HHS Public Access

Author manuscript

Lang Speech. Author manuscript; available in PMC 2015 December 21.

Published in final edited form as:

Lang Speech. 2015 June ; 58(0 2): 190–203.

Inferring difficulty: Flexibility in the real-time processing of disfluency

Daphna Heller[1], **Jennifer E. Arnold**[2], **Natalie M. Klein**[3], and **Michael K. Tanenhaus**[4]

[1]University of Toronto [2]University of North Carolina, Chapel Hill [3]Office of Research Protections, US Army Medical Research and Materiel Command [4]University of Rochester

Abstract

Upon hearing a disfluent referring expression, listeners expect the speaker to refer to an object that is previously-unmentioned, an object that does not have a straightforward label, or an object that requires a longer description. Two visual-world eye-tracking experiments examined whether listeners directly associate disfluency with these properties of objects, or whether disfluency attribution is more flexible and involves situation-specific inferences. Since in natural situations reference to objects that do not have a straightforward label or that require a longer description is correlated with both production difficulty and with disfluency, we used a mini artificial lexicon to dissociate difficulty from these properties, building on the fact that recently-learned names take longer to produce than existing words in one's mental lexicon. The results demonstrate that disfluency attribution involves situation-specific inferences; we propose that in new situations listeners spontaneously infer what may cause production difficulty. However, the results show that these situation-specific inferences are limited in scope: listeners assessed difficulty relative to their own experience with the artificial names, and did not adapt to the assumed knowledge of the speaker.

Keywords

disfluency; reference; inferences; artificial words; eye-tracking

As cognitive load increases, speakers become more disfluent (e.g., Bortfield, Leon, Bloom, Schober, & Brennan, 2001; Goldman-Eisler, 1968; Siegman, 1979). Disfluency increases when speakers are planning utterances (e.g., Beattie, 1979; Clark & Fox Tree, 2002; Clark & Wasow, 1998), at major breaks in discourse structure (Swerts, 1998; Swerts & Geluykens, 1994), and when speakers are unsure of the answer to a question (Brennan & Williams, 1995; Smith & Clark, 1993) or must choose between a number of alternatives (Schachter, Christenfeld, Ravina, & Bilous, 1991; Schachter, Rauscher, Christenfeld, & Tyson Crone, 1994). Disfluency also increases when speakers experience difficulty with a specific aspect

Correspondence should be addressed to Daphna Heller, Department of Linguistics, University of Toronto, 100 St. George Street, Toronto ON M5S 3G3, CANADA. daphna.heller@utoronto.ca.

Daphna Heller, Department of Linguistics, University of Toronto; Jennifer E. Arnold, Department of Psychology, University of North Carolina at Chapel Hill; Natalie Klein, Office of Research Protections, US Army Medical Research and Materiel Command; Michael K. Tanenhaus, Department of Brain and Cognitive Sciences, University of Rochester.

of their utterance. For example, speakers are more disfluent (i) when lexical retrieval is more difficult because the word being planned is low in frequency or in contextual probability (Beattie & Butterworth, 1979; Goldman-Eisler, 1968), (ii) when the syntactic structure being planned is the less frequent one (Cook, Jaeger & Tanenhaus, 2009), or (iii) when the referring expression being planned is for an object that that has not been mentioned recently or that is unconventional and thus hard to describe (Arnold & Tanenhaus, 2011).

Listeners, in turn, utilize disfluency patterns to generate expectations about how the utterance will unfold. Similar to effects in production, disfluency biases listeners against words that are high in contextual probability, i.e. draws their attention to the less expected continuations (Corley, MacGregor & Donaldson, 2007). When processing reference, a disfluent referring expression that contains a filler, such as *"thee uh… "*, leads listeners to develop the expectation that the disfluent expression will refer to (a) an object that has not been previously referred to (Arnold, Tanenhaus, Altmann & Fagnano, 2004); (b) an object that does not have a conventional name (e.g. a squiggly shape), thus requiring the planning of a longer description (Arnold, Hudson Kam & Tananhaus, 2007; Watanabe, Hirose, Den, & Minematsu, 2008); and (c) an object that does not have a conventional name *and* has not been previously referred to (Barr, 2001). That is, listeners associate disfluent speech with aspects of language that are perceived as difficult for the speaker to plan and therefore cause disfluency.

In standard situations, unconventional objects present difficulty in producing referring expressions because they require both a novel conceptualization and the planning of longer descriptions. The goal of the current study is to dissociate planning difficulty from conceptualization and length. To this end, we used several sets of novel shapes, teaching participants artificial names for them during a training session. We hypothesized that reference to these shapes may be perceived as difficult, because it has been shown that recently-learned names take longer to produce than established words in one's native language (Costa, Santesteban & Ivanova, 2006). At the same time, the unnamed shapes in our study had properties that have been previously associated with disfluency. First, they required both a novel conceptualization and longer descriptions (Arnold et al., 2007; Barr, 2001), and, second, they were not referred to during the training session (Arnold et al., 2004; Barr, 2001).

This novel situation allows us to examine the flexibility of drawing inferences about the attribution of disfluency. Specifically, if, upon hearing disfluent speech in this novel situation, listeners expect reference to objects with properties that have been previously associated with disfluency (i.e. objects that have not been previously referred to, objects that do not have a conventional name and/or objects that require a longer description), then they should develop an expectation that the upcoming referent would be an unnamed shape. If, however, listeners are more flexible and they actively assess planning difficulty in this new situation, and they perceive the artificial names as difficult to retrieve and produce, then they should develop the expectation that the upcoming referent would be a named shape.

Previous work has shown that listeners exhibit some flexibility in the real-time attribution of disfluency. For example, when listeners receive instructions from multiple speakers, they are

sensitive to the identity of the speaker in determining which objects have already been referred to, with disfluency biasing towards objects that have not been referred to by the speaker uttering the disfluent speech (Barr & Seyfeddinipur, 2009). Another study has shown that when listeners were told that the speaker had object agnosia and thus experienced difficulty naming ordinary objects, they did not show a bias toward objects lacking a conventional name when processing disfluency (Arnold et al., 2007). However, this flexibility has certain limitations. For example, in situations where disfluency was preceded by construction noise, listeners said they thought the noise was distracting to the speaker, and yet their real-time processing still attributed the disfluency to the speaker trying to name an object that lacks a conventional name (Arnold et al., 2007). This raises the possibility that while disfluency involves some situation-specific inferences, it is tied to objects with certain properties, such as not being referred to previously (i.e., being previously unmentioned), or objects that lack a conventional name or require a longer description. Our goal in the current study is to examine whether listeners are flexible in their attribution of disfluency, such that they can dissociate disfluent speech from the properties with which it tends to occur.

A secondary goal of the current study is to build on previous findings, and ask whether listeners use disfluency information flexibly, integrating information about the knowledge of a speaker even when it differs from their own knowledge. To this end, we included another set of shapes, for which listeners learned an artificial name during training, but were subsequently told that the speaker did *not* learn that name. This set up a contrast between the participant's own knowledge (3 artificial names) and the knowledge that they assumed the speaker to have (2 artificial names). If listeners' processing of disfluency is based on complex inferences about the speaker, then these shapes will be treated like unnamed shapes, because the speaker is not expected to know that name. But if listeners assess the source of disfluency based on their own experience, then those shapes will pattern with the other named shapes, because the listeners were trained on all names.

EXPERIMENT 1

Participants first learned artificial names for three types of shapes in a passive comprehension task. They were subsequently told that they would have to follow instructions recorded by a naïve participant who had received the same training, except that this participant had only learned two of these names (they were told which ones). The comprehension phase employed the Visual World Paradigm (Cooper 1974; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995), where participants followed instructions to click on shapes as their eye-movements were recorded. Displays contained two pairs of shapes in two colors – see Figure 1. This visual context was chosen because (i) in order to refer successfully to one of the four shapes, a color adjective must be used, e.g. "*the blue plinuk*" or "*the blue one with the hook on the bottom*", and (ii) the color adjective renders the referring expression temporarily ambiguous, in the sense that the color adjective is equally compatible with two shapes (cf. Arnold et al., 2007). The form of disfluencies was a lengthened "*thee*" followed by the filler "*uh*", which have been argued to be collateral signals used by speakers to manage conversation (Clark, 1996; Clark & Fox Tree, 2002; Barr & Seyfeddinipur, 2009). The disfluent determiner occurred right before the color

adjective, e.g. "*Click on theee uh blue…*", so the interval of adjective processing allowed us to examine how disfluency affects listeners' expectations about upcoming reference.

To examine our first question about flexibility in disfluency attribution, we used displays with two named shapes and two unnamed shapes. If disfluency is directly associated with objects that lack a name and require a longer description, then more looks are expected to the unnamed shape of the mentioned color upon processing of the color adjective that follows a disfluent determiner, e.g. *Click on thee uh red*. If, however, disfluency processing is more flexible and listeners adapt to the particular situation, listeners may expect disfluency before the recently-learned names, so we should see more looks to the shape with the artificial name. To examine our second question about whether these effects are modulated by assumptions about the speaker, we used displays with two named shapes and two shapes with listener-privileged names. If listeners integrate information about the speaker's knowledge in their attribution of disfluency, these trials should create the same pattern as the matched-perspective trials, because, from the speaker's perspective, these trials contain two named and two unnamed shapes. If, however, listeners assess the production difficulty for names from their own perspective, listener should not prefer one shape over the other, because shapes with listener-privileged names will be treated like other named shapes. In addition to measuring listeners' tacit biases during real-time processing, we asked them during debriefing about which shapes they thought caused more disfluency (in actuality, disfluencies appeared equally likely before reference to the different objects).

The motivation for creating sets of shapes rather than individual shapes was twofold. First, this allowed a situation where all the shapes used during testing were new – a particular shape was shown only once in the experiment, so each individual shape appeared either in the training or in the testing phrase. More important, this created a situation where shapes without names could reasonably be referred to with varying descriptions during testing; recurring shapes would have created the expectation that the speaker will reuse the same description, as is natural in conversation (Clark & Wilkes-Gibbs, 1986).

Methods

Participants—We report data from 28 native English participants from Rochester, NY. Three additional participants were tested but excluded from analysis, because of equipment problems (n=1) or because they did not believe our cover story (n=2). Participants were paid $7.50 each for their participation.

Materials—Four sets of 72 shapes were created. The shapes in each set were similar enough for a name to apply to all of them, but the sets differed from each other so individual shapes could reasonably only belong to one set – see Figure 2. For each participant, three bi-syllabic made-up names (phonetically: /plinʌk /, /kɑbɪt /, /duvɪt /) were randomly assigned to three shape sets. To avoid biases related to name-shape combinations, the names rotated through the shapes in a modified Latin square design, creating 12 lists. Shapes were black during training, and blue, red or green with a black object in the middle (heart, square, triangle, or star) during the visual-world testing– see again Figure 1.

Named shapes were always referred to with their names. Shapes with no name, and shapes with listener-privileged names, were referred to with a description that had a repetitive frame and used general terms, like "*the [color] one with the…*" or "*the [color] shape that has…*". Half of the descriptions mentioned the black object (e.g. "*with the heart*") and half mentioned another property (e.g. "*that has arms*"). To avoid intonational cues about the length of the upcoming referring expression, we cross-spliced the beginning of the instruction up to the end of the color from instructions containing descriptions to instructions containing names.

Three factors were manipulated in a 2×2×2 within-subjects design: Perspective (matched vs. mismatched) × Fluency (fluent vs. disfluent) × Reference (name vs. description). Perspective was manipulated by changing the shapes in the display: in the matched-perspective conditions named shapes were coupled with unnamed shapes, and in the mismatched conditions named shapes were coupled with listener-privileged shapes. Fluency was manipulated by changing the form of the definite article: the fluent article was "*thuh*", whereas the disfluent article was prolonged and followed by a filler: "*thee uh*". Finally, Reference was manipulated by changing the target in the display, which, in turn, determined the type of referring expression to be used. Specifically, named shapes were referred to with their names, while unnamed shapes, as well as shapes with listener-privileged names, were referred to with descriptions. All in all, there were four trials in each condition, yielding thirty-two experimental trials.

In addition, there were four fillers that served as practice trials: two fillers contained two pairs of named shapes, and another two fillers contained two unnamed shapes and two listener-privileged shapes. Because of the rotation of names across shapes (see above), twelve lists were created, such that each list had different shapes with different descriptions. In all lists, the four randomly-ordered fillers appeared first, and were followed by the pseudo-randomly ordered critical trials.

Procedure—Participants were told that the experiment tested "how people give and follow instructions when the other person is not there." We told them that "participants are randomly assigned by the computer to production or comprehension" and that "we use the instructions produced by previous participants for those assigned to comprehension." This gave a natural explanation for disfluency (Arnold et al., 2004, 2007), and allowed participants to assess the training experienced by the speaker.

Phase 1: Name training—Participants first learned three names. On each trial, two shapes appeared on the screen: the target was randomly drawn from one named set and a distractor was randomly drawn from another named set. An instruction in a male voice said: "*Click on the [name]*". Independent of the response, the distractor disappeared and the target stayed on the screen for 500 ms. Participants progressed to the next phase when they performed perfectly on one block (eighteen trials, six for each name). The second phase of training was identical, except it involved four shapes in every display, one from each set (unnamed shapes were never referred to).

Phase 2: Perspective training—Participants were informed that they were "randomly assigned by the computer to comprehension", and were shown which two shapes their speaker learned the names for. To train on the knowledge mismatch with the speaker, on each trial one shape appeared on the screen, and participant answered the question "Did the speaker learn a name for this shape?" by clicking true or false. An error sound played if participants were wrong in their answer. Participants advanced to the next phase after performing perfectly on one block (five shapes from each of the four sets; twenty shapes altogether). To ensure that participants were not more familiar with shared names, no names were used during this phase of training.

Phase 3: Visual-World comprehension—During this phase, participants' eye-movements were recorded using an EyeLinkII head-mounted eye-tracker. On each trial, four pictures appeared on the screen and the instruction played over speakers. The trial ended when the participant clicked on any shape (no feedback was given). Importantly, this phase utilized a female voice to emphasize the contrast with the male voice in training and to make the cover story believable.

Phase 4: Questionnaire—Participants' awareness of the experimental manipulations was assessed following Arnold et al. (2007). We started by asking general questions about the experiment. We then asked specifically whether the participant had thought that the speaker was reading instructions or that she was told to sound disfluent during the comprehension phase. Finally, we asked about participants' attribution of disfluency: whether they thought the speaker was more likely to be disfluent before named or unnamed shapes (even though, as per our design, disfluency rates were equal across the named- and unnamed-target conditions).

Results and Discussion

Participants quickly learned the names during name training (*Phase 1*): they completed an average of 2.71 blocks in the first part and 1.46 blocks in the second part in order to meet the criterion of 100% accuracy on a block of eighteen trials. They were also quick to learn in the perspective training (*Phase 2*), requiring an average of 1.94 blocks to perform perfectly on a block of fifteen trials.

Our main analysis involved assessing the effect of disfluency on processing reference (*Phase 3*). To this end, we examined fixations that occurred during the processing of the ambiguous color adjective, which is processed after the (dis)fluent determiner, but before the disambiguating information from the noun. Because it takes 200 ms to program and launch a saccade (Hallett, 1986), we focus on the interval from 200 ms after adjective onset to 200 ms after noun onset.

We begin by examining the matched-perspective conditions: Figure 3 plots the proportion of fixations to the four objects in each of the four matched-perspective conditions: fluent-name, fluent-description, disfluent-name, disfluent-description (eleven trials, or 1%, were excluded because the participant did not click on the target). Qualitatively, during fluent instructions (e.g. "*Click on the*") listeners were as likely to look at the named and unnamed shapes of the mentioned color, until they heard disambiguating information from the noun which revealed

the referent (see the left panel of Figure 3). Interestingly, following disfluency (e.g., "*Click on theee uh*"), listeners looked more to the named shape of the mentioned color as compared with the unnamed shape of that color (see the right panel of Figure 3). Note that this pattern is observed in both disfluent conditions, creating an anticipation of the intended referent when the named shape was the target (disfluent-name condition), and a bias *against* the intended referent when the intended referent was the unnamed shape (disfluent-description condition). In all cases, all four objects were considered equally before the onset of the color adjective.

To quantify this pattern, we calculated "target-advantage ratio": the likelihood of looking to the target over the likelihood of looking to either object that matches the color – this measure reflects which of the two potential referents listeners prefer, if any. We focused on the interval of adjective processing – a window of 312 ms, spanning 200 ms after the onset of the adjective to 200 ms after the (average) onset of the noun. For this interval in each trial, we first computed the probability of fixating the target and the probability of fixating the competitor, and then created the target-advantage-ratio for the trial: the probability of fixating the target over the probability of fixating and target or the competitor (Heller, Gronder & Tanenhaus, 2008; Wolter, Gorman & Tananhaus, 2011). Target-advantage-ratios were quasi-logit transformed before being submitted to an ANOVA (see Agresti 2002, Jaeger 2008).

These target-advantage ratios were submitted to a 2×2 (Fluency × Reference) repeated-measures ANOVA. Because our design counter-balanced across lists both name-shape combinations and which shapes were named or unnamed, we only conducted by-subjects analyses (cf. Raaijmakers, Schrijnemakers, & Gremmen, 1999).There was a main effect of reference, $F(1,27) = 13.65$, $p = .001$: target-advantage ratios were higher in the name conditions, indicating that participants had an overall preference to look at named shapes. In addition, the Fluency × Reference interaction was significant, $F(1,27) = 5.52$, $p < .05$. Planned comparisons confirmed that the target-advantage ratios did not differ in the fluent conditions, $F < 1$, indicating that a fluent determiner did not bias listeners toward a certain kind of shape. In contrast, target-advantage ratios differed significantly when the determiner was disfluent, $F(1,27) = 19.25$, $p < .001$. This effect reveals that disfluency biased listeners to anticipate named shapes, which was the target in the disfluent-name condition, and the competitor in the disfluent-description condition. The target-advantage ratios, in log-odds space, are plotted in the left panel of Figure 4 (these are the same conditions as the ones plotted in Figure 3). Recall that in log-odds space, chance (50%) is 0, so positive values indicate the anticipation of the target, whereas negative values indicate the anticipation of the competitor. This bias towards the named shapes contrasts with previous work, in which disfluency biased participants towards objects that did not have a straightforward label and required a longer description. This pattern shows that disfluency processing is performed by a flexible mechanism, where listeners draw inferences that are situation specific. We propose that in this situation listeners perceived the artificial names as harder to produce, expecting them to present more difficulty to the speaker.

Next, we examine the mismatched perspective conditions, which addressed the question of whether listeners relativize their inference to the assumed knowledge state of the speaker

when it differs from their own. The target-advantage ratios in the mismatched conditions, transformed into log odds space, are plotted in the right panel of Figure 4 (for considerations of space we do not present the fixation plots). Recall that displays in the mismatched-perspective conditions contained two named shapes and two shapes with the listener-privileged name, so if listeners adapt to the speaker's assumed knowledge of names, disfluency should introduce a bias towards the shared-name shape, as in the matched-perspective conditions. But in contrast to the interaction we observe for the matched-perspective conditions, the mismatched-perspective conditions do not show any biases away from chance. In other words, upon hearing a disfluent determiner "*thee uh*" listeners were not biased to look at a named shape (with a shared name) over a listener-privileged shape. Instead, they treated the two type of shapes equally, consistent with their own experience (since the participants knew the names for both). Indeed, the Fluency × Reference ANOVA on logit-transformed target advantage ratios did not reveal any effects, Fs < 1. This pattern indicates that listeners did not assign a special status to shapes with listener-privileged names, and treated them like named shapes with shared names. Thus, while the matched-perspective conditions showed that the attribution of disfluency is flexible and involves situation-specific inferences, these inferences seem to reflect listeners' own experience with the artificial names, rather than the assumed knowledge state of the speaker. This suggests a limitation on the ability of listeners to adapt their inferences about the attribution of disfluency.

Finally, when participants were asked directly about their attribution of disfluency during the questionnaire (*Phase 4*), 86% of participants said that the speaker was more disfluent with the unnamed shapes. This conscious attribution of disfluency sharply contrasts with the pattern of eye-movements observed during real-time processing, where disfluency biased listeners towards the named shapes. This pattern may reflect participants' conscious awareness that disfluent speech normally occurs before longer descriptions, whereas their implicit processing mechanism is more flexible. This disconnect is reminiscent of the finding in Arnold et al., (2007) where 71% of the participants said that construction noise was the cause of disfluency, and yet participants' eye-movements showed sensitivity to whether objects had lacked a conventional name and required a longer description.

EXPERIMENT 2

When processing disfluent speech, listeners in Experiment 1 were biased towards shapes with recently-learned names over those without names. This occurred despite the fact that our unnamed shapes had properties that are usually associated with disfluency: they were previously-unmentioned (as they were never referred to during training), and they did not have a readily available label (as they were not assigned an artificial name). Thus, they required a longer referring expression. We concluded that this bias, which is the reverse of what has otherwise been observed in the literature, indicates that the attribution of disfluency relies on a flexible mechanism that allow listeners to make situation specific inferences. This explanation assumes that listeners assessed the artificial names as harder to produce than longer descriptions that involved a novel conceptualization. If this is correct, it should be possible to eliminate or even reverse the bias by changing the experience participants have with the artificial name. But if the bias in Experiment 1 was due to some

property of the artificial names, perhaps because novel words are perceived as inherently difficult, then changing the experience with the names should not affect the bias.

Experiment 2 changed the experience of listeners with names, examining how this affects the attribution of disfluency in real time, but also consciously. Specifically, we enhanced the name training stage by adding an active task in which participants practiced producing the names. In addition, the perspective training was enhanced by adding a task where participants had to practice giving instructions to someone with the assumed knowledge of their speaker.

Methods

Participants—We report data from 22 native English speakers from Rochester, NY, none of whom participated in Experiment 1. Two additional participants were excluded from analysis because they did not believe our cover story. Participants were paid $7.50.

Materials—The materials were identical to Experiment 1.

Procedure—The cover story was identical to Experiment 1.

Phase 1: Name training: The first two stages of the name training were as in Experiment 1. A third stage had participants produce names in isolation. On each trial, one shape appeared on the screen, and participants had to say its name. There were five shapes from each named set, yielding fifteen shapes altogether (i.e. both shapes that will later turn out to have shared names and listener-privileged names). No feedback was given.

Phase 2: Perspective training: The first part of the perspective training was identical to Experiment 1. In addition, participants produced instructions for four-shape displays that were similar to the visual-world comprehension displays in that they contained two pairs of identical shapes that contrast in color. Each of the four shape types was the (highlighted) target in three displays – two with shared names, one unnamed and one with a listener-privileged name. Participants were instructed to produce instructions that could be understood by a person with the same knowledge of names as their speaker. No feedback was given at this stage.

Phase 3: Visual-world comprehension: The eye-tracking task was as in Experiment 1.

Phase 4: Questionnaire: The debriefing stage was identical to Experiment 1.

Results and Discussion

As in Experiment 1, participants learned the names quickly, reaching the criterion of 100% performance on a whole block at an average of 2.14 blocks in the first part and 1.45 blocks in the second part. In addition, they performed well during name production, achieving 94% accuracy.

During perspective training, participants learned the perspective quickly, completing an average of 1.50 blocks to meet the 100% performance criterion. They then went on to

produce instructions (49 trials, or 13% of the data, were discarded because the participant clicked to end the trial before saying anything). First, the production task indicates that participants knew which shapes were named and which were not: when referring to shapes with shared names, participants virtually always used names (99% of trials), and when referring to unnamed shapes, they almost never produced a name (1% of trials with names). However, this task reveals that not all participants understood the perspective manipulation: 6 out of the 22 participants (27%) consistently used a name to refer to shape whose name was supposed not to be known to their partner (the overall likelihood of using names was 23%).

The eye-movement data was analyzed as in Experiment 1 (eight trials, or 1% of the data, were excluded because the participant did not click on the target). For ease of comparison with Experiment 1, we present the target-advantage ratios, transformed into log-odds space, during the interval of the processing of the ambiguous color adjective in Figure 5 – 0 is chance, a positive value represents a bias to the target, and a negative value a bias towards the same-color competitor. The matched-perspective condition did not show any biases from chance, contrasting with the interaction pattern obtained in Experiment 1. This was confirmed by the lack of any effects in a Fluency × Reference ANOVA performed on the quasi-logit transformed target advantage ratios (Fs < 1). Turning to our second question about sensitivity to the speaker's assumed knowledge of names, a Reference × Fluency ANOVA on target advantage ratios in the mismatch-perspective condition revealed no effects (main effect of Fluency: $F(1,21) = 3.18$, $p = .08$; main effect of Reference and Fluency × Reference : Fs < 1) – see again Figure 5. Note that because the added training completely eliminated the bias in the matched-perspective conditions (rather than reversing it, for example), this pattern is not informative. That is, this pattern is compatible with two different explanations: (i) that shared-name shapes and listener-privileged shapes were both treated as named, as in Experiment 1, and (ii) that the participants treated listener-privileged names like unnamed shapes, and, as in the matched-perspective condition, did not find naming shapes more difficult than producing a description.

Like in Experiment 1, there was a striking dissociation between listeners' real-time behavior and their explicit attribution of disfluency. When asked directly about the source of disfluency during debriefing, participants were just as likely to suggest the unnamed shapes (87%) as they were in Experiment 1. The fact that training did not affect conscious attributions of disfluency provides support to the proposal that these judgments reflect listeners' awareness of the linguistic context of disfluency, in contrast with the tacit mechanism which is more flexible and sensitive to situation-specific variables.

A final analysis focused on a subset of the data: the matched-perspective conditions with disfluent instructions from the two experiments. A 2×2 ANOVA examined the effect of Experiment (1, with less training vs. 2 with more training: a between-subjects factor) on Reference (name vs. description: a within-subjects factor). The effect of experiment was not significant, $F < 1$, but there was a main effect of reference, $F(1,48) = 5.76$, $p = .02$. Crucially, the Experiment × Reference interaction was significant, $F(1,48) = 7.80$, $p < .01$, demonstrating that referential expectation was affected differentially by the two levels of training. This pattern confirms that listeners' processing of disfluency does not reflect a

direct association of disfluent speech and objects with artificial names. Instead, it demonstrates the flexibility of the system, and is consistent with our proposal that real-time disfluency attribution depends on the perceived difficulty of name production in a particular situation.

GENERAL DISCUSSION

Using sets of novel shapes and a mini-artificial lexicon, we created a situation where some shapes were assigned a name that was readily available for reference, while others were not, which meant they required a novel conceptualization and the planning of a longer description. In Experiment 1, we found that a disfluency biased listeners towards the objects with the novel names. This is the first result to show that disfluency can bias listeners *against* objects that lack a conventional name and require a longer description, properties that have been previously correlated with disfluency (Arnold et al., 2007; Barr, 2001). Instead we propose that the training conditions in Experiment 1 were such that the newly-learned artificial names were perceived as more difficult to produce than longer descriptions for the unnamed shapes. This proposal is further supported by the fact that the bias toward objects with artificial names disappeared when we changed listeners' experience with the artificial names in Experiment 2 (we didn't change any of the properties of the stimuli themselves). Specifically, when listeners actually practiced producing the names, they were as likely to consider named and unnamed shapes following disfluency. We propose that in this situation listeners perceived the production of names or descriptions as difficult to a similar extent. Note that in previous research it was not always the case that listeners perceived one referent as harder to name than another: when processing disfluent instructions from a speaker with object agnosia, listeners were equally likely to look at objects with and without a conventional name (Arnold et al., 2007).

The contrast in the disfluency effect between the two experiments demonstrates that listeners' attribution of disfluency utilizes a flexible, situation-specific mechanism, and is not a direct mapping on disfluent speech to objects with certain properties. We hypothesize that the basis for these inferences is listeners' spontaneous assessment of what is expected to present production difficulty in the situation, but further research is required to demonstrate that listeners' inferences about difficulty are indeed what is driving inference about disfluency attribution. Interestingly, while the real-time processing of disfluency was affected by the different training in the two experiments, their conscious attribution of disfluency was not: in both experiments listeners thought that the unnamed shapes gave rise to more disfluent speech, which was diverged from the on-line behavior in both experiments. It seems plausible that these conscious attributions are sensitive to the linguistic contexts in which disfluencies are usually found and not to an assessment of the current situation alone.

While participants adapted to the new situation, their situation-specific inferences did not take into account the speaker's assumed knowledge of names when it differed from their own, but rather reflected their own experience with the artificial names. This seems surprising, because Barr and Seyfeddinipur (2009) found that listeners who processed disfluency did adapt to the speaker in assessing which objects were previously-unmentioned.

Moreover, when participants learn artificial names with a partner, they clearly distinguish between privileged and shared names in the form of their utterances (Heller, Gorman & Tanenhaus, 2012). Although this study is not directly comparable because it focuses on production, the combined results are consistent with the proposal that perspective-taking is mediated by shared experience that creates source memories for specific knowledge (Horton & Gerrig, 2005). In fact, it has been directly demonstrated that shared experience is crucial for perspective taking (Brown-Schmidt, 2009). Future research will address the question of whether listeners will integrate information about the speaker's knowledge into their disfluency attribution when the knowledge is acquired by shared experiences.

In sum, the current study demonstrates the flexibility of online disfluency processing, which depends on situation-specific inferences, but it also shows limitations on the information that is used in drawing these inferences.

Acknowledgement

For their help in creating stimuli and collecting and coding the data, we are grateful to Tom Covey, Rachel Sussman, Neil Bardhan, Kate Pirog Revill and Dana Subik. This research was partially supported by NSF grant BCS-0745627 to J. Arnold and NIH grant HD- 27206 to M. K. Tanenhaus. The views expressed in this article are those of the author(s) and do not reflect the official policy of the Department of the Army, the Department of Defense, or the U.S. Government.

REFERENCES

Agresti, A. Categorical Data Analysis. 2nd edition. Hoboken, New Jersey: John Wiley and Sons, Inc.; 2002.

Arnold JE, Hudson Kam CL, 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 & Cognition. 2007; 33:914–930.

Arnold, JE.; Tanenhaus, MK. Disfluency isn't just um and uh: the role of prosody in the comprehension of disfluency. In: Gibson, E.; Perlmutter, N., editors. The processing and acquisition of reference. Boston, MA.: MIT Press; 2011. p. 197-218.

Arnold JE, Tanenhaus MK, Altmann R, Fagnano M. The old and thee, uh, new. Psychological Science. 2004; 15:578–581. [PubMed: 15327627]

Barr, DJ. Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In: Cavé, C.; Guaïtella, I.; Santi, S., editors. Oralité et gestualité: Interactions et comportements multimodaux dans la communication. Paris: L'Harmattan; 2001. p. 597-600.

Barr DJ, Seyfeddinipur M. The role of fillers in listener attributions for speaker disfluency. Language and Cognitive Processes. 2009; 25:441–455.

Beattie GW. Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation. Linguistics. 1979; 17:61–78.

Beattie GW, Butterworth BL. Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. Language and Speech. 1979; 22:201–211.

Bortfeld H, Leon SD, Bloom JE, Schober MF, Brennan SE. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. Language and Speech. 2001; 32:229–259.

Brennan SE, Williams M. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. Journal of Memory and Language. 1995; 34:383–398.

Brown-Schmidt S. Partner-specific interpretation of maintained referential precedents during interactive dialog. Journal of Memory and Language. 2009; 61:171–190. [PubMed: 20161117]

Clark, HH. Using language. Cambridge: Cambridge University Press; 1996.

Clark HH, Fox Tree JE. Using *uh* and *um* in spontaneous speaking. Cognition. 2002; 84:73–111. [PubMed: 12062148]

Clark HH, Wilkes-Gibbs D. Referring as a collaborative process. Cognition. 1986; 22:1–39. [PubMed: 3709088]

Clark HH, Wasow T. Repeating words in spontaneous speech. Cognitive Psychology. 1998; 37:201–242. [PubMed: 9892548]

Cook, SW.; Jaeger, TF.; Tanenhaus, MK. Producing less preferred structures: More gestures, less fluency. In: Taatgen, NA.; van Rijn, H., editors. Proceedings of the 31st Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society; 2009. p. 62-67.

Cooper RM. The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. Cognitive Psychology. 1974; 6:84–107.

Corley M, MacGregor LJ, Donaldson DI. It's the way that you, er, say it: Hesitations in speech affect language comprehension. Cognition. 2007; 105:658–668. [PubMed: 17173887]

Costa A, Santesteban M, Ivanova I. How Do Highly Proficient Bilinguals Control Their Lexicalization Process? Inhibitory and Language-Specific Selection Mechanisms Are Both Functional. Journal of Experimental Psychology: Learning, Memory and Cognition. 2006; 32:1057–1074.

Goldman-Eisler, F. Psycholinguistics: Experiments in spontaneous speech. London: Academic Press; 1968.

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.

Heller D, Gorman KS, Tanenhaus MK. To name or to describe: shared knowledge affects choice of referential form. TopiCS in Cognitive Science. 2012; 4:295–305.

Heller D, Grodner D, Tanenhaus MK. The role of perspective in identifying domains of reference. Cognition. 2008; 108:831–836. [PubMed: 18586232]

Horton WS, Gerrig RJ. Conversational common ground and memory processes in language production. Discourse Processes. 2005; 40:1–35.

Jaeger TF. Categorical Data Analysis: Away from ANOVAs (transformation or not) and toward Logit Mixed Models. Journal of Memory and Language. 2008; 59:434–446. [PubMed: 19884961]

Raaijmakers JGW, Schrijnemakers JMC, Gremmen F. How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. Journal of Memory and Language. 1999; 41:416–426.

Siegman, AW. Cognition and hesitation in speech. In: Siegman, AW.; Feldstein, S., editors. Of speech and time: Temporal speech patterns in interpersonal contexts. Hillsdale, NJ: Erlbaum; 1979. p. 151-178.

Schachter S, Christenfeld N, Ravina B, Bilous F. Speech disfluency and the structure of knowledge. Journal of Personality and Social Psychology. 1991; 60:362–367.

Schachter S, Rauscher FH, Christenfeld N, Tyson Crone K. The vocabularies of academia. Psychological Science. 1994; 5:37–41.

Smith VL, Clark HH. On the course of answering questions. Journal of Memory and Language. 1993; 32:25–38.

Swerts M. Filled pauses as markers of discourse structure. Journal of Pragmatics. 1998; 30:485–496.

Swerts M, Geluykens R. Prosody as a marker of information flow in spoken discourse. Language and Speech. 1994; 37:21–43.

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. Science. 1995; 268:1632–1634. [PubMed: 7777863]

Watanabe M, Hirose K, Den Y, Minematsu N. Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. Speech Communication. 2008; 50:81–94.

Wolter L, Gorman K, Tananhaus MK. Scalar reference, contrast and discourse: Separating effects of linguistic discourse from availability of the referent. Journal of Memory and Language. 2011; 65:299–317. [PubMed: 21927536]

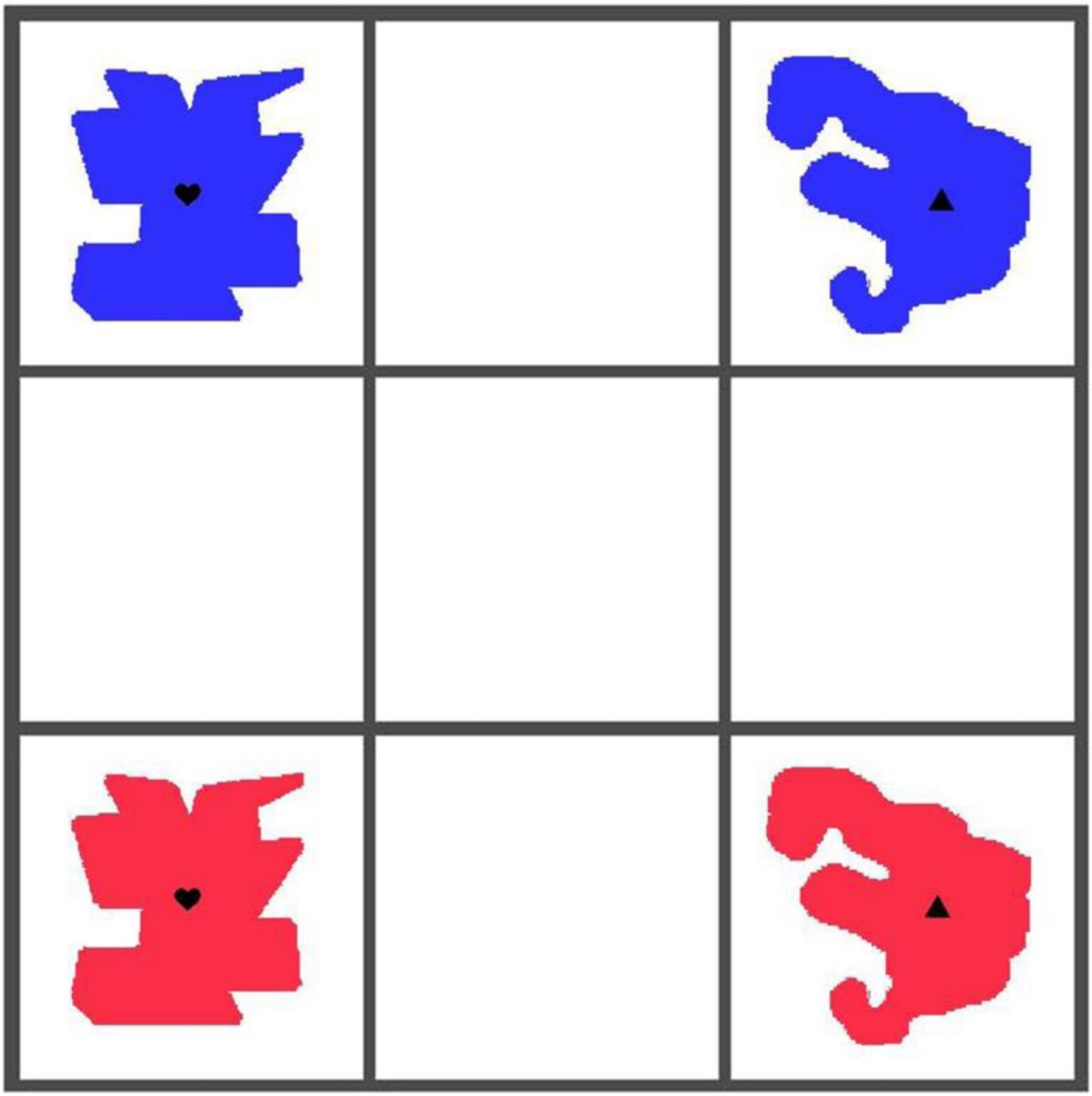

Figure 1.
Sample display for the visual-world comprehension task. The top shapes would be of one color (e.g. blue), whereas the bottom one were a different color (e.g., red).

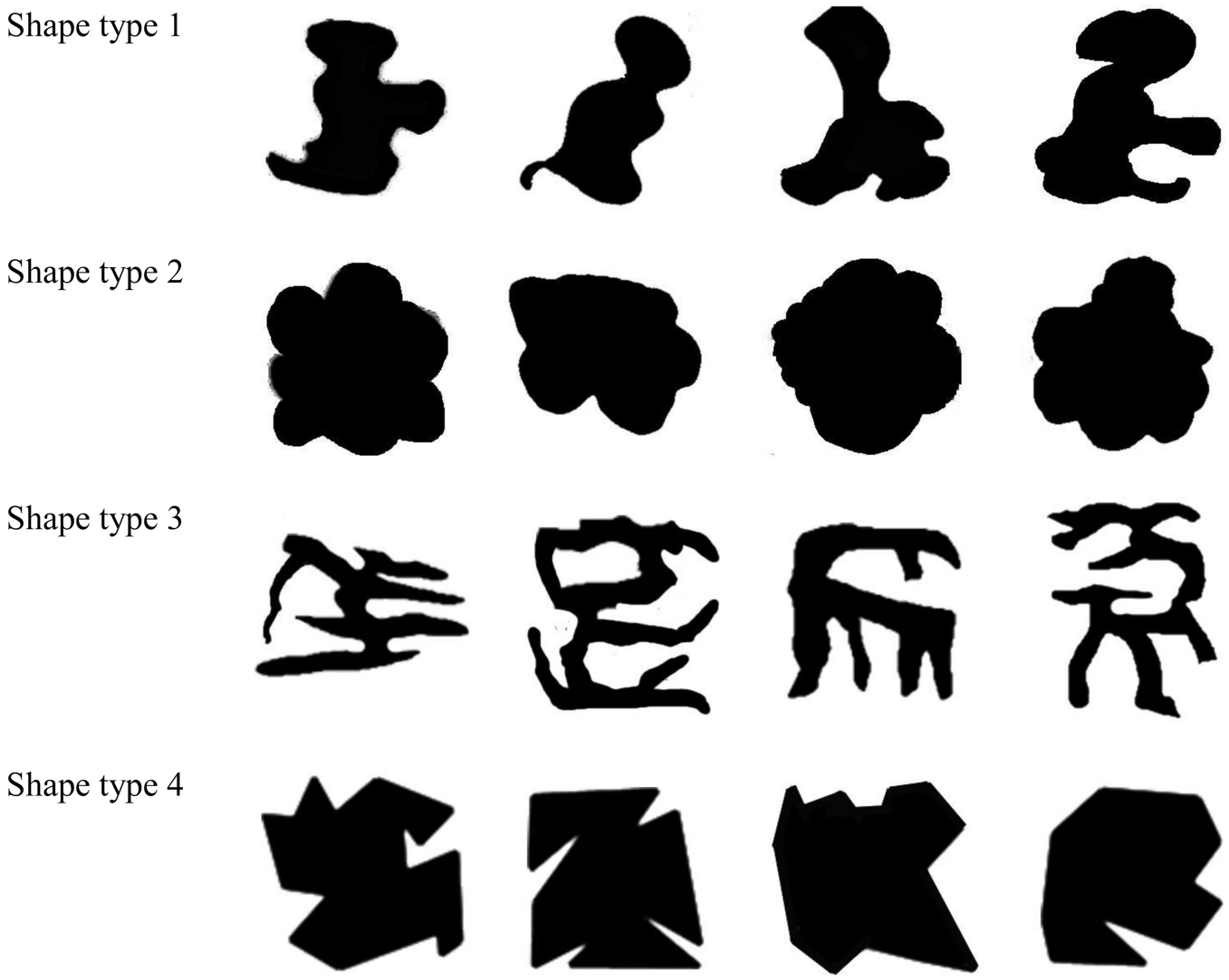

Figure 2.
Four exemplars from each of the four types of novel shapes. The full sets consisted of 72 exemplars each.

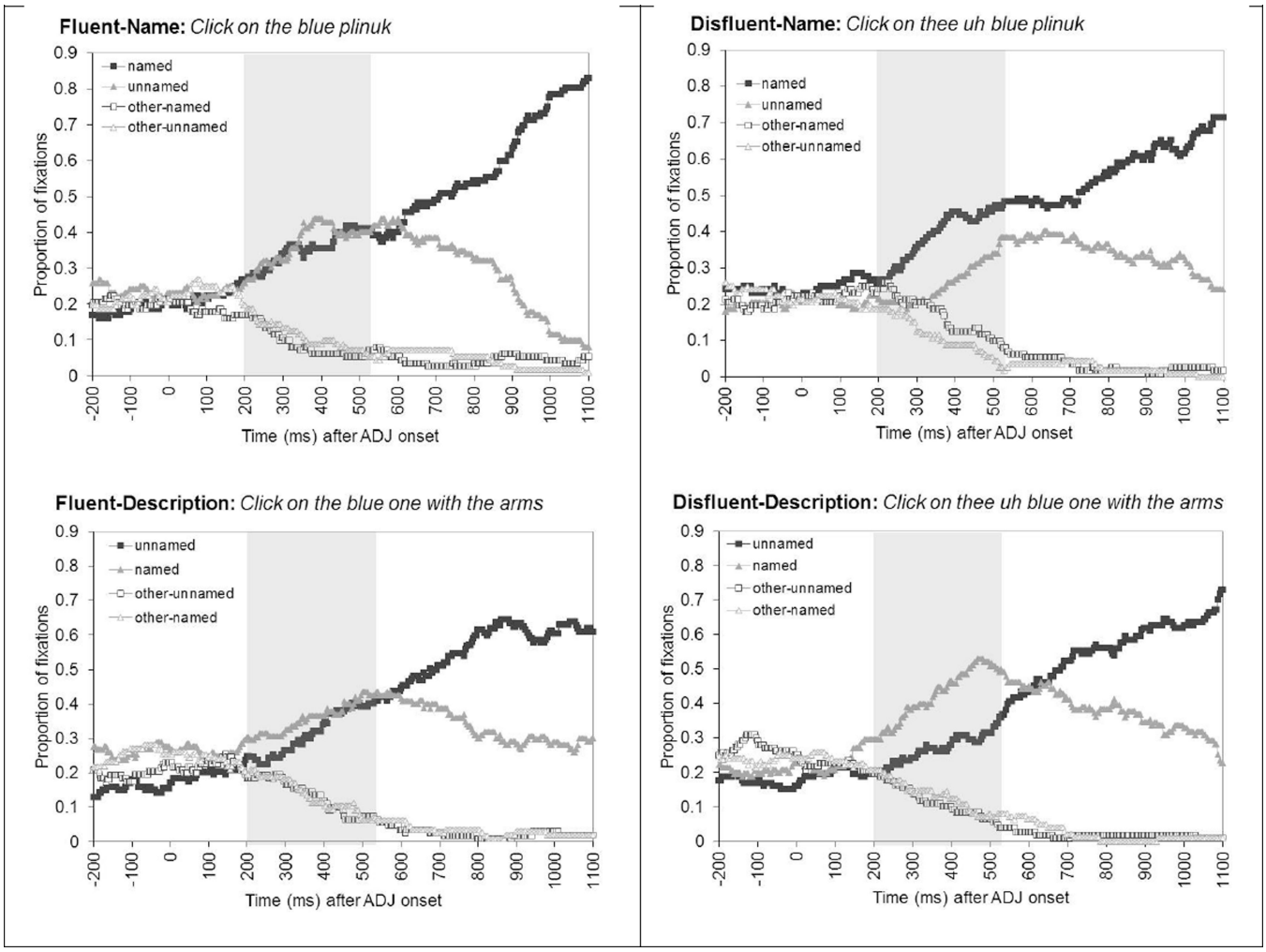

Figure 3. Proportion of fixations to the four objects in the display in the matched-perspective conditions, Experiment 1

Trials are aligned to the onset of the color adjective (e.g., *red*), at 0ms. The average noun onset (e.g. *plinuk* or *one*) is 312ms. The grey panel indicates the interval of adjective processing 200–512ms.

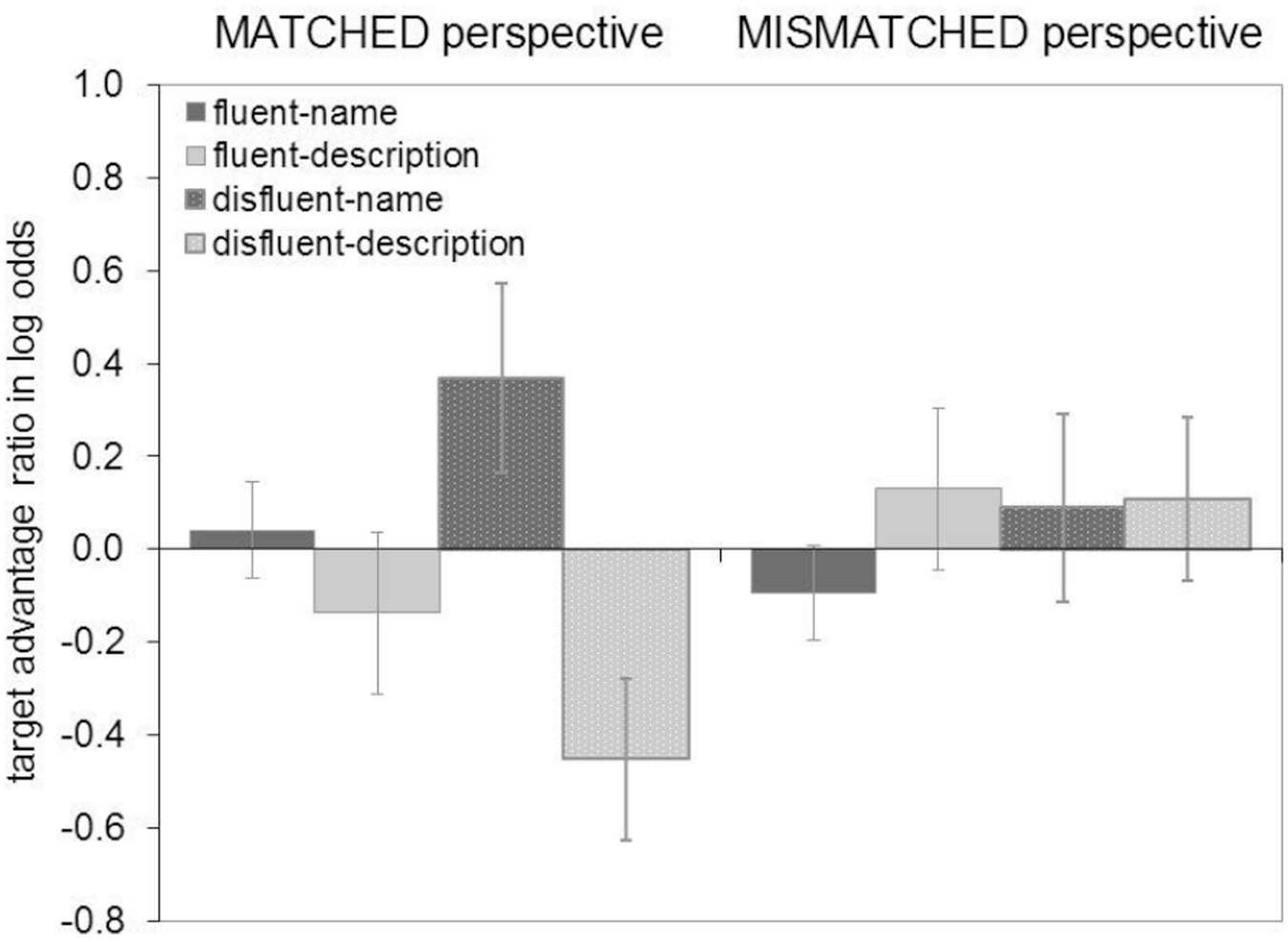

Figure 4.

Target advantage ratios in log-odds space across all the conditions in Experiment 1. A positive value indicates a bias for the target, while a negative value indicates a bias toward the same-color competitor (0 is chance).

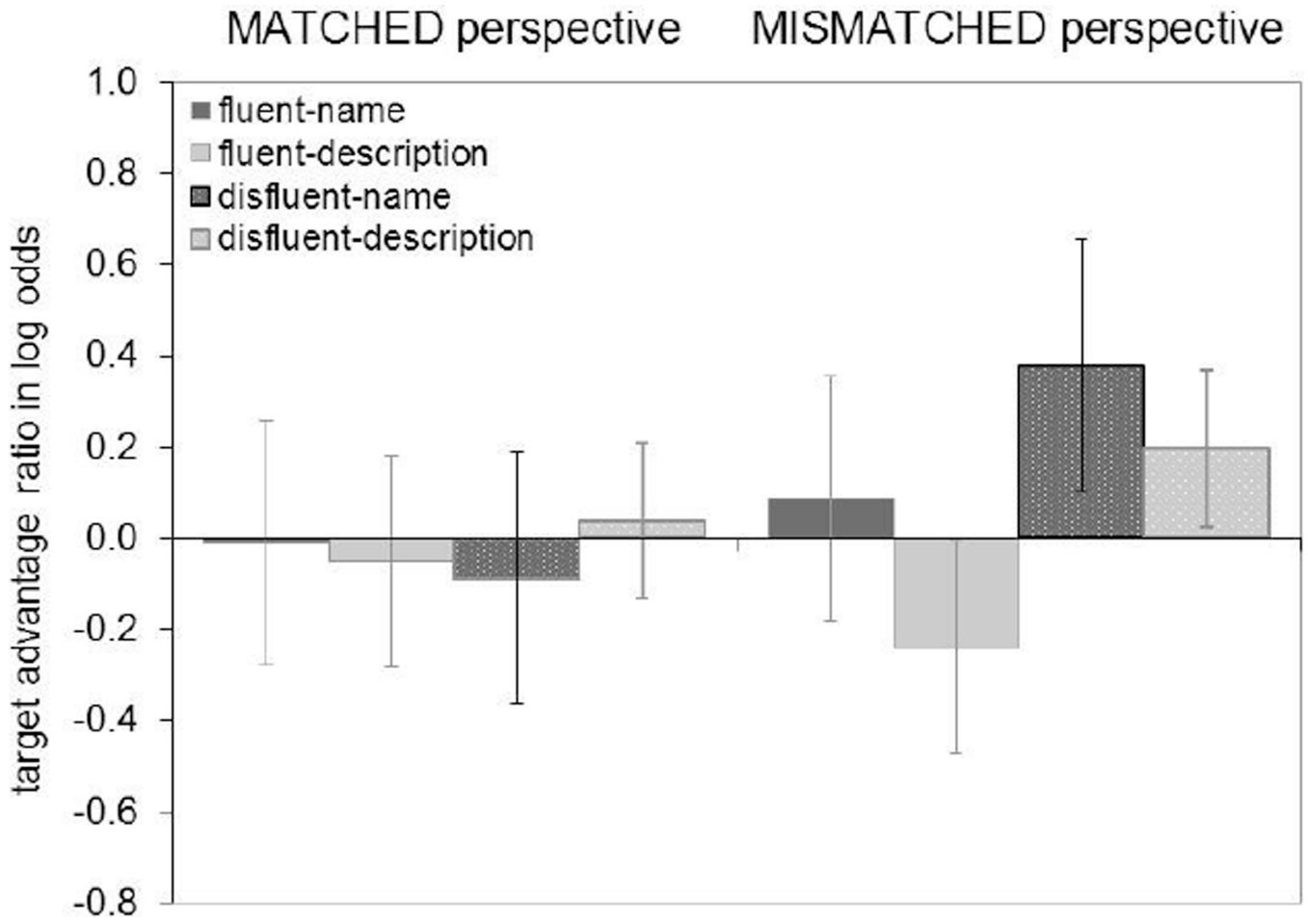

Figure 5.

Target-advantage ratios (in log-odds space), Experiment 2. A positive value indicates a bias for the target, while a negative value indicates a bias toward the same-color competitor (0 is chance).