenter's instructions. Each set of images was rearranged 5 times in a row, for a total of 160 rounds. Unlike Experiments 1a–b, on rounds 1–4, only three of the four images were described by the experimenter; the remaining image remained unmentioned and the participant did not re-arrange this picture. On the fifth (critical) round, the experimenter named all 4 images. The image that remained unmentioned in rounds 1–4 was either the competitor image or an unrelated image, thus across all conditions, on round 5, the target image always had an established name.

This design feature is similar to the test trial structure used by Kronmüller and Barr (2007) who introduced a new, unmentioned image prior to the test trial. In the present study, the unmentioned image was present throughout all 5 rounds of matching, and was not a 'new' image, which, as Brennan and Hanna (2009) argue, might have attracted fixations, minimizing partner effects. Leaving one of the 4 images unnamed during rounds 1–4 provided a plausible reason for the experimenter to use a new name in the broken-precedent condition, thus making

<sup>8</sup>An initial 16 participants (not reported here) were run in a pilot version of the task during which the experimenters were learning the procedure and how to use the eye-tracker.

it not immediately obvious that the precedent had been broken (see Kronmüller & Barr, 2007).

## Materials

**Visual stimuli:** The visual stimuli were 128 hard to name images, half of which were used as target and competitor images. The set of images included those used in Experiments 1a–b and other, similar images. On each trial, two of the four images were black and white, two were colorized. The target and competitor pair of images was either the black and white pair or the colorized pair. As in Experiments 1a–b, the target and competitor images were selected such that they could be described using a common term (e.g. *crushed*).

**Auditory stimuli:** On rounds 1–4, the experimenter described the target image and two other images; either the competitor and an unrelated image, or the two unrelated images. Experimenters came up with names for the unrelated images on the fly. The name for the target image and competitor image, if named, were provided in a prompt on the experimenter's computer screen. Unlike Experiment 1a, in this experiment the prompt included a full image description. Experimenters were instructed to say this description, and then add extra descriptive words if necessary. As in Experiments 1a–b, for a given target, the critical description on the fifth trial was the same for maintained-precedent and broken-precedent trials. The expression variable (maintained vs. broken precedent) was manipulated by changing the target description on rounds 1–4. For a maintained precedent trial, the expression would be the same for rounds 1–5. For a broken precedent trial, the expression on rounds 1–4 would be different from that on round 5 (Table 6).

Providing the experimenter with the full description of the target and competitor images allowed systematic control of the point-of-disambiguation of the target expression. For example, the critical expression *the crushed up bit of?* is temporarily consistent with both the target referent (cookie crumbs) and the competitor (paper). In fact, target and competitor images were swapped across lists, and the critical expression to describe the paper when it was the target was *the crushed up bit of paper*.

**List rotations:** The experiment manipulated two within-participants variables, partner (same or different) and expression (maintained or broken), yielding four conditions. The 32 critical items were rotated through the 4 conditions across 8 lists, which also counterbalanced target and competitor images. Each list was presented in one of two random orders. Each participant was run on a single list. For half of the critical trials, the image that was left unmentioned during rounds 1–4 was the competitor; for the other half of critical trials, the image left unmentioned was one of the unrelated images; this variable was not systematically manipulated so it was not included in the analyses<sup>9</sup>.

## Analysis and Results

**Characteristics of Critical Instructions—**The experimenter's description of each image was transcribed. Trials on which the experimenter made an error or a disfluent repair were excluded from analysis (this eliminated 1% of critical trials).

A prosodic analysis on a subset (25%) of the critical (5<sup>th</sup> round) descriptions of target pictures was conducted to test for prosodic differences in the first three words of the critical instructions across the experimental conditions. Of the four experimenters who took part in this experiment, only two ran enough subjects and provided enough codeable trials to take part in this analysis.

<sup>9</sup>The basic pattern of results appeared similar regardless of which item was left out.

The analysis revealed no condition differences in pitch (maximum, minimum or average), intensity or duration.

As in Experiment 1a, the point-of-disambiguation occurred earlier for maintained precedents and when the original partner was speaking. There was also a significant expression  $\times$  partner interaction (see Appendix A). Because the entirety of the target expression was scripted, and there were no prosodic differences in the first few words, these differences likely resulted from differences in production rate towards the end of critical utterances.

For maintained precedents and original partners, the average point-of-disambiguation was 856 ms following the onset of the first content word in the critical expression (hereafter “adjective onset”), vs. 940ms for maintained expressions and new partners,  $t(31)=3.56$ ,  $p<.01$ ,  $t(31)=6.72$ ,  $p<.0001$ . For broken precedents, the average point-of-disambiguation was 973ms for original partners, compared to 938ms for new partners,  $t(31)=1.2$ ,  $p=.24$ ,  $t(31)=4.23$ ,  $p<.0001$ . Despite these differences, disambiguating information occurred quite late in all conditions, thus fixations driven by the disambiguating information would not be expected around 1056 ms at the earliest, corresponding to the fourth analysis region. This is well after the time range (200–600ms) in which early partner-specific effects would be expected to occur.

**Eye-tracking Analyses—**The participants’ interpretation of the critical referring expression was analyzed by examining the latency to fixate the target, as well as target advantage scores over time, defined in the same way as in Experiments 1a–b. In this experiment there was a significant effect of the type of critical image (black and white vs. colorized), however it did not affect the critical partner  $\times$  expression interaction, which is the focus of the analyses. For clarity, this factor is not included in the primary analyses; a summary of the image type effect is presented in Appendix B.

**Latency to fixate the target:** The latency of fixating the target was analyzed in an ANOVA with partner (same or different) and expression (broken or maintained precedent) as factors (Table 7). Target fixation latencies were much faster in this experiment compared to Experiment 1a, most likely due to the small number of images on the screen (Figure 9). A main effect of expression was due to shorter latencies to fixate the target when the referring precedent was maintained. Unlike Experiment 1a, the partner by expression interaction was not significant. To compare the results of this experiment with Metzing and Brennan (2003) and those of Experiment 1a, we conducted planned, pairwise comparisons to test for a partner-specific precedent effect. When precedents were maintained, target identification times were significantly faster when the original partner was speaking,  $t(31)=1.72$ ,  $p<.05$ ;  $t(31)=1.93$ ,  $p<.05$ , both one-tailed. There was no difference when precedents were broken,  $ts<.2$ .

The lack of a significant partner by expression interaction in this experiment may be due to the simpler display and the resulting faster target identification times. However, another difference from Experiment 1a is that Experiment 2 was twice as long, thus participants experienced 8 broken precedents in Experiment 2, compared to only 4 in Experiment 1a (by contrast, Metzing & Brennan, 2003 had only 2 broken precedent trials). Thus, addressees may have learned that their partner did not adhere to referring precedents, curtailing partner-specific interpretation processes. In fact, in a similar experiment with young children, Matthews, Lieven and Tomasello (2008) found a partner-specific effect for broken-precedents only on the first broken precedent trial; by the second trial, the children appeared to have learned that the speaker did not adhere to referring precedents. Consistent with this hypothesis, we observed a near-significant interaction between partner, expression, and 1<sup>st</sup> vs. 2<sup>nd</sup> half of the experiment,  $F(1,31)=3.71$ ,  $p=.06^{10}$ . For trials in the first half of the experiment, the partner  $\times$  expression interaction was significant,  $F(1,31)=4.72$ ,  $p<.05$ : When precedents were maintained, target identification was 161ms faster when the original partner was speaking compared to a new



partner (490ms [SE=25] vs. 651ms [SE=85], respectively),  $t(31)=1.77$ ,  $p<.05$ , one-tailed. In contrast, when precedents were broken, target identification was a non-significant 120ms *slower* when the original partner was speaking, compared to a new partner (864ms [SE=68] vs. 744ms [SE=57], respectively),  $t(31)=1.26$ ,  $p=.11$ , one-tailed. In contrast, the partner  $\times$  expression interaction did not approach significance in the second half of the experiment,  $F(1,6)$ .

**Target advantage scores**—Target advantage scores were analyzed in the same manner as Experiments 1a–b (Figure 10). Results of the ANOVA with partner and expression as factors are presented in Table 8.

**Before the point-of-disambiguation:** The first region to show significant condition effects was region 2, which captured fixations made in response to the onset of the first content word in the critical description, and well before the point-of-disambiguation (Figure 11). A main effect of partner at region 2 (marginal by items) was due to faster interpretation with the original partner. A main effect of expression was due to faster interpretation of maintained precedents. These effects were qualified by a significant partner  $\times$  expression interaction. Replicating the findings of Experiment 1a, addressees interpreted maintained precedents faster when spoken by the person who established the precedent, compared to a new speaker,  $t(31)=2.82$ ,  $p<.01$ ,  $t(31)=3.09$ ,  $p<.01$ . In contrast, there was no partner effect for broken precedents,  $t<.15$ .

However, the lack of a partner-specific effect for broken precedents may reflect the fact that participants learned that their partners did not adhere to referring precedents. Consistent with this hypothesis, at region 2, the partner  $\times$  expression interaction significantly interacted with experiment half (1<sup>st</sup> vs 2<sup>nd</sup>),  $F(1,31)=5.69$ ,  $p<.05$ . During the first-half trials, the partner $\times$ expression interaction was significant,  $F(1,31)=7.36$ ,  $p<.05$ : When precedents were maintained, target advantage scores were significantly higher when the original partner was speaking compared to a new partner (.232 [SE=.04] vs. .098 [SE=.05], respectively),  $t(31)=2.29$ ,  $p<.05$ . And, replicating the very early broken-precedent effect observed in Experiment 1, when precedents were broken, target advantage scores were significantly *lower* when the original partner was speaking, compared to a new partner (−.031 [SE=.04] vs. .076 [SE=.04], respectively),  $t(31)=1.81$ ,  $p<.05$ , one-tailed. In contrast, during the second half of the experiment, the partner  $\times$  expression interaction was not significant,  $F(1,31)=0.0$ ,  $p=.99$ . The partner  $\times$  expression  $\times$  half interaction did not approach significance at any of the remaining time regions.

At region 3, there was a main effect of expression, but no interaction with partner (Table 8), likely due to the late emergence of a maintained precedent benefit for new speakers beginning around 600ms post-stimulus. Thus, addressees accessed learned expression-referent associations later with a new partner compared to the original partner.

**Late effects:** Regions 4–5 captured fixations that were driven by the point-of-disambiguation. At region 4, the effect of expression remained significant. Similar to Experiment 1a, regions 4 and 5 both showed an effect of partner that was due to higher target advantage scores with the new speaker, possibly due to late identification of the target in these conditions, prompted by the point-of-disambiguation.

<sup>10</sup>By-items analyses are not possible because which half of the experiment a given item appeared in was not systematically manipulated, thus, across lists, some items occurred in only one half, others in both halves.

## Conclusion

Participants in this experiment interpreted their partner's temporarily ambiguous expressions with respect to referential precedents that were shared with that partner. These precedents, or conceptual pacts allowed addressees to resolve the ambiguity well before the point-of-disambiguation. Across the entire experiment, when a precedent for a potential referent existed, but it was not shared with the current speaker, participants were less likely to rely on this precedent. This result is consistent with the findings of Experiment 1a, and demonstrates that from the earliest moments of understanding, addressees used partner-specific information to guide interpretation of entrained terms.

Experiment 2 was twice as long as Experiment 1a, with 8 trials during which the original partner broke a referring precedent. Addressees were sensitive to these instances, and apparently learned *not* to expect the original partner to maintain precedents. In the first half of the experiment, in analyses of both target fixation latencies and early time-course analyses, we observed significant partner  $\times$  expression interactions. When expressions were maintained, addressees fixated the target significantly earlier, and were more likely to initially consider the target more than competitors with the original partner, compared to a new partner. When precedents were broken, from the earliest moments of understanding, addressees generated fewer target fixations and more competitor fixations when the original partner was speaking compared to a new partner. These early partner-specific effects demonstrate that an addressee's initial interpretation of an utterance takes into account partner-specific information about referring history. These effects emerged despite the fact that there was an unmentioned image in the scene, thus ruling out the alternative explanation of the broken-precedent effect in Experiment 1a that addressees were cued to take partner-specific information into account when a new term was used in contexts where every image had already been mentioned (see Kronmüller & Barr 2007).

By contrast, during the second half of the experiment, the partner  $\times$  expression interaction was not significant for either dependent variable. The fact that addressees learned to ignore precedents is consistent with the original conception of referential conceptual pacts as flexible agreements for how to refer (Brennan & Clark, 1996), which should be sensitive to speaker consistency and contextual variables that may prompt speakers to abandon precedents. When it became clear that the original speaker did not adhere to referring precedents, addressees no longer made assumptions about how he or she would refer.

## General Discussion

The results of the experiments presented here demonstrate that in interactive dialog settings, an addressee's initial interpretation of his or her partner's utterance is sensitive to the identity of the speaker and their experience with that speaker. These partner-specific effects emerged early in the process of interpretation, during the first critical analysis region (beginning 180–200ms after critical adjective onset). There was no evidence of an initial perspective-free processing stage (Kronmüller & Barr, 2007). In contrast, in non-interactive settings we failed to replicate the partner-specific interpretation pattern, suggesting that interpretation processes are qualitatively different in interactive compared to non-interactive settings.

The finding that addressees used partner-specific information to guide interpretation of temporarily ambiguous expressions is consistent with previous work showing partner-specific interpretation of broken precedents (Metzing & Brennan, 2003; Brennan & Hanna, 2009; Matthews, Lieven, & Tomasello, 2008), as well a large body of work showing that addressees take the perspective of their interlocutor into account as they interpret their utterances (Hanna, Tanenhaus, & Trueswell, 2003; Hanna & Tanenhaus, 2004; Brown-Schmidt, Gunlogson & Tanenhaus, 2008; Heller, Grodner & Tanenhaus, 2008). These results are also consistent with

work in language production showing that speakers are sensitive to the perspective of their partner (Haywood, Pickering, & Branigan, 2005; Lockridge & Brennan, 2002), and specifically whether that partner is familiar with previously entrained terms (Brennan & Clark, 1996; Horton, 2007; Issacs & Clark, 1987; Wilkes-Gibbs & Clark, 1992).

The mechanism underlying the observed partner-specific processing advantage for maintained precedents may be largely due to automatic, associative processes that link specific individuals, referential forms, and referents (Horton & Gerrig, 2005b; Horton, 2007; also see Duff, Hengst, Tranel, & Cohen, 2006). Whether the contextual information that guides referential processing is limited to three-way person-name-referent associations, or whether other contextual information, such as the presence of an eavesdropper, also plays a role, is a question for future work. The partner-specific penalty for interpretation of broken precedents may result from an added assumption that when an entrained term is mutually known<sup>11</sup>, the partner will continue to use it (see E. Clark, 1987; Metzing & Brennan, 2003). The fact that this effect was eliminated in the second half of Experiment 2 suggests that addressees are sensitive to whether speakers adhere to referring precedents when generating these assumptions. The earliness of the partner-specific effects for both maintained and broken precedents suggests that partner-specific information is continuously available to the language processing system, and that it does not require explicit or time-consuming efforts to be employed.

Finally, a central theme of this article is that insights into language processes that are inherently interactive will likely require investigations of language use in natural, interactive settings. An open question regarding the present findings is what aspect of the interaction led to the positive results. One feature of naturally produced speech is that speakers mark what information is familiar to them (Bard, et al., 2000). We saw evidence of this in the form of earlier points-of-disambiguation in conditions with original partners and maintained expressions. However, this is unlikely to be the locus of the early partner  $\times$  expression interactions, since participants in Experiment 1b did not show this effect, yet the expressions they listened to contained these features. Further, the critical partner  $\times$  expression interactions emerged before the earliest disambiguating information occurred.

A more likely possibility is that the live speakers in Experiments 1a and 2 were more salient and distinct than the pre-recorded voices in Experiment 1b (and in previous work using pre-recorded stimuli). When information sources are more distinct, memory for the source of that information improves (for a review, see Johnson, Hashtroudi, & Lindsay, 1993), thus addressees in the Experiments 1a and 2 may have had more success distinguishing the two speakers and generating accurate three-way associations between individual speakers, expressions, and referents. In contrast, participants in Experiment 1b may have blended the representations of the two speakers, or assumed that both speakers shared the addressee's own perspective. The opportunity to interactively establish the referring precedents may also have played a role, as information is thought to enter common ground only when established through interactive grounding processes (Clark & Brennan, 1991; Clark & Schaefer, 1989), thus addressees may not assume precedents are part of common ground in the absence of an interaction.

What do these results imply for future work on partner-specific language processes and language processing work in general? Live, face-to-face interaction limits experimental control and adds associated challenges to running an experiment, and for some experimental questions, interactive paradigms may be intractable. On the other hand, true interaction may be a necessary

<sup>11</sup>Results by Shintel & Keysar (2007) suggest that entrained terms may reflect shared, rather than mutual knowledge (see Clark & Marshall, 1978), though a lack of a processing benefit for maintained precedents and stimulus differences between the conditions complicate interpretation of their results.

pre-condition to studying certain types of language processes, particularly those which involve inherently interactive representations. And in theory, because language is deeply rooted in face-to-face interaction, all aspects of language processing may be affected to some degree by the interaction itself. Perhaps a reasonable goal for future work would be to pair well-controlled, standard experiments with investigations of the same processes in natural settings, such as analyses of spoken corpora, or on-line conversational paradigms like the one used here.

Another possibility is to employ pseudo-interactive paradigms in which participants listen to pre-recorded utterances, but are led to believe that they are interacting with a live person in another room. However, previous work using this approach failed to find evidence of partner-specific interpretation patterns (e.g. Barr & Keysar, 2002, experiment 3). And in related work using similar paradigms, information about the speaker's perspective has not been integrated into the on-line interpretation of referring expressions (Barr, 2008b), whereas in other work using interactive paradigms, it has (Heller, Grodner, & Tanenhaus, 2008; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008). Thus, some pseudo-interactive paradigms may not provide enough of the genuine interactivity and subtle cues such as laughs, sighs, and prosodic information which are characteristic of a true, face-to-face conversation.

Finally, in the experiments presented here, addressees interpreted partially-scripted expressions produced by experimenters who had previous experience with the task. A similar approach was taken by Metzger and Brennan (2003), and Matthews, Lieven, and Tomasello (2008), who also used interactive settings and observed partner-specific effects. These interactive experimental settings may be comparable to teaching or instructional contexts in which the speaker is more familiar with the topic of conversation than the addressee. An open question is how partner-specific effects might change in a broader range of natural, spontaneous conversations outside of experimental contexts. The results of the experiments presented here suggest that the more natural and interactive the conversation is, the stronger partner-specific effects will be.

## References

- Arnold JE, Tanenhaus MK, Altmann RJ, Fagnano M. The old and thee, uh, new. *Psychological Science* 2004;15:578–582. [PubMed: 15327627]
- Arnold JE, Hudson Kam C, Tanenhaus MK. If you say thee uh- you're describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2007;33:914–930.
- Arnold JE, Eisenband JG, Brown-Schmidt S, Trueswell JC. The immediate use of gender information: eyetracking evidence of the time-course of pronoun resolution. *Cognition* 2000;76:B13–B26. [PubMed: 10822045]
- Bard EG, Anderson AH, Sotillo C, Aylett M, Doherty-Sneddon G, Newlands A. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language* 2000;42:1–22.
- Barr DJ, Keysar B. Anchoring comprehension in linguistic precedents. *Journal of Memory and Language* 2002;46:391–418.
- Barr DJ. Analyzing 'visual world' eye-tracking data using multilevel logistic regression. *Journal of Memory and Language* 2008a;59:457–474.
- Barr DJ. Pragmatic expectations at linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition* 2008b;109:18–40. [PubMed: 18760407]
- Brennan SE, Clark HH. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1996;22:1482–1493.
- Brennan SE, Hanna JE. Partner-specific adaptation in Dialog. *Topics in Cognitive Science* 2009;1:274–291.
- Bromme R, Jucks R, Wagner T. How to Refer to 'Diabetes'? Language in Online Health Advice. *Applied Cognitive Psychology* 2005;19:569–586.

- Brown-Schmidt S, Gunlogson C, Tanenhaus MK. Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition* 2008;107:1122–1134. [PubMed: 18164700]
- Chambers CG, Tanenhaus MK, Eberhard KM, Filip H, Carlson GN. Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language* 2002;47:30–49.
- Clark, EV. The principle of contrast: A constraint on language acquisition. In: MacWhinney, B., editor. *Mechanisms of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum; 1987.
- Clark, HH.; Brennan, SE. Grounding in communication. In: Resnick, LB.; Levine, J.; Teasley, SD., editors. *Perspectives on Socially Shared Cognition*. APA; 1991.
- Clark, HH.; Marshall, CR. Reference diaries. In: Waltz, DL., editor. *TINLAP-2: Theoretical issues in natural language processing-2*. New York: Association for Computing Machinery; 1978. p. 57-63.
- Clark, HH.; Marshall, CR. Definite reference and mutual knowledge. In: Joshi, AK.; Webber, BL.; Sag, IA., editors. *Elements of Discourse Understanding*. Cambridge University Press; 1981.
- Clark, HH. *Arenas of language use*. Chicago: University of Chicago Press; 1992.
- Clark, HH. *Using language*. Cambridge, UK: Cambridge University Press; 1996.
- Clark HH, Wilkes-Gibbs D. Referring as a collaborative process. *Cognition* 1986;22:1–39. [PubMed: 3709088]
- Clark HH, Krych MA. Speaking while monitoring addressees for understanding. *Journal of Memory and Language* 2004;50:62–81.
- Clark HH, Schaefer EF. Contributing to Discourse. *Cognitive Science* 1989;13:259–294.
- Duff MC, Hengst J, Tranel D, Cohen NJ. Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience* 2006;9:140–146.
- Eberhard KM, Spivey-Knowlton MJ, Sedivy JC, Tanenhaus MK. Eye-movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research* 1995;24:409–436. [PubMed: 8531168]
- Garrod S, Anderson A. Saying what you mean in dialog: A study in conceptual and semantic co-ordination. *Cognition* 1987;27:181–218. [PubMed: 3691025]
- Grodner, D.; Sedivy, J. The effect of speaker-specific information on pragmatic inferences. In: Pearlmuter, N.; Gibson, E., editors. *The Processing and Acquisition of Reference*. Cambridge, MA: MIT Press; (in press)
- Hallett, PE. Eye movements. In: Boff, KR.; Kaufman, L.; Thomas, JP., editors. *Handbook of perception and human performance*. New York: Wiley; 1986. p. 10.1-10.112.
- Hanna JE, Tanenhaus MK. Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science* 2004;28:105–115.
- Hanna JE, Tanenhaus MK, Trueswell JC. The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language* 2003;49:43–61.
- Haywood SL, Pickering MJ, Branigan HP. Do speakers avoid ambiguities during dialog? *Psychological Science* 2005;16:362–366. [PubMed: 15869694]
- Heller D, Grodner D, Tanenhaus MK. The role of perspective in identifying domains of reference. *Cognition* 2008;108:831–836. [PubMed: 18586232]
- Horton WS. The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes* 2007;22:1114–1139. [PubMed: 18584063]
- Horton WS, Gerrig RJ. The impact of memory demands on audience design during language production. *Cognition* 2005a;96:127–142. [PubMed: 15925573]
- Horton WS, Gerrig RJ. Conversational common ground and memory processes in language production. *Discourse Processes* 2005b;40:1–35.
- Issacs EA, Clark HH. References in conversation between experts and novices. *Journal of Experimental Psychology: General* 1987;116:26–37.
- Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. *Psychological Bulletin* 1993;114:3–28. [PubMed: 8346328]
- Keysar B, Barr DJ, Balin JA, Paek TS. Definite Reference and Mutual Knowledge: Process Models of Common Ground in Comprehension. *Journal of Memory and Language* 1998;39:1–20.



- Keysar B, Barr DJ, Balin JA, Brauner JS. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science* 2000;11:32–38. [PubMed: 11228840]
- Krauss RM, Weinheimer S. Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology* 1966;4:343–346. [PubMed: 5969163]
- Kraut RE, Lewis SH, Swezey LW. Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology* 1982;43:718–731.
- Kronmüller E, Barr DJ. Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language* 2007;56:436–455.
- Lockridge CB, Brennan SE. Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review* 2002;9:550–557. [PubMed: 12412896]
- MacDonald MC, Pearlmutter NJ, Seidenberg MS. Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review* 1994;101:676–703. [PubMed: 7984711]
- Markman AB, Makin VS. Referential communication and category acquisition. *Journal of Experimental Psychology: General* 1998;127:331–354. [PubMed: 9857492]
- Matthews, DE.; Lieven, EVM.; Tomasello, M. What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Proceedings of the LONDIAL workshop on the semantics and pragmatics of dialog (SEMDIAL)*; London, UK. 2008 Jun.
- Metzing C, Brennan SE. When conceptual pacts are broken: partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language* 2003;49:201–213.
- Nadig AS, Sedivy JC. Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science* 2002;13:329–336. [PubMed: 12137135]
- Powell RS, O'Neal EC. Communication Feedback and Duration as Determinants of Accuracy, Confidence, and Differentiation in Interpersonal Perception. *Journal of Personality and Social Psychology* 1976;34:746–756.
- Schober MF, Clark HH. Understanding by Addressees and Overhearers. *Cognitive Psychology* 1989;21:211–232.
- Shintel H, Keysar B. You said it before and You'll say it again: Expectations of Consistency in Communication. *Journal of Experimental Psychology: Learning, Memory and Cognition* 2007;33:357–369.
- Tanenhaus, MK.; Trueswell, JC. Sentence comprehension. In: Miller, J.; Eimas, P., editors. *Handbook of cognition and perception*. San Diego, CA: Academic Press; 1995.
- Traum, DR. Computational Models of Grounding in Collaborative Systems; Working notes of AAAI Fall Symposium on Psychological Models of Communication; 1999 Nov. p. 124-131.
- Wilkes-Gibbs D, Clark HH. Coordinating Beliefs in Conversation. *Journal of Memory and Language* 1992;31:183–194.

## Acknowledgments

Thank you to Jennifer E. Arnold for providing her stimuli which were used in both experiments. Special thanks to three anonymous reviewers for numerous helpful comments on an earlier draft. The author was supported in part by NIH training grant T32 MH19990-07 to J. Kathryn Bock.

## Appendix A

### Analyses of the point-of-disambiguation

| <i>Critical noun onsets in Experiment 1a.</i> |              |           |            |           |           |              |
|---|--------------|-----------|------------|-----------|-----------|--------------|
| Source of variance                            | Participants |           | Items      |           | MinF'     |              |
|   | <i>MSe</i>   | <i>F1</i> | <i>MSe</i> | <i>F2</i> | <i>df</i> | <i>MinF'</i> |
| Partner                                       | 89562        | 3.64      | 11790      | 8.97*     | 1,61      | 2.59         |
| Expression                                    | 65232        | 42.26*    | 36008      | 24.32*    | 1,34      | 15.44*       |
| P × E   | 56121        | .77       | 33900      | 1.00      | 1,52      | .44          |

| <i>Point-of-disambiguation in Experiment 2.</i> |              |           |            |           |           |              |
|---|--------------|-----------|------------|-----------|-----------|--------------|
| Source of variance                              | Participants |           | Items      |           | MinF'     |              |
|   | <i>MSe</i>   | <i>F1</i> | <i>MSe</i> | <i>F2</i> | <i>df</i> | <i>MinF'</i> |
| Partner   | 14718        | 1.33      | 2299       | 9.31*     | 1,40      | 1.16         |
| Expression                                      | 2630         | 39.79*    | 1755       | 57.73*    | 1,60      | 23.55*       |
| P×E   | 8002         | 14.32*    | 1432       | 82.06*    | 1,41      | 12.19*       |

Note.  $df_1=1,47$ ,  $df_2=1,15$ .

\*  $p<.05$

Note.  $df_1=1,31$ ,  $df_2=1,31$ .

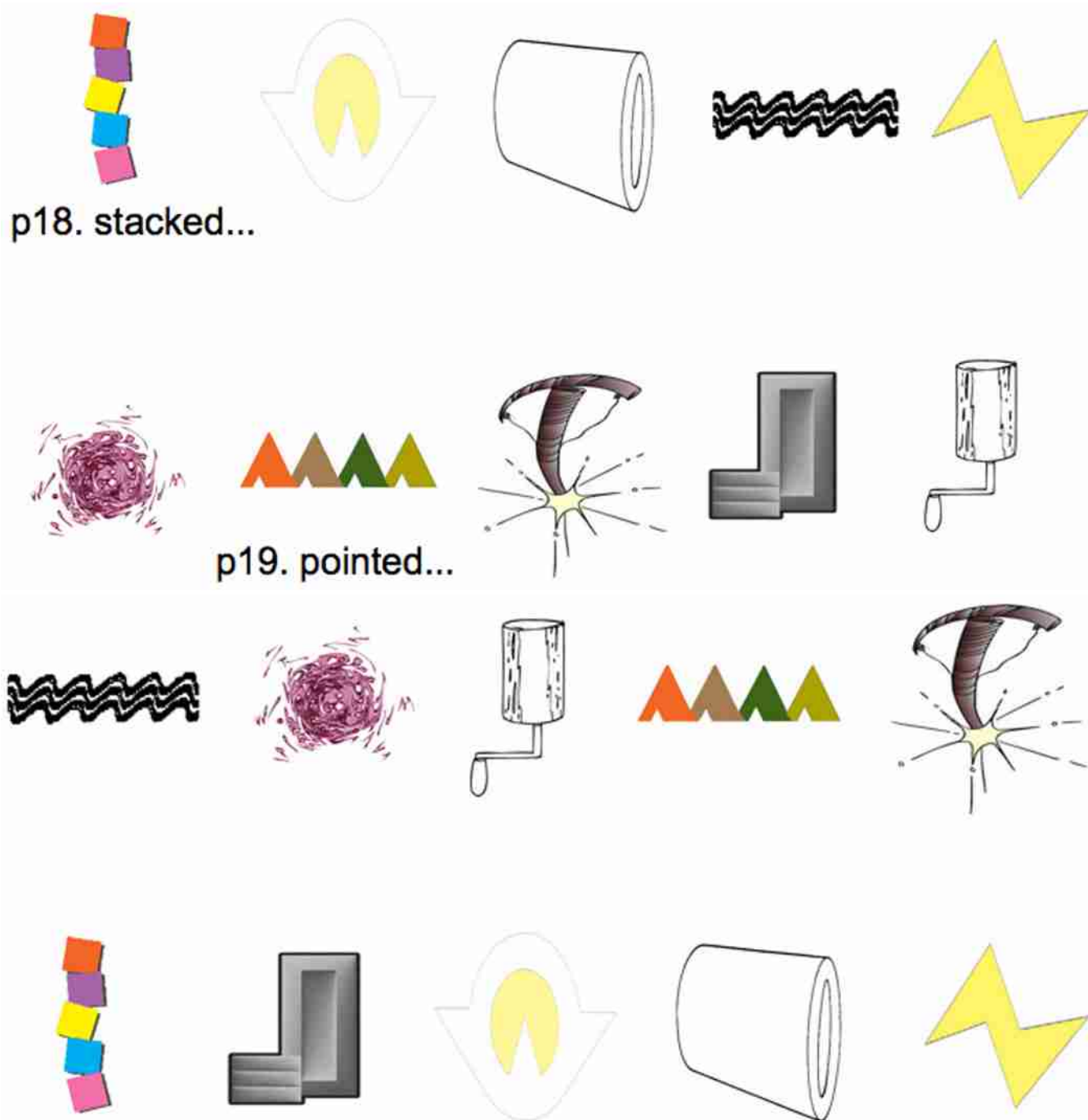
\*  $p<.05$

## Appendix B

### Experiment 2. Analyses including the picture factor

An analysis of the latency to fixate the target which included picture type as a factor revealed an effect of image type, due to faster target fixations for colorized images, suggesting that the mapping between adjective and image was faster or better learned for colorized targets ( $F_s>6$ ). A partner × image type interaction ( $F_s>4$ ) was due to a marginal effect of partner for colorized targets, with faster interpretation when the original partner was speaking,  $F1(1,31)=3.59$ ,  $p=.07$ ,  $F2(1,15)=3.76$ ,  $p=.07$ . For black and white targets, there was no main effect of partner,  $F_s<1$ .

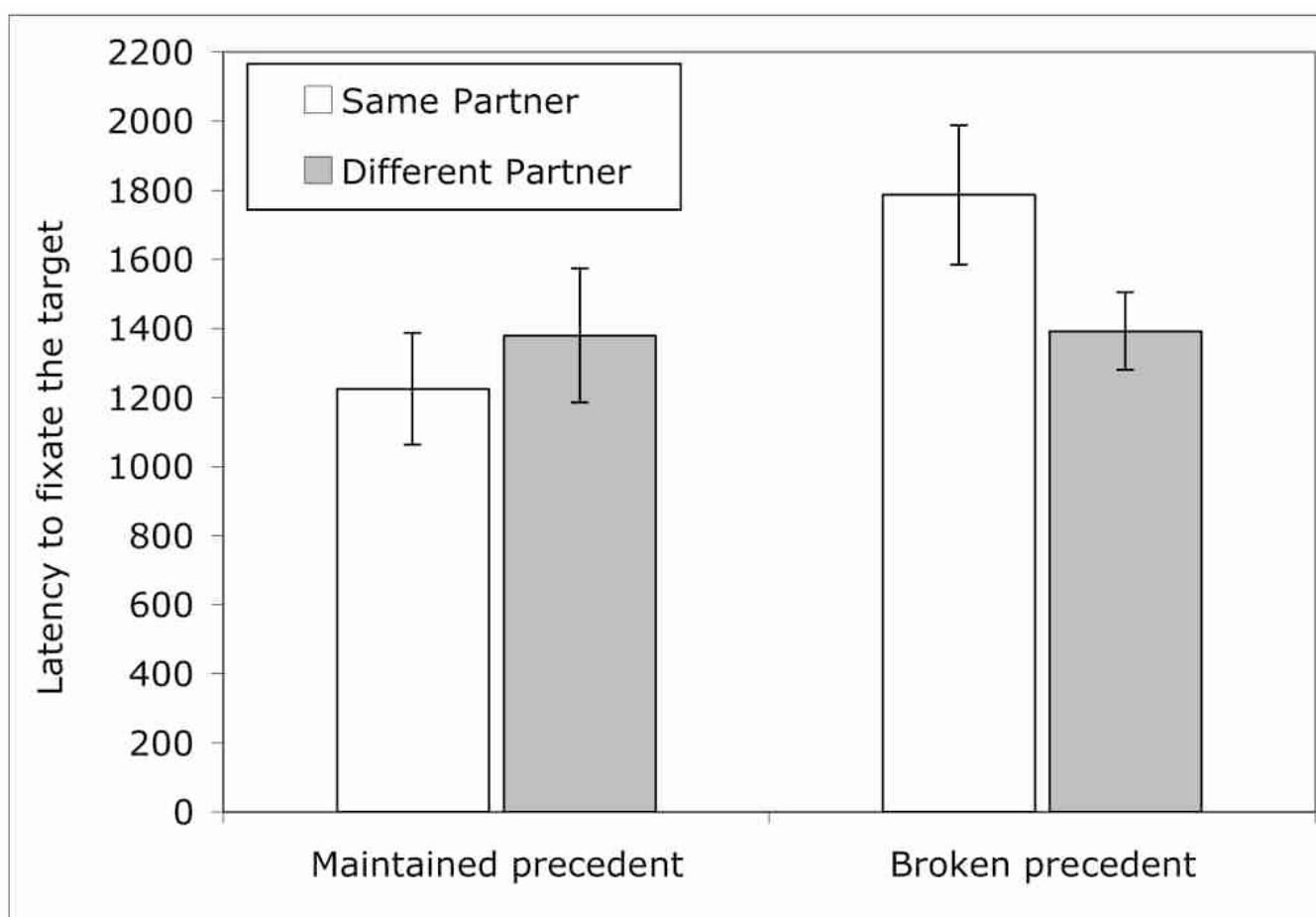
Analysis of target advantage scores including picture type as a factor revealed a significant effect of image type at region 2, due to higher target advantage scores for colorized targets. A significant partner × type interaction at region 5 was due to higher target advantage scores for new compared to old speakers with colorized targets  $F1(31)=11.58$ ,  $p<.01$ ,  $F2(15)=14.10$ ,  $p<.01$ . There was no partner effect for black and white targets,  $F_s<.2$ .



**Figure 1.**  
a–b. Experiment 1a, example screens for the experimenter (Fig. 1a) and participant (Fig. 1b) on the fourth (critical) round of matching. The experimenter's screen shows the critical adjectives for the target (top left) and competitor pictures.



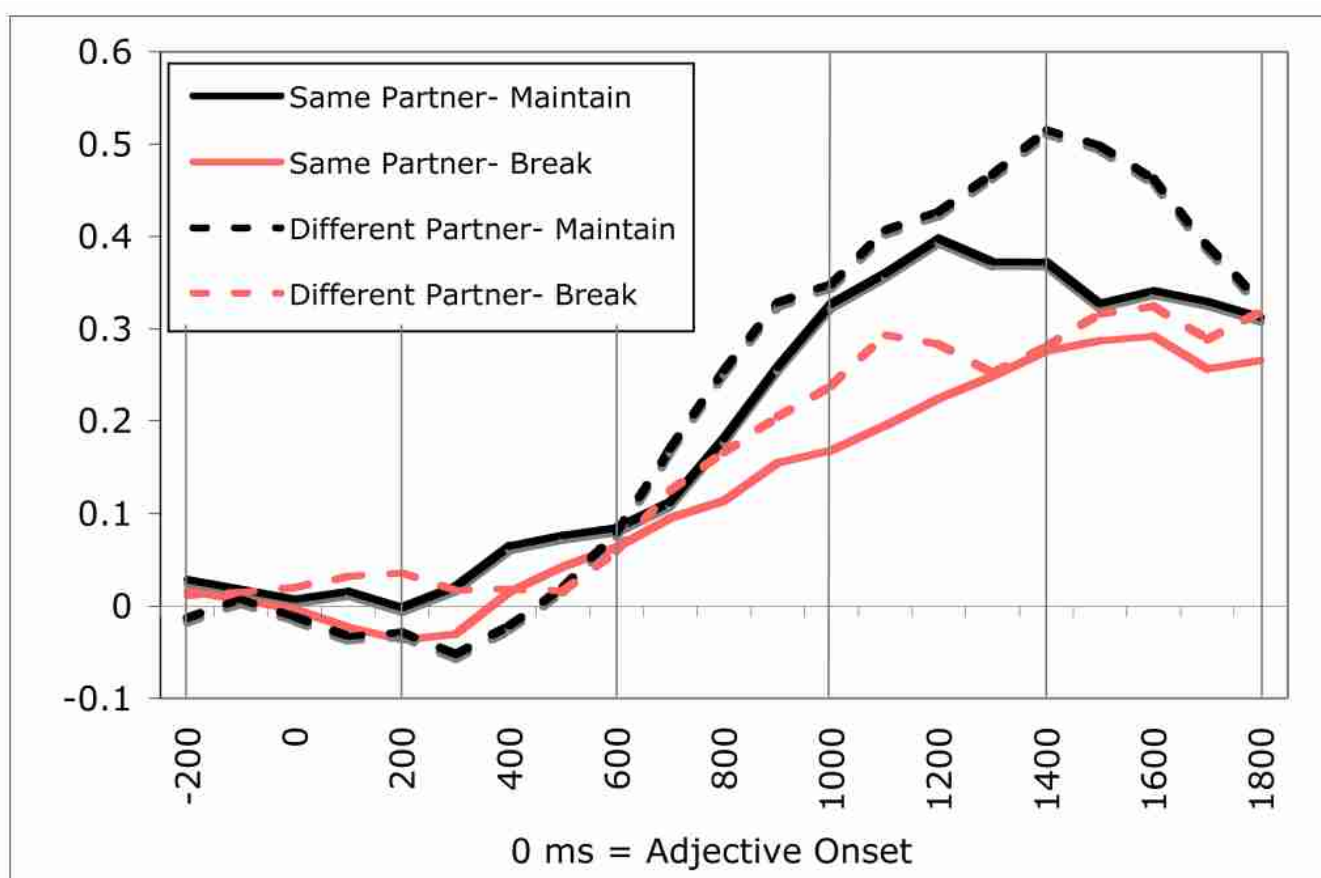
NIH-PA Author Manuscript



**Figure 3.**

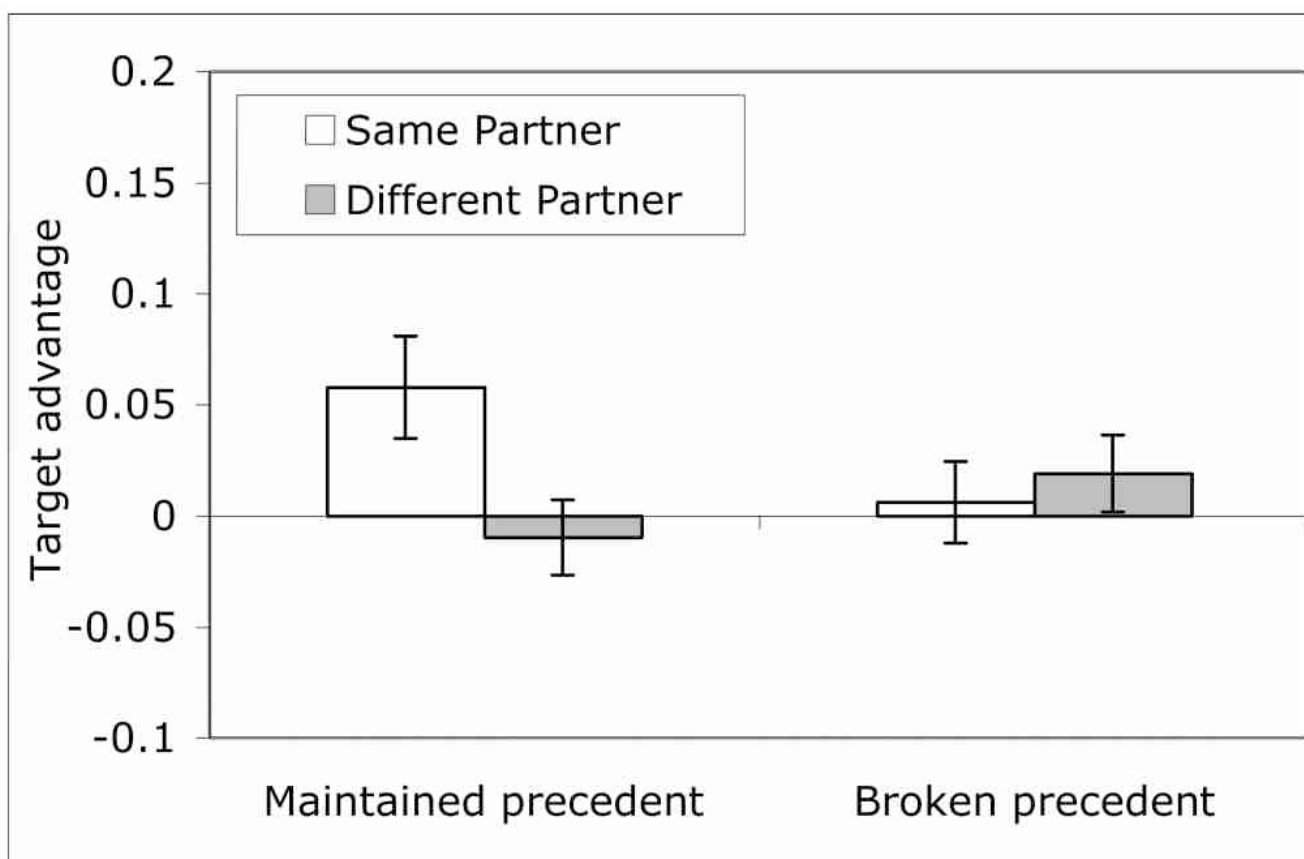
Experiment 1a, latency of first target look, for fixations beginning 200 ms after adjective onset. Error bars indicate standard error of the mean.





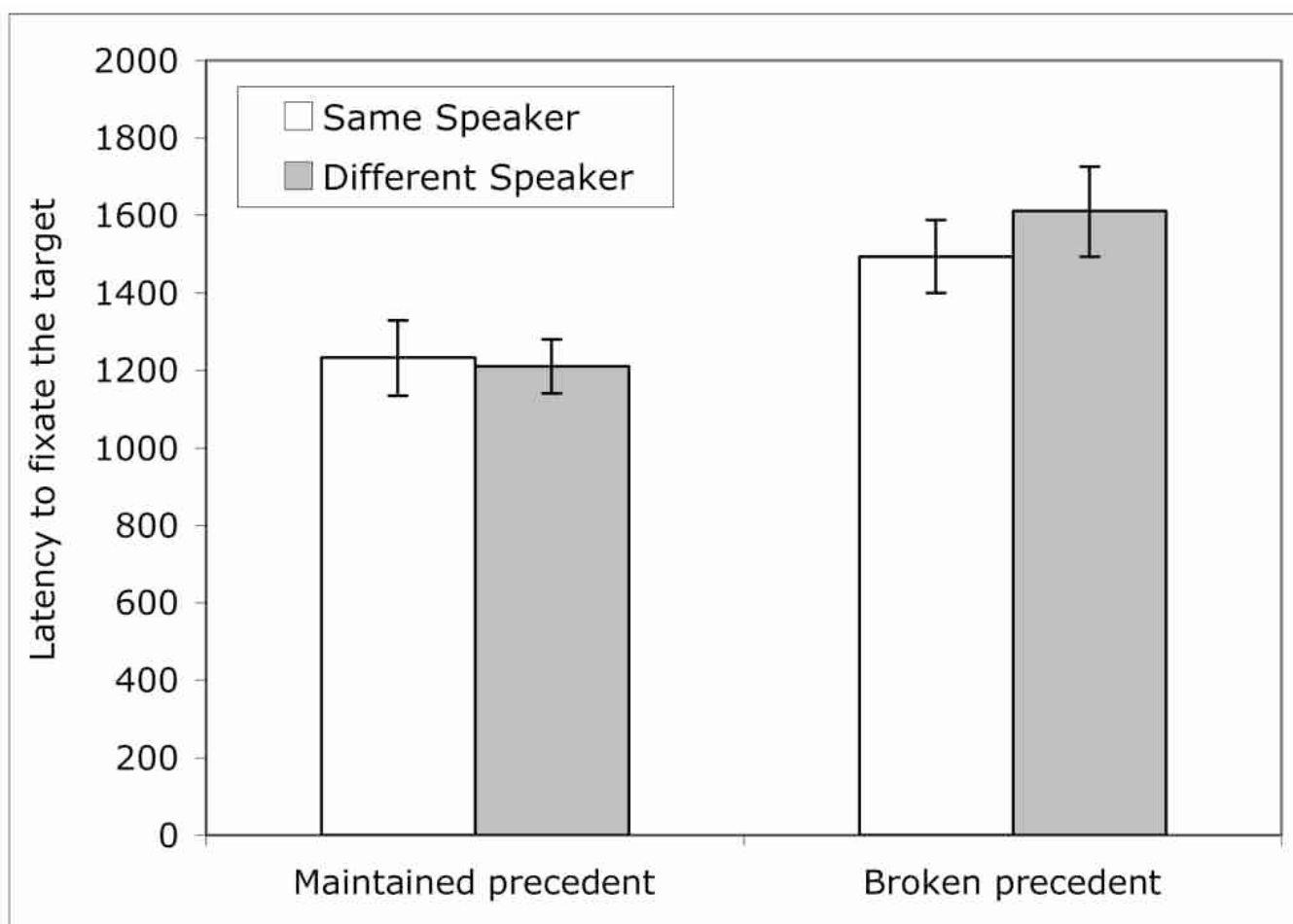
**Figure 4.**

Experiment 1a, target advantage scores over time. 0ms = Critical adjective onset. Vertical lines indicate analysis regions.

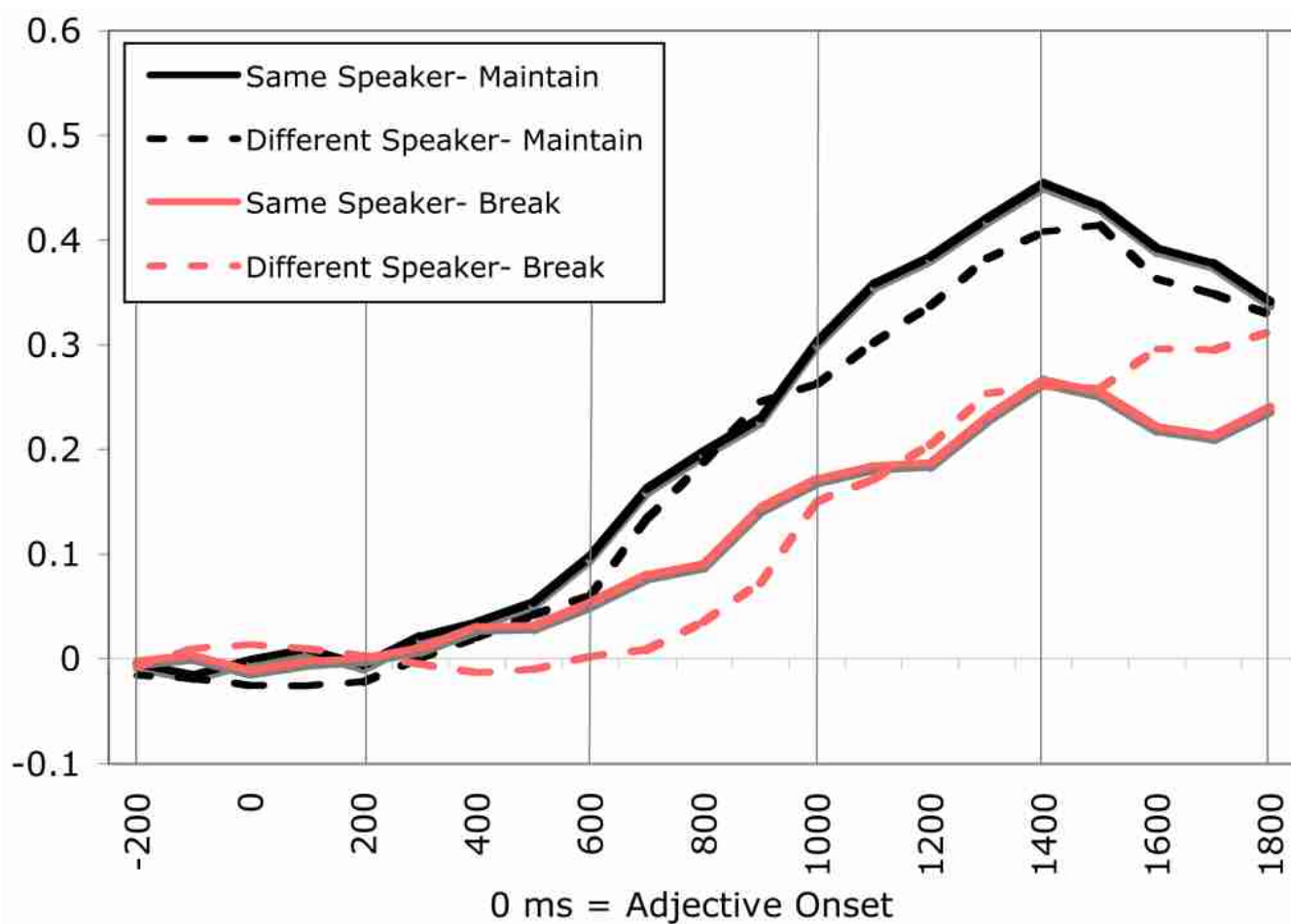


**Figure 5.**

Experiment 1a, average target advantage scores at region 2 (200–600ms post-stimulus). Error bars indicate standard error of the mean.



**Figure 6.** Experiment 1b, latency of first target look, for fixations beginning 200 ms after adjective onset. Error bars indicate standard error of the mean.



**Figure 7.** Experiment 1b, target advantage scores over time. 0ms = Critical adjective onset. Vertical lines indicate analysis regions.

## Figure8a



*The crushed up bit of  
cookie crumbs*



[describe]



*The sheet of paper  
that's crinkled*



[describe]

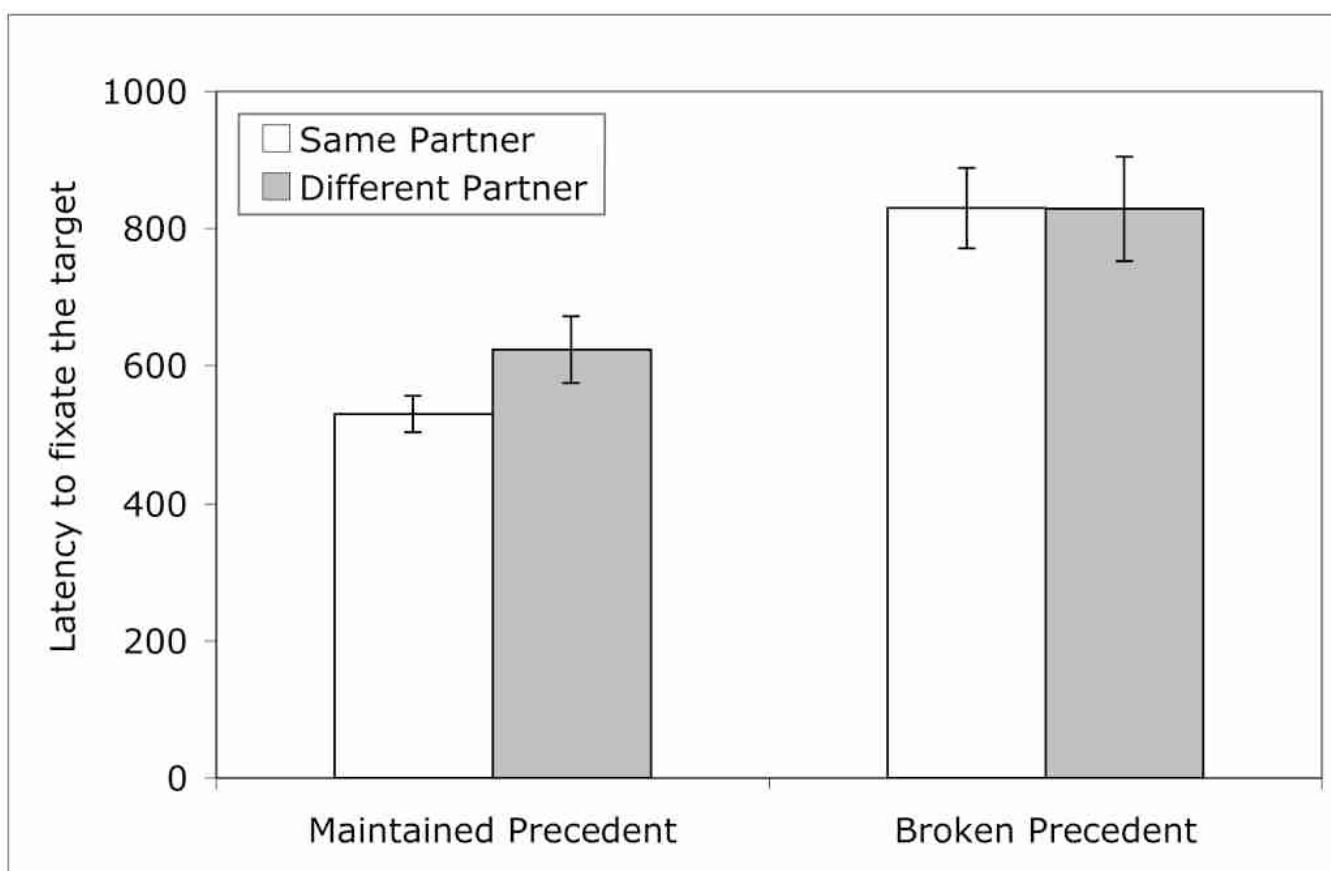


## Figure 8b

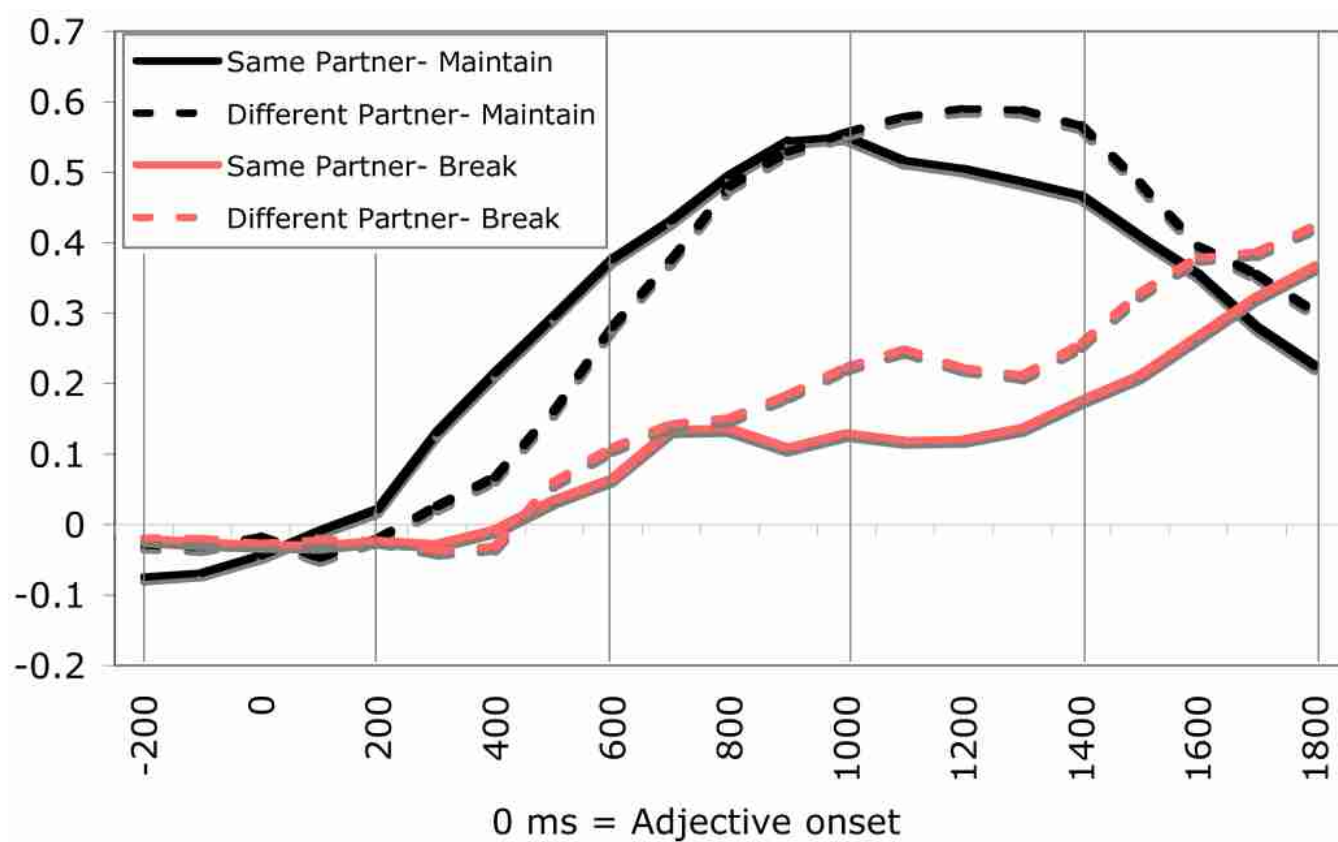


**Figure 8.**

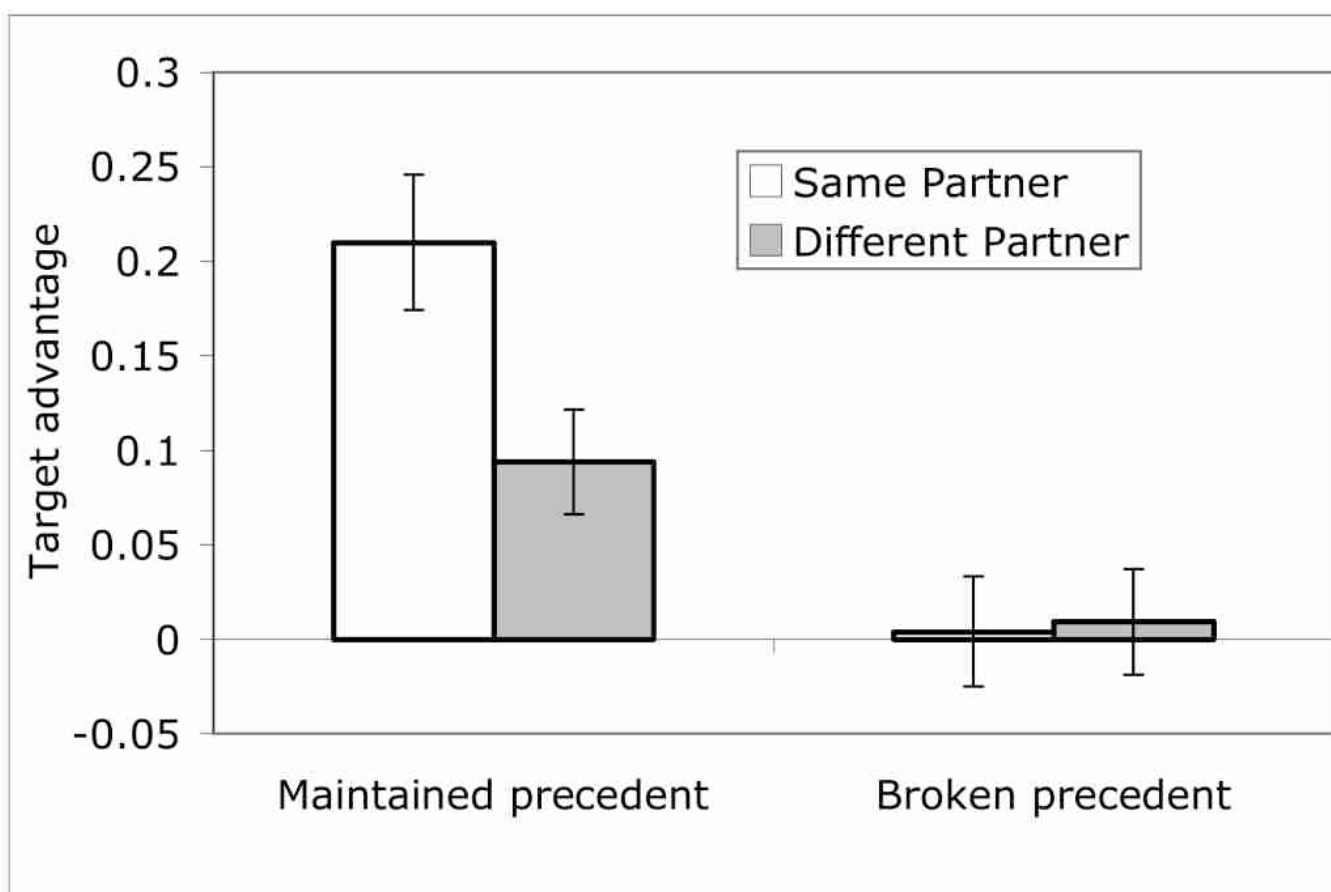
Experiment 2, example screens for the experimenter (Fig. 8a) and participant (Fig. 8b) on the 5<sup>th</sup> (critical) round of matching. The experimenter's screen shows the order in which the pictures should be described (top to bottom) and the descriptions which should be used for the target and competitor images.



**Figure 9.** Experiment 2, latency of first target look, for fixations beginning 200 ms after adjective onset. Error bars indicate standard error of the mean.



**Figure 10.** Experiment 2, target advantage scores over time. 0ms = Adjective onset. Vertical lines indicate analysis regions.



**Figure 11.**

Experiment 2, average target advantage scores at region 2 (200–600 ms post-stimulus). Error bars indicate standard error of the mean.

**Table 1**

Experiment 1a, example target and competitor descriptions.

| Maintained precedent condition |  |   |
|--------------------------------|--|---|
| Round                          | Target description                           | Competitor description                  |
| 1                              | the <u>multicolored</u> pile of kiddy blocks | the pointed teepees all in a row        |
| 2                              | the <u>multicolored</u> kiddy blocks         | the <u>pointed</u> row of teepees       |
| 3                              | the <u>multicolored</u> blocks               | the <u>pointed</u> teepees              |
| 4                              | the <u>multicolored</u> blocks               | the <u>pointed</u> teepees              |
| Broken precedent condition     |  |   |
| 1                              | the <u>stacked</u> cubes about to fall over  | the <u>pointed</u> teepees all in a row |
| 2                              | the <u>stacked</u> cubes falling over        | the <u>pointed</u> row of teepees       |
| 3                              | the <u>stacked</u> cubes                     | the <u>pointed</u> teepees              |
| 4                              | the <u>multicolored</u> blocks               | the <u>pointed</u> teepees              |

Note. In this example, the target image was the blocks, and the competitor was the teepees. The critical descriptive word (prompted on the Experimenter's screen) is underlined.

Table 2

Mean square error (*MSe*) and *F* tests for latency to fixate the target in Experiment 1a.

| Source of variance           | Participants |           | Items      |           | <i>MinF'</i> |              |
|------------------------------|--------------|-----------|------------|-----------|--------------|--------------|
|                              | <i>MSe</i>   | <i>F1</i> | <i>MSe</i> | <i>F2</i> | <i>df</i>    | <i>MinF'</i> |
| Partner<br>Expression<br>P×E | 1158107      | .60       | 307249     | .84       | 1,53         | .35          |
|                              | 914838       | 4.31*     | 405536     | 2.50      | 1,34         | 1.58         |
|                              | 1791065      | 2.02      | 302990     | 5.89*     | 1,62         | 1.50         |

Note. *df*<sub>1</sub>=1,47; *df*<sub>2</sub>=1,15.

\* *p*<.05



**Table 3**

Mean square error (*MSe*) and *F* tests for target advantage scores in Experiment 1a.

| Time region              | Participants |      |        | Items |        |      | MinF'  |  |
|--------------------------|--------------|------|--------|-------|--------|------|--------|--|
|                          |              |      |        |       |        |      |        |  |
| Source of variance       |              | MSe  | F1     | MSe   | F2     | df   | MinF'  |  |
| Region 1: -200 to 200ms  | Partner      | .023 | .013   | .005  | .035   | 1,62 | .01    |  |
|                          | Expression   | .016 | .071   | .011  | .132   | 1,58 | .05    |  |
|                          | P × E        | .018 | 1.96   | .005  | 3.00   | 1,58 | 1.19   |  |
|                          |              |      |        |       |        |      |        |  |
| Region 2: 200 to 600ms   | Partner      | .023 | 1.53   | .006  | 1.26   | 1,41 | .69    |  |
|                          | Expression   | .014 | .436   | .010  | .011   | 1,16 | .01    |  |
|                          | P × E        | .015 | 5.01*  | .002  | 10.80* | 1,60 | 3.42~  |  |
|                          |              |      |        |       |        |      |        |  |
| Region 3: 600 to 1000ms  | Partner      | .035 | 1.60   | .015  | 2.55   | 1,56 | .98    |  |
|                          | Expression   | .053 | 5.52*  | .013  | 5.41*  | 1,45 | 2.73   |  |
|                          | P × E        | .038 | .21    | .017  | .097   | 1,30 | .07    |  |
|                          |              |      |        |       |        |      |        |  |
| Region 4: 1000 to 1400ms | Partner      | .066 | 2.96~  | .021  | 3.14~  | 1,47 | 1.52   |  |
|                          | Expression   | .056 | 22.42* | .017  | 21.26* | 1,44 | 10.91* |  |
|                          | P × E        | .051 | .013   | .010  | .003   | 1,22 | .00    |  |
|                          |              |      |        |       |        |      |        |  |
| Region 5: 1400 to 1800ms | Partner      | .073 | 3.21~  | .012  | 7.82*  | 1,61 | 2.28   |  |
|                          | Expression   | .069 | 6.57*  | .022  | 6.03*  | 1,43 | 3.14~  |  |
|                          | P × E        | .043 | 2.67   | .023  | .66    | 1,23 | .53    |  |
|                          |              |      |        |       |        |      |        |  |

Note. Region 1 is used as a baseline. *df*<sub>1</sub>=1,47; *df*<sub>2</sub>=1,15.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$d \leq 10$   
 $d \leq 50$   
\*

**Table 4**  
Mean square error (*MSe*) and *F* tests for latency to fixate the target in Experiment 1b.

| Source of variance | Participants |           | Items      |           | <i>MinF'</i> |              |
|--------------------|--------------|-----------|------------|-----------|--------------|--------------|
|                    | <i>MSe</i>   | <i>F1</i> | <i>MSe</i> | <i>F2</i> | <i>df</i>    | <i>MinF'</i> |
| Partner            | 456305       | .23       | 75026      | .65       | 1,62         | .17          |
| Expression         | 481625       | 10.90*    | 126652     | 14.57*    | 1,52         | 6.24*        |
| P × E              | 265235       | .85       | 98583      | .69       | 1,41         | .38          |

Note. *df*<sub>1</sub>=1,47; *df*<sub>2</sub>=1,15.

\* *p*<.05

**Table 5**

Mean square error (*MSe*) and *F* tests for target advantage scores in Experiment 1b.

| Time region              |  | Participants                   |                        |                      | Items                  |                      |                      | MinF' |  |  |
|--------------------------|--|--------------------------------|------------------------|----------------------|------------------------|----------------------|----------------------|-------|--|--|
| Source of variance       |  | MSe                            | F1                     | MSe                  | F2                     | df                   | MinF'                |       |  |  |
| Region 1: -200 to 200ms  |  | Partner<br>Expression<br>P × E | .10<br>1.04<br>1.20    | .005<br>.002<br>.003 | .09<br>2.51<br>1.27    | 1,43<br>1,43<br>1,61 | .05<br>.74<br>.62    |       |  |  |
| Region 2: 200 to 600ms   |  | Partner<br>Expression<br>P × E | 1.16<br>1.29<br>.23    | .003<br>.006<br>.005 | 2.59<br>1.11<br>.22    | 1,61<br>1,42<br>1,44 | .80<br>.60<br>.11    |       |  |  |
| Region 3: 600 to 1000ms  |  | Partner<br>Expression<br>P × E | 1.89<br>15.11*<br>1.00 | .008<br>.013<br>.007 | 3.23<br>15.72*<br>1.35 | 1,57<br>1,46<br>1,52 | 1.19<br>7.70*<br>.57 |       |  |  |
| Region 4: 1000 to 1400ms |  | Partner<br>Expression<br>P × E | .44<br>20.86*<br>.76   | .012<br>.016<br>.018 | .51<br>24.16*<br>.70   | 1,49<br>1,49<br>1,44 | .24<br>11.19*<br>.36 |       |  |  |
| Region 5: 1400 to 1800ms |  | Partner<br>Expression<br>P × E | .10<br>10.23*<br>2.17  | .012<br>.027<br>.004 | .17<br>9.60*<br>5.07*  | 1,57<br>1,44<br>1,61 | .06<br>4.95*<br>1.52 |       |  |  |

Note. Region 1 is used as a baseline. *df*<sub>1</sub>=1,47; *df*<sub>2</sub>=1,15.

$p \leq .05$   
\*

**Table 6**

Experiment 2, example descriptions for target images.

| <i>Maintained precedent condition</i> |  |                                    |
|---------------------------------------|--|------------------------------------|
| <b>Rounds</b>                         | <b>Target description</b>                  | <b>Competitor description</b>      |
| 1–4                                   | <u>the crushed up bit of</u> cookie crumbs | the sheet of paper that's crinkled |
| 5                                     | <u>the crushed up bit of</u> cookie crumbs | the sheet of paper that's crinkled |
| <i>Broken precedent condition</i>     |  |                                    |
| 1–4                                   | the pile of crumbled pebbles               | the sheet of paper that's crinkled |
| 5                                     | <u>the crushed up bit of</u> cookie crumbs | the sheet of paper that's crinkled |

*Note.* In this example, the target image is the cookie crumbs/pebbles, and the competitor is the crinkled paper. The Experimenter was prompted with each entire noun phrase. The temporarily ambiguous portion of the target description (given the scene) is underlined.



Table 7

Mean square error (*MSe*) and *F* tests for latency to fixate the target in Experiment 2.

| Source of variance | Participants |           | Items      |           | <i>MinF'</i> |              |
|--------------------|--------------|-----------|------------|-----------|--------------|--------------|
|                    | <i>MSe</i>   | <i>F1</i> | <i>MSe</i> | <i>F2</i> | <i>df</i>    | <i>MinF'</i> |
| Partner            | 116411       | .60       | 63477      | .85       | 1,60         | .35          |
| Expression         | 61959        | 33.07*    | 59806      | 31.77*    | 1,62         | 16.20*       |
| P × E              | 88833        | .82       | 91713      | 1.12      | 1,61         | .47          |

Note. *df*<sub>1</sub>=1,31, *df*<sub>2</sub>=1,31.

\* *p*<.05

**Table 8**  
Mean square error (*MSe*) and *F* tests for target advantage scores in Experiment 2.

| <i>Time region</i>              | Participants |           | Items      |           | <i>MinF'</i> |              |
|---------------------------------|--------------|-----------|------------|-----------|--------------|--------------|
|                                 | <i>MSe</i>   | <i>F1</i> | <i>MSe</i> | <i>F2</i> | <i>df</i>    | <i>MinF'</i> |
| <i>Region 1: -200 to 200ms</i>  | Partner      | .05       | .029       | .02       | 1,52         | .01          |
|                                 | Expression   | .34       | .027       | .18       | 1,57         | .12          |
|                                 | P × E        | .01       | .029       | .00       | 1,31         | 0            |
| <i>Region 2: 200 to 600ms</i>   | Partner      | 4.19*     | .029       | 3.87~     | 1,62         | 2.01         |
|                                 | Expression   | 18.02*    | .022       | 29.24*    | 1,59         | 11.15*       |
|                                 | P × E        | 4.40*     | .028       | 4.77*     | 1,62         | 2.29         |
| <i>Region 3: 600 to 1000ms</i>  | Partner      | .00       | .034       | .00       | 1,31         | 0            |
|                                 | Expression   | 67.82*    | .041       | 84.56*    | 1,61         | 37.64*       |
|                                 | P × E        | .95       | .026       | 1.63      | 1,58         | .6           |
| <i>Region 4: 1000 to 1400ms</i> | Partner      | 6.68*     | .035       | 6.79*     | 1,62         | 3.37         |
|                                 | Expression   | 71.67*    | .042       | 98.35*    | 1,61         | 41.46*       |
|                                 | P × E        | .10       | .026       | .08       | 1,61         | 0.04         |
| <i>Region 5: 1400 to 1800ms</i> | Partner      | 7.26*     | .027       | 7.55*     | 1,62         | 3.70~        |
|                                 | Expression   | 3.23~     | .093       | 1.88      | 1,58         | 1.19         |
|                                 | P × E        | .15       | .032       | .14       | 1,62         | .07          |

Note. Region 1 is used as a baseline.  $df_1 = 1, 31$ ;  $df_2 = 1, 31$ .

\*  $p \leq .05$   
~  $p \leq .10$