

Does cognitive control modulate referential ambiguity resolution? A remote visual world study

Valerie J. Langlois, Tal Ness, Albert E. Kim & Jared M. Novick

To cite this article: Valerie J. Langlois, Tal Ness, Albert E. Kim & Jared M. Novick (25 Jun 2024): Does cognitive control modulate referential ambiguity resolution? A remote visual world study, Language, Cognition and Neuroscience, DOI: [10.1080/23273798.2024.2369182](https://doi.org/10.1080/23273798.2024.2369182)

To link to this article: <https://doi.org/10.1080/23273798.2024.2369182>



View supplementary material 



Published online: 25 Jun 2024.



Submit your article to this journal 



View related articles 



View Crossmark data 

REGULAR ARTICLE



Does cognitive control modulate referential ambiguity resolution? A remote visual world study

Valerie J. Langlois^a, Tal Ness^b, Albert E. Kim^a and Jared M. Novick^b

^aInstitute for Cognitive Science and Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, USA; ^bProgram in Neuroscience and Cognitive Science and Department of Hearing and Speech Sciences, University of Maryland, College Park, MD, USA

ABSTRACT

Cognitive control facilitates the resolution of representational conflict during language comprehension when incompatible interpretations vie for selection due to linguistic ambiguity. However, it remains unknown if control is required only when conflict arises between multiple, strongly supported interpretations, or even when there is substantial evidence for one interpretation and competition from weak alternatives. We investigated referential ambiguities such as “She will eat the red ...”, in which listeners temporarily consider multiple red objects as potential referents, including those that are partially consistent with the input (e.g. heart satisfies “red” but not “eat”). We introduce a remote visual-world paradigm to track listeners’ interpretive commitments via webcam, combined with a cross-task adaptation manipulation of cognitive-control engagement. We replicated subtle competition effects in referential ambiguity but found that upregulated cognitive control did not modulate competition. This suggests that a competing representation must reach a certain activation threshold before conflict arises, requiring cognitive control.

ARTICLE HISTORY

Received 21 September 2023
Accepted 5 May 2024

KEYWORDS

Cognitive control; referential ambiguity; webcam-based eye-tracking; language processing; visual world paradigm

Introduction

During real-time language comprehension, listeners often consider interpretations that differ from the one intended, due to pervasive ambiguity in linguistic messages at multiple levels of representation, including phonological, syntactic, semantic, and referential. Understanding how humans resolve such ambiguities has been a central focus of psycholinguistics research for decades. There is growing evidence that cognitive control plays an important role in some forms of ambiguity resolution, by biasing processing toward one contextually relevant interpretation when multiple conflicting interpretations are simultaneously active, enabling the rapid resolution of ambiguities that might otherwise take longer to resolve (Ness et al., *in press*). However, it is unclear whether cognitive control is involved in the resolution of all forms of ambiguity, or only in specific sorts of ambiguity that induce a conflict between two strongly supported but mutually incompatible representations. The current study tested whether cognitive control is involved in a form of referential ambiguity exemplified by the situation in which a listener hears a sentence fragment like “She will eat the red ...”, in a context containing multiple red objects that compete for reference (e.g. a red heart and a red pear).

Here, the listener must resolve the ambiguity over which object is likely to be the argument of the verb. However, we posit that such sentences do not engender any significant conflict because one referent receives strong support from multiple sources of evidence (e.g. pears are both edible and red) while the other does not (e.g. hearts are red, but inedible). To preview our results, we find that for these sorts of temporary referential ambiguities, listeners do consider both red objects as potential referents, as evidenced by their eye movement patterns; yet cognitive control is not necessarily involved in resolving such ambiguity.

Cognitive control is recruited to resolve conflict during language comprehension

In prominent accounts of cognitive control, the term “conflict” refers to situations where there are multiple strongly supported but incompatible ways of interpreting a stimulus. For instance, one type of conflict, known as “prepotent conflict”, occurs when individuals must suppress their dominant way of characterising a stimulus in favour of less favoured alternatives (Botvinick et al., 2001). The canonical example of this is the Stroop task, where participants name the font colour in which a

CONTACT Jared M. Novick  jnovick1@umd.edu  Department of Hearing and Speech Sciences, University of Maryland, 0100 Lefrak Hall, College Park, MD 20742

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23273798.2024.2369182>.

© 2024 Informa UK Limited, trading as Taylor & Francis Group

colour word is presented. Response times are longer, and accuracy is lower, if the colour word and its font colour do not match (e.g. the word “red” displayed in blue), compared to when there is no such conflict (e.g. “blue” displayed in blue). This decrease in performance is attributed to the need to override prepotent, automatic reading processes in favour of perceptual colour naming.

Conflict can also manifest as “under-determined conflict”, where multiple response options are equally plausible, leading to conflict in their selection. For example, in a verb generation task (e.g. Thompson-Schill et al., 1997), participants produce verbs in response to nouns or objects that are provided. Some objects elicit multiple alternative verbs because the stimulus itself does not create a single compelling response (high conflict; e.g. ball → kick, throw, catch, roll), while others prompt only one strongly linked verb (low conflict; e.g. scissors → cut). Response times are longer in the high-conflict condition due to the need to select from multiple strongly active alternatives. Resolving both types of conflict – prepotent conflict and underdetermined response conflict – necessitates cognitive control (e.g. Botvinick et al., 2001).

We have proposed that cognitive control is central to language comprehension, aiding in the selection of the most appropriate analysis when linguistic representations conflict (Ness et al., *in press*; Novick et al., 2005, 2010). One example of representational conflict during language processing is when comprehenders must revise early processing decisions in view of late-arriving, incompatible evidence. For instance, consider the command, “Put the frog on the napkin into the box”. Here, the phrase “on the napkin” is temporarily ambiguous: it could specify the goal (where the frog should be put), or a modifier (the frog to be moved somewhere is currently on a napkin). Findings from eye-tracking studies show that listeners provisionally assign the goal interpretation to “on the napkin” as it unfolds, and then revise to the modifier interpretation as “into the box” arrives (Novick et al., 2008; Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999). This early commitment to the napkin-as-goal analysis is strongly supported by the verb “put”, because it requires a goal (“Put the frog” is ungrammatical). The late-arriving input “into the box” provides strong support for the napkin-as-modifier interpretation, conflicting with listeners’ original analysis. In our model, cognitive control is needed to bias processing toward the napkin-as-modifier analysis over the “prepotent” napkin-as-goal analysis, enabling the listener to resolve the conflict and recover from their initial misinterpretation (Ness et al., *in press*).

Indeed, an individual’s ability to revise depends on their state of cognitive control (e.g. Hsu et al., 2021;

Hsu & Novick, 2016). Hsu and Novick (2016) experimentally manipulated people’s cognitive-control engagement through the performance of the Stroop task (i.e. Stroop-Incongruent vs. – Congruent trials), immediately before they followed “Put” instructions. On Stroop-Incongruent trials (e.g. “red” displayed in blue), the mismatch between the font colour and the word meaning induces representational conflict, which engages cognitive control more so than Stroop-Congruent trials (e.g. “blue” displayed in blue). Listeners looked at the correct goal (e.g. the box) earlier, and committed fewer action errors involving the incorrect goal (e.g. an empty napkin in the scene), after Stroop-Incongruent than – Congruent trials. The idea here is that the upregulation of cognitive control via Stroop persists for long enough to influence performance on the ensuing sentence trial, suggesting that increased cognitive-control engagement facilitates syntactic ambiguity resolution. We refer to this effect as “cross-task adaptation of cognitive control” (for a model of how this is achieved, see Ness et al., *in press*).

Other studies using this paradigm have found that increased cognitive control affects the resolution of different forms of linguistic conflicts, such as when strong syntactic and semantic cues point to incompatible ways of assigning thematic roles, e.g. “The fox was chased by the rabbit”. In this case, the syntax dictates that the rabbit and fox are (respectively) the Agent and Theme of chasing, but world knowledge suggests that it was probably the other way around (Ferreira, 2003). Thothathiri et al. (2018) showed that upregulated cognitive control via Stroop facilitated thematic role assignment under such conditions. Similarly, Ovans et al. (2022a) tested how cognitive-control engagement impacted the processing of *semantic attraction* sentences (e.g. “The bathroom floor was mopping yesterday”). Here, the more plausible “floor”-as-Theme interpretation conflicts with the syntactically-licensed “floor”-as-Agent one. Such sentences typically generate a P600 ERP effect (compared to “The bathroom floor was mopped yesterday”), suggesting morphosyntactic editing activity (“mopping” → “mopped”) to accommodate the likelier semantics-driven analysis. Crucially, there was a *larger* P600 effect following Stroop-Incongruent versus – Congruent trials. This indicates that cognitive control biased the outcome of the conflict toward the more plausible interpretation, which increased morphosyntactic repair attempts.

Ambiguity and conflict

In proposing that cognitive control aids in the resolution of conflict, we have defined conflict as involving more

than one strongly supported but incompatible representation of the input (consistent with Botvinick et al., 2001), which can lead to difficulty in selecting a single representation (Ness et al., *in press*; Novick et al., 2005, 2010). In the fox and rabbit sentences, for example, listeners face a choice between a highly plausible but syntactically unsupported interpretation, and a syntactically obligatory but implausible interpretation. The claim is that cognitive control is necessary for helping the comprehender to efficiently select a single, best-supported analysis in a situation of representational conflict. An underdeveloped but important implication of our model is that although conflict typically originates in ambiguity, some forms of ambiguity engender no conflict and therefore do not require cognitive control. That is, not all situations involving ambiguity constitute representational *conflict*.

An example of ambiguity without conflict occurs when the linguistic cues accumulated so far provide significantly more probabilistic support for one interpretation than they do for another. For example, Kukona et al. (2014) and Nozari et al. (2016) asked participants to listen to sentences like “She will eat the red pear”, while viewing a scene in which two of four objects were compatible with the adjective’s meaning (e.g. a pear and a heart are both red; see Figure 1A). Although the heart is locally compatible with the adjective’s meaning, the earlier sentential evidence, when combined with the adjective, favours the correct interpretation (pear) because the verb “eat” requires an edible object, which the pear satisfies but the heart does not. Thus, while this sentence contains a temporary referential ambiguity, the ambiguity does not fit neatly into the traditional categories of prepotent or underdetermined conflict as defined in classic models of cognitive control.

Despite support for the correct pear interpretation, the looking pattern from listeners’ eye-movement record indicates that the heart receives some consideration as a candidate for the upcoming verb-argument role (e.g. Altmann & Kamide, 1999). Shortly after hearing the adjective “red”, they looked at both adjective-compatible objects (heart and pear) more than adjective-incompatible objects in the scene, before looks converged on the target object (pear) as the final noun was heard. Given the evidence that these referential ambiguities engender some degree of competition, even though they do not involve conflict between two strongly supported interpretations, it is conceivable that non-conflict ambiguities still require assistance from cognitive control. In fact, one study concluded that cognitive control is involved in the resolution of such referential ambiguities (Nozari et al., 2016). They found that individuals with stronger performance on a cognitive control task (Flanker) showed smaller competition effects in the “red pear” sentences (fewer looks to the heart) than individuals with weaker performance. This correlational data suggests that individual cognitive control ability may be related to the resolution of referential ambiguity, and therefore implicates cognitive control in adjudicating between competing referents, even when there is no flagrant conflict between representations. However, the correlational finding does not indicate a causal link, nor can it eliminate the role of other factors that might mediate the observed relationship.

In the current study, we examined the relationship between cognitive control and referential ambiguity resolution by using a cross-task adaptation approach, as described above, to directly examine whether there is a causal impact of cognitive control engagement on the resolution of non-conflict referential competition.

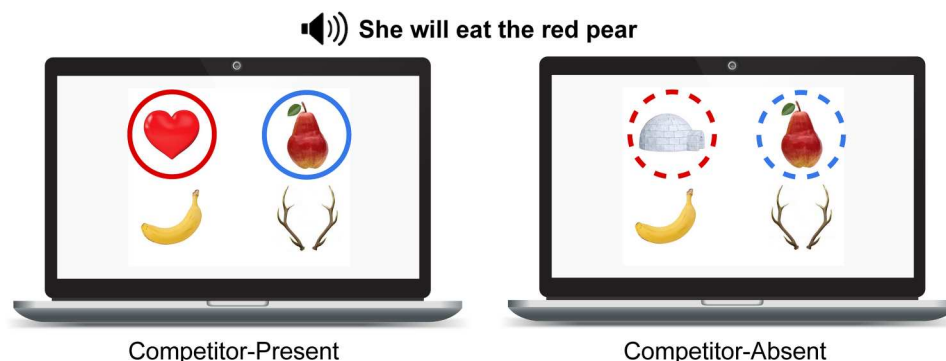


Figure 1. Example of a Competitor-Present (A) and Competitor-Absent (B) scene during a sentence listening trial. In this example, “She will eat the red pear”, the Target referent is the pear (circled in solid blue in the Competitor-Present scene and dashed blue in the Competitor-Absent scene). The Competitor is the red heart (circled in solid red), which is replaced with a Non-Competitor igloo in the Competitor-Absent condition (circled in dashed red). Circles are for illustration purposes only; participants did not see them.

The current study

We tested whether the engagement of cognitive control, as reflected in cross-task adaptation, would be observed in forms of referential ambiguity that do not involve conflict. We combined the cross-task adaptation technique with sentence listening in a visual-world paradigm (Hsu et al., 2021; Hsu & Novick, 2016; Patra et al., 2023; Thothathiri et al., 2018). Participants listened to sentences like “She will eat the red pear” while viewing scenes consisting of four objects (Kukona et al., 2014; Nozari et al., 2016). In the Competitor-Present condition (Figure 1A), two of the four objects were compatible with the adjective (e.g. *pear* as Target, and *heart* as Competitor). We expected to find looks to both the Target and the Competitor within the critical adjective time window, consistent with competition for reference, as observed previously. In the Competitor-Absent condition (Figure 1B), participants heard the same sentences but viewed a scene containing only one adjective-compatible object (e.g. only the pear; the red heart was replaced by a white igloo). In this condition, we expected looks to fall predominantly on the Target object, soon after the verb, due to the lack of referential competition.

In order to manipulate the engagement status of cognitive control, we interleaved sentence comprehension trials with trials of the Stroop task. Incongruent Stroop trials were designed to engage cognitive control at a higher level than Congruent trials. If cognitive-control engagement modulates the resolution of referential ambiguity, then the competition between the Target and Competitor objects should be attenuated (more looks to the Target during the adjective window) when control is upregulated, following Incongruent versus Congruent Stroop trials in the Competitor-Present Condition (cross-task adaptation of cognitive control). Alternatively, this form of ambiguity may not engage cognitive control, due to a lack of representational conflict, in which case we expect no attenuation of the competition effect following Incongruent Stroop.

We tracked comprehenders’ developing interpretations of each sentence with a variant of the visual-world paradigm: webcam eye-tracking, which we dub the *remote visual world paradigm* (Ovans et al., 2022a, 2022b). Participants completed the experiment remotely, with stimuli presented on their Internet-connected home computers. Trained research assistants hand-coded participants’ looks to the four quadrants in the visually presented scenes from frame-by-frame playback of the webcam video recordings. The webcam eye-tracking method holds significant promise for conducting

eye-gaze studies with remote participation, which could yield larger and more diverse samples, and enable operation through laboratory disruptions like the COVID-19 pandemic.

The experiment did not use an automatic function for coding eye position over time, even though recent studies have experienced success with this approach (Degen et al., 2021; Prystauka et al., 2024; Slim & Hartsuiker, 2023; Vos et al., 2022). Prystauka et al. (2024) observed subtle referential competition effects like those of Kukona et al. (2014) in webcam-recorded data, using the automatic webcam eye-tracking functionality in the Javascript library Webgazer.js (Papoutsaki et al., 2016). Although Webgazer.js has grown in popularity to replicate a number of in-lab effects, we encountered issues that reduced the feasibility of using its eye-tracking functionality in its current implementation. The calibration function in Webgazer requires significant computational resources on participants’ computers in order to operate in real-time, which may render some participants unable to complete calibration. In addition to concerns about the functioning of the calibration algorithm, we discovered additional issues related to the accurate recording of latency information during internet-based experiment administration, which could be mitigated with a hand-coded approach but not in the current implementation of Webgazer. The challenges of using automatic eye-tracking on webcam data are discussed in more detail in the Method and Discussion.

We opted to code gaze direction by hand. Previous studies have demonstrated that hand-coding gaze direction, collected by video camera, is feasible to observe sentence processing effects in a laboratory-based, visual-world setting (e.g. Snedeker & Trueswell, 2004). However, it remains unknown whether human-coded gaze direction is sufficiently accurate to yield sensitivity to the subtle referential ambiguity effects that were expected here, especially when collected via webcam over the internet. The remote webcam participation arrangement introduces several sources of variability, including participants’ screen size, camera quality, lighting, internet bandwidth, and environmental distractions, which might introduce noise in the data that is absent during in-lab testing.

Overall, the study had the following objectives: (1) To test whether cognitive-control engagement directly impacts the resolution of non-conflict referential ambiguity, and (2) To determine whether our remote visual world paradigm with manual coding is sensitive to subtle competition effects, despite numerous departures from the conditions and procedures of in-lab eye-tracking.

Method

Participants

Eighty-seven native English-speaking volunteers from the University of Maryland participated for course credit (23 men, 61 women, 3 transgender/non-binary; mean age = 18.86 years, range = 18–22 years). We excluded participants who did not complete the Stroop trials ($N=4$) or had poor video quality that made coding difficult ($N=18$). Sixty-five participants were included in the final analyses. All procedures were approved by the University of Maryland Institutional Review Board.

Materials

Stroop

Print colours included blue, green, and yellow. Colour words included *blue*, *green*, *yellow*, *brown*, *red*, and *orange*. For the 24 Congruent trials, print colour matched the colour words (e.g. “blue” displayed in blue), while the colour and word did not match for the 24 Incongruent trials (e.g. “red” displayed in blue).

Sentences

Each stimulus sentence consisted of a subject noun phrase, a main verb, and a noun phrase that began with an adjective (full list of sentences in Supplementary Appendix A). Sentences were modelled after Kukona et al. (2014) and Nozari et al. (2016). They were recorded in a sound-attenuated booth by a female native-English speaker using consistent intonation across sentences, which was ensured through visual inspection of the waveform to minimise inferences from prosodic cues (e.g. Kurumada et al., 2014). The sampling rate was 44.1 kHz.

Scenes

A visual scene consisting of four objects in separate quadrants was associated with each sentence (Figure 1). In the Competitor-Present condition, the scene included a Target object that met the selectional restrictions of the main verb and the adjective (e.g. a red pear), and a Competitor object that met the restrictions of only the adjective (e.g. a red heart). Scenes in the Competitor-Absent condition were identical to the Competitor-Present scenes, except that the Competitor was replaced by an object that was incompatible with both the verb and the adjective (e.g. a white igloo; see Table 1). Each scene also contained a second object that was a plausible argument for the verb but not the adjective (verb competitor, e.g. “banana”); this followed the studies of

Table 1. Visual objects accompanying the sentence “She will eat the red pear”, by condition.

Scene type	Target	Competitor / non-competitor	Verb competitor	Unrelated
1 Competitor-Present	Pear	Heart	Banana	Antlers
2 Competitor-Absent	Pear	Igloo	Banana	Antlers

Kukona et al. (2014) and Nozari et al. (2016). However, we did not manipulate the verb competitor as a factor within the design as it was not relevant to our hypotheses.

Objects were rated by a separate group of participants to confirm a strong semantic fit between the adjective (e.g. red) and both the Target (e.g. pear) and Competitor (e.g. heart), and also to confirm a poor semantic fit between the verb (e.g. eat) and Competitor (e.g. heart) (see Supplemental Appendix B).

Design

Participants listened to sentences accompanied by visual scenes, which were interleaved with trials of the Stroop task in a 2×2 factorial design (as depicted in Figure 2). There were 48 critical Stroop-to-Sentence sequences, each of which could occur with two levels of Stroop condition (Incongruent vs. Congruent), and with two levels of referential scene condition: either an adjective competitor or a non-competitor object (Competitor-Present vs. Competitor-Absent). Thus, there were four critical types of Stroop-to-Sentence sequences, each occurring 12 times: Stroop-Congruent + Competitor-Present; Stroop-Congruent + Competitor-Absent; Stroop-Incongruent + Competitor-Present; and Stroop-Incongruent + Competitor-Absent. The critical Stroop-to-sentence sequences were surrounded, but never interrupted, by 48 filler sentences and 72 filler Stroop trials, which reduced the possibility of predicting the next trial type at any point in the experiment. The four possible sequences, Stroop-Stroop, Stroop-Sentence, Sentence-Stroop, and Sentence-Sentence, occurred within the experiment at frequencies of 20%, 35%, 35%, and 10%, respectively. Filler sentences had the same structure and complexity as the stimulus sentences (e.g. “She will order the nice steak”), but all objects within the scene were not constrained by either the verb or the adjective. Four experimental lists were created. Each list presented $\frac{1}{4}$ of the critical sentences in each of the four experimental condition configurations, and each sentence was assigned to all four conditions across the four lists (Latin Square design). A pseudo-random ordering was established

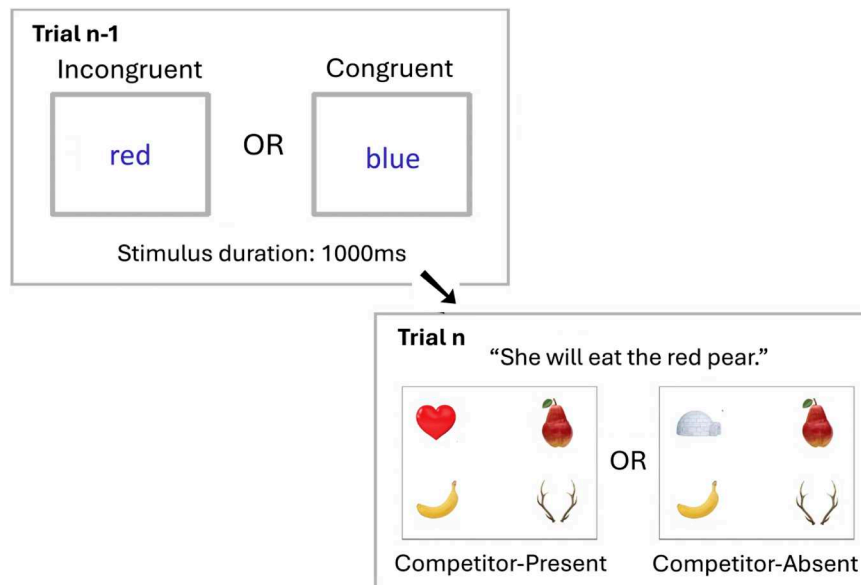


Figure 2. Experimental design. Incongruent or Congruent Stroop trials were presented with a duration of 1000 ms, which preceded sentences accompanied by a Competitor-Present or Competitor Absent scene.

for the four lists, and this order was then reversed to create four additional order-reversed lists, yielding eight total lists.

Procedure

Experimental sessions were run on the Penn Controller for Internet-Based Experiments platform (PCIBex; Zehr & Schwarz, 2018). Participants completed the experiment remotely via the web browser. After participants provided consent and access to their webcam for recording, the browser automatically entered the full screen. Participants were shown the output of their webcam video and instructed to position their heads in the centre of the video with their eyes unobstructed and well-lit during the experiment. Participants were told to make sure their speakers or headphones were playing sound. To check whether the video feed was mirror-reversed, participants were instructed to look at an image on one side of the screen, and say either "left" or "right". Participants then read instructions for the Stroop and sentence listening tasks. Following the instructions, they practiced both tasks in the same block (five Stroop trials, seven sentence listening trials).

Stroop trials

Each Stroop trial began with the presentation of a black cross, which disappeared after the participant clicked it using their mouse/trackpad, initiating the presentation of a colour word (stimulus duration 1000 ms), above the cross. Participants were instructed to name the colour of the print aloud, quickly and clearly. Responses

were recorded by the microphone built into their computer or headphones. Participants received no feedback about response accuracy.

Sentence comprehension trials

Each sentence comprehension trial began with the presentation of the visual scene. After the scene had been displayed for 1000–1500 ms (mean = 1382 ms), auditory sentence playback was initiated. The range of durations for the scene preview period stemmed from variability in the initiation of sentence playback on participants' computers; the duration of the preview period was random and uncorrelated with experimental conditions. After sentence playback was completed, participants were instructed to click on the object within the visual scene that matched the spoken description. After clicking, or after 2 s without response following playback completion, a yes/no comprehension question about the sentence content appeared on-screen.

Data coding

Stroop response times (RT) were extracted by automatically identifying the onset of human speech within each participant's Stroop-trial audio recording (Matlab's detectSpeech function). Stroop accuracy was documented by human coders, who listened to each audio recording.

Looks to the objects within the scene from the critical sentence trials were coded by four research assistants, following a method developed by Snedeker and

Trueswell (2004). Each data coder viewed the participant's webcam video frame by frame using Datavyu (<https://datavyu.org/>), monitored for changes in gaze direction (mainly saccades), and recorded the time and gaze direction at the beginning of each new eye fixation from the coder's perspective. Gaze direction was coded into seven categories – “top left”, “top right”, “bottom left”, “bottom right”, “center”, “offscreen”, and “unsure” (Figure 3). The onset of a code was where the eye-gaze landed (thus saccades were absorbed by the previous fixation). Blinks were marked when the eyelids were too closed for gaze direction to be discernible. Codes marked as “unsure”, “offscreen”, and “blinks” were labelled as trackloss. Coders were blind to the experimental condition for each video (they had no knowledge about the objects in the accompanying visual scene or about the preceding Stroop trial), which eliminated any possibility of hypothesis-driven bias in data coding. 50% of subjects were double-coded to check for reliability, and to confirm that there was no difference in coding strategy between coders. On average, coders agreed on gaze direction and onset time of changes in gaze direction (within ~100 ms) 90% of the time. The first author checked participants whose trials had less than a 90% match rate between codes, and made a decision on codes that disagreed.

Data preprocessing

Based on the coding of every change in gaze direction as described above, we next labelled each frame for gaze direction (e.g. upper left), by extrapolating the gaze direction label used for each new fixation to every subsequent frame until the next change in gaze direction. After each frame was labelled for gaze direction, we then converted the label from the coder's perspective to the participant's perspective, and added a label indicating which of the four objects was fixated (Target, Competitor, Non-Competitor, Verb Competitor, Unrelated), yielding a looks-to-objects time series. Note that

we assumed that a gaze into a particular quadrant of the scene (e.g. upper left) could be interpreted as a fixation on the stimulus object within that quadrant (e.g. Competitor), even though we did not track fixation location accurately enough to be certain that this was true (Snedeker & Trueswell, 2004). Any trial with more than 20% track loss was excluded from the analysis. We excluded 165 trials (5%), which left 3113 trials in the final analysis. In order to quantify the degree of referential competition for each object, we calculated the proportion of looks to each object out of all looks to either the object and its ostensible competitor. For instance, we calculated the following ratios:

$$\begin{aligned} &\text{Proportion of Looks to Competitor} \\ &= \frac{\text{Looks to Competitor}}{\text{Looks to Competitor} + \text{Looks to Target}} \end{aligned}$$

$$\begin{aligned} &\text{Proportion of Looks to Non-competitor} \\ &= \frac{\text{Looks to Non-competitor}}{\text{Looks to Non-competitor} + \text{Looks to Target}} \end{aligned}$$

Preprocessed data and a guide to coding eye gaze are available at <https://osf.io/ngz5q/>.

In order to analyse the relationship between sentence playback and looks to objects in the scene, we coregistered the auditory sentence recording with the looks-to-objects time series. This step was necessary because random delays in the onset of sentence playback, described above, meant that we lacked an accurate record of sentence onset time in the raw data. We used Matlab's function `detectSpeech` to mark the onset of speech in the audio recording extracted from participants' computer microphones and to align the sentence audio playback with the looks-to-objects time series. For participants who used headphones to listen to sentence stimuli ($N = 18$), we had no audio recording of the sentence playback; for these participants, we instead used the onset time for each stimulus item averaged across participants who used their computer speakers. In order to allow finer granularity in the alignment of the adjective word onset with the video

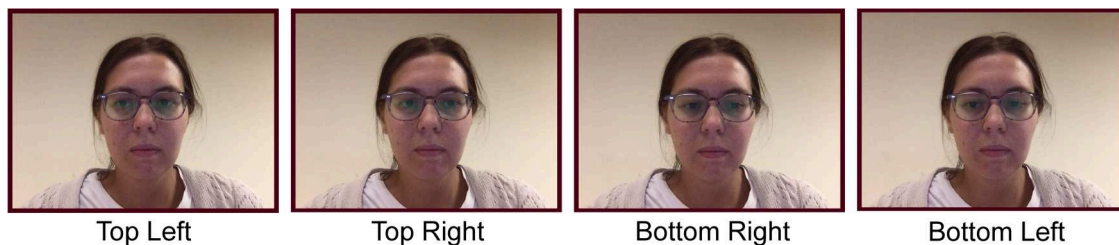


Figure 3. Still images of the first author looking at the four quadrants within a typical visual scene in the experiment. Gaze direction labels were marked from the coder's perspective, and then translated into the participant's perspective during data preprocessing (e.g. the top left label is a look to the top right object on the participant's screen).

recording, each looks-to-objects time series was upsampled to 1000 frames per second (fps). After coregistration, the looks-to-objects time series was downsampled from 1000 to 20 fps.

Statistical analyses

We used a cluster-based permutation test to assess whether the proportion of looks was greater in the Competitor in the Competitor-Present scenes than looks to the Non-competitor in the Competitor-Absent scenes (Ito & Knoeferle, 2023; Maris & Oostenveld, 2007). Our analysis performed a linear mixed effects model on the Competitor vs. Non-competitor difference in each 50 ms time sample from 0 to 1500 ms after the onset of the adjective, yielding a time-series of t -values. We then identified clusters (minimum cluster size = 2) of time-adjacent t -values corresponding to $p < .05$ (uncorrected) and calculated a cluster-level summed t -statistic for each cluster. This same procedure was performed on 5000 randomizations of the observed data, which were generated by randomly permuting condition labels (Competitor-Present, Competitor-Absent, Stroop-Congruent, and Stroop-Incongruent). We retained clusters whose cluster-level summed t -statistic had a Monte Carlo $p < .05$, relative to the null distribution generated by the random permutations.

In order to assess whether the Stroop trial type modulated the Competitor vs. Non-Competitor effect, we performed a cluster-based permutation test in the same

way as described above, except that the test statistic used to build clusters was the t -value associated with the Stroop type X Competitor Present interaction term in the linear mixed effects model.

Results

Stroop

As shown in Figure 4, response latencies were longer, and accuracy was lower, in Incongruent than Congruent Stroop trials. These observations were confirmed with paired t -tests [latency: $t(66) = -13.419$, $p < .001$; accuracy: $t(66) = 4.7614$, $p < .001$]. Such effects demonstrate that the Incongruent Stroop trials engendered conflict, as we predicted, even when administered remotely.

Referential competition

As can be seen in Figure 5, participants on average looked more at the Competitor (e.g. heart) in the Competitor-Present scenes compared to the Non-competitor (e.g. igloo) in the Competitor-Absent scenes within a window of time that began shortly after adjective onset and lasted until ~ 1200 ms after adjective onset. This pattern suggests that the Competitor was considered a potential referent, competing with the target object in the same scene (indeed, as can also be seen in Figure 5, looks to the Target were lower in the Competitor-Present scenes as compared to the Competitor-Absent scenes within the same time window). This

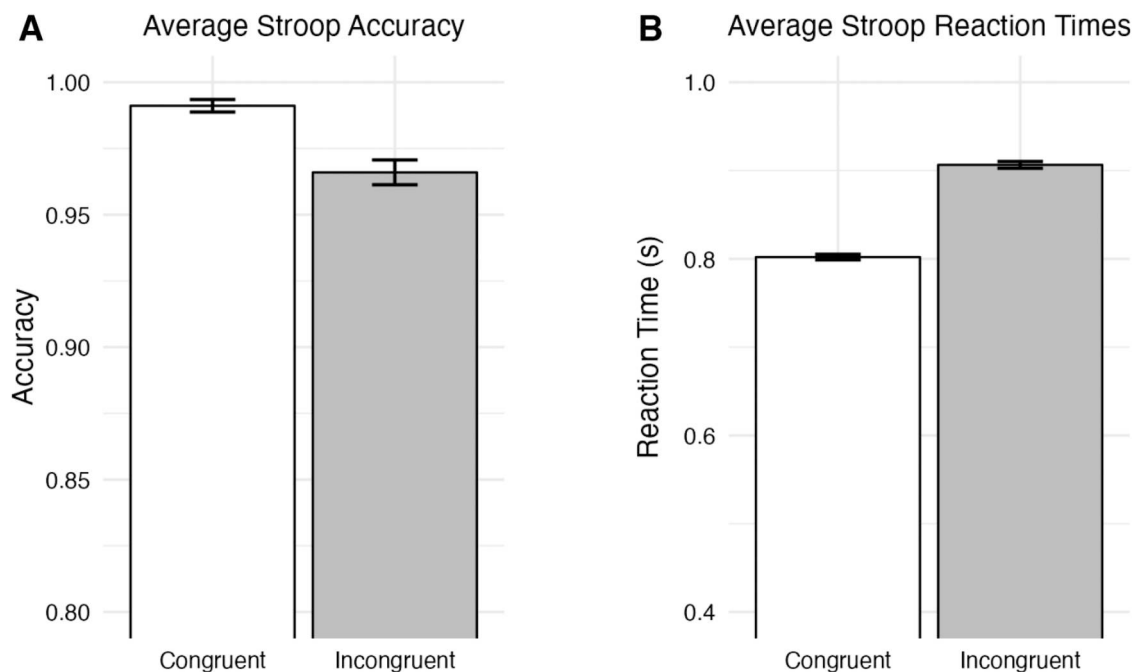


Figure 4. Accuracy (A) and response times (B) for the Stroop trials. Error bars represent 95% within-subject confidence intervals.

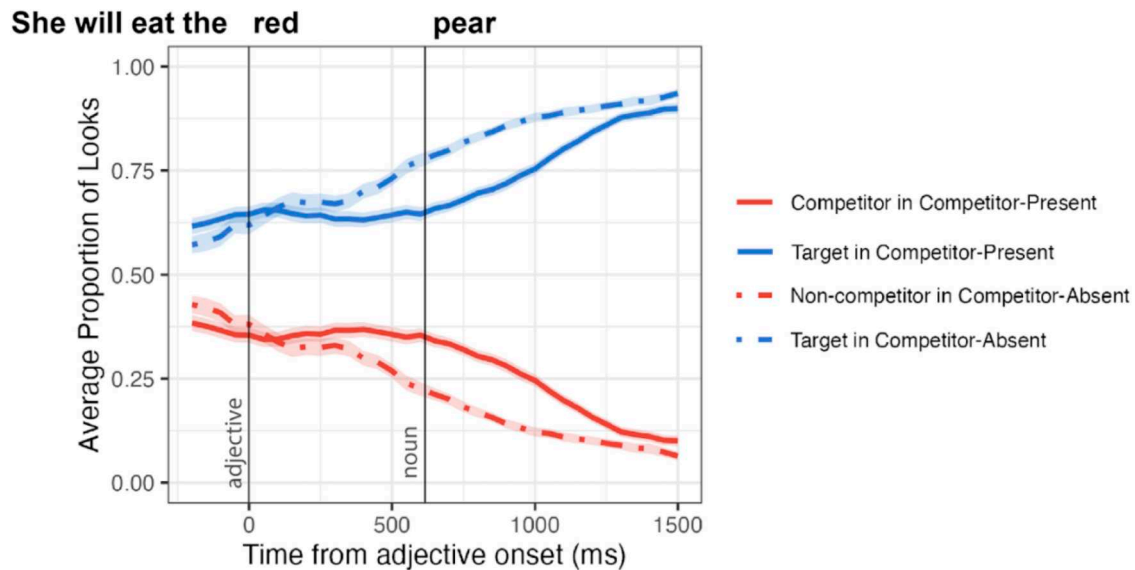


Figure 5. The average proportion of looks to each of the objects in their respective condition: the Competitor object in the Competitor-Present scenes (red solid line), the Non-competitor in the Competitor-Absent scenes (red dotted line), the Target object in the Competitor-Present (blue solid line) and Competitor-Absent (blue dotted line) scenes. Vertical black lines mark the onset of the adjective and noun (averaged across items) in the sentence. Shaded region represents 95% within-subject confidence interval.

difference resulted in a significant cluster of time bins in which proportions of looks to the Competitor exceeded those to the Non-Competitor, spanning 300 to 1200 ms post-adjective onset ($p < .001$). Importantly, this replicates subtle competition effects observed in the lab, including the time-course of the competition's rise and fall, using our remote visual world paradigm with hand coding of eye position over time (Kukona et al., 2014; Nozari et al., 2016).

Effect of prior Stroop type on online referential ambiguity resolution

We tested the hypothesis that cognitive-control engagement can aid the resolution of reference competition by examining looks at the Competitor and the Non-Competitor object following Incongruent and Congruent Stroop. If cognitive control does reduce competition effects, we would expect to see fewer looks to the Competitor object following Incongruent than Congruent Stroop. As can be seen in Figure 6A, however, there is no obvious impact of prior Stroop type on looks to the Competitor. A cluster-based permutation test on the interaction between Scene Type (Competitor-Present vs. Competitor-Absent) and prior Stroop trial type (Incongruent vs. Congruent) did produce a significant cluster from 1150 to 1350 ms. But this interaction was driven by the difference in the proportion of looks in the Competitor-Absent scenes. As shown in Figure 6B, there were more looks within this time window to the

Non-competitor (igloo) following Stroop-Incongruent trials ($p = .015$), but Stroop type did not modulate looks to the Competitor (heart) in Competitor-Present scenes (no clusters were identified).

Confirmatory Bayesian analysis

In view of the null interaction between the Prior Stroop trial type and Competitor-Present vs. Absent scenes, we assessed the ratio of the likelihood of the null interaction compared to the likelihood of the alternative hypothesis with a post hoc Bayesian analysis to create credible intervals and Bayes factors (BF; Wagenmakers et al., 2010; using the *brms* package in R; Bürkner, 2017). A larger BF value (> 10) indicates a higher likelihood of the null hypothesis accounting for the data. In addition, following Dempsey et al. (2020), we evaluated the parameter values obtained from the model's posterior distribution by using the region of practical equivalence (ROPE), which is a range of parameter values for which the differences between the likelihood of the null and alternative hypotheses are too small to be considered distinguishable from one another. Here, we set the ROPE to a range of -0.1 to 0.1 of a standardised parameter (Kruschke, 2018). We found the shortest interval within the posterior distribution that had a 95% probability of containing the true parameter value (HDI; highest density interval). Using the 95% HDI, we calculated the percentage of the interval that lies within the ROPE. If the 95% HDI falls within the ROPE, there is minimal

She will eat the red pear

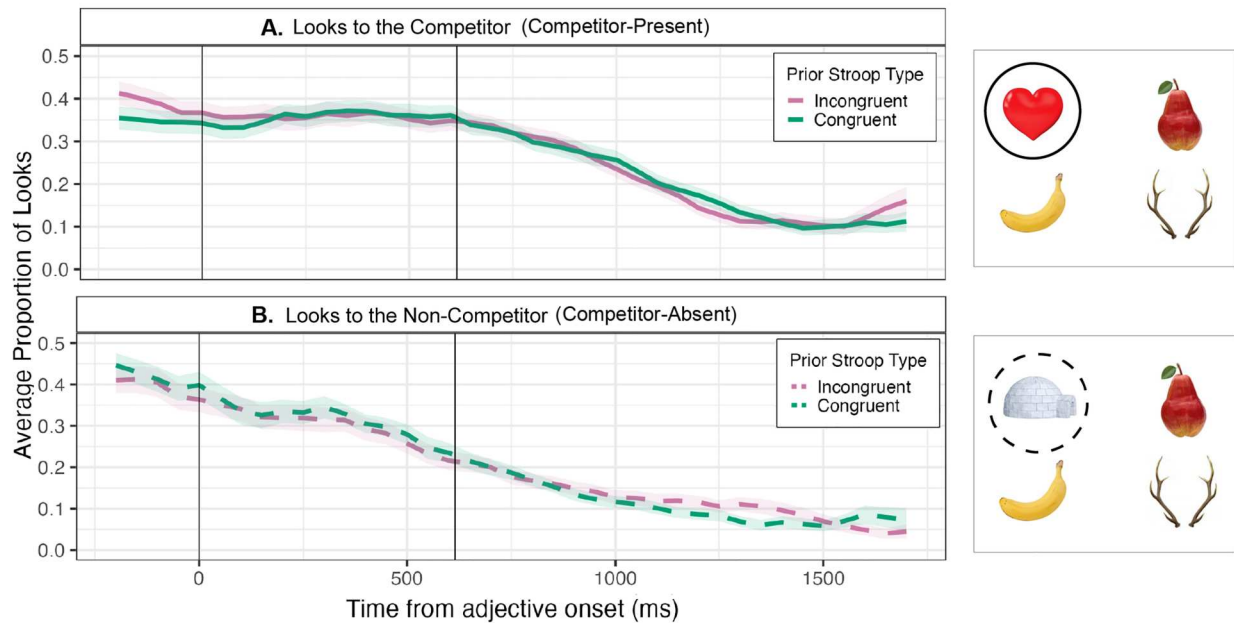


Figure 6. Average proportion of looks to the Competitor in the Competitor-Present (panel A, solid lines) and Competitor-Absent (panel B, dotted lines) scene type following Stroop-Incongruent trials (pink) and Stroop-Congruent trials (green). Shaded region represents 95% within-subject confidence interval.

evidence to support or reject the null hypothesis (Kruschke, 2015).

Our Bayesian model was fitted to the data with fixed effects of Competitor-Present vs. – Absent, prior Stroop trial type, the interaction, and a maximal random effects structure. We used an uninformative prior since we did not have any prior data regarding cognitive control engagement in sentences with referential ambiguity. As seen in Table 2, the model estimates that the null hypothesis is much more likely than the alternative, as indicated by the large BF value and 57% of the HDI in the ROPE, providing further evidence for no interaction between Stroop trial type and Scene type. In contrast, there is strong support for an effect of Scene type, which has a BF value of 0, and thus indicates a higher likelihood of a difference between Competitor-Present

and Competitor-Absent scenes. Furthermore, none of the HDI falls within the ROPE, indicating that the null and alternative hypotheses are distinguishable.

Discussion

We tested whether cognitive control is involved in the resolution of temporary referential ambiguities during spoken language comprehension. Participants listened to sentences like “She will eat the red pear”, while viewing a scene that included two red objects (e.g. a heart and a pear). The adjective *red* provided evidence that the heart could be the referent of an upcoming noun, but the accumulated linguistic evidence weighed probabilistically in favour of the pear, which was compatible with both the adjective and the preceding verb *eat*. Because one object was substantially more consistent with the available linguistic cues than the other, we described this type of ambiguity as lacking representational conflict.

Using a novel modification of the visual-world paradigm in which subjects participated remotely using their own computers, we found clear evidence of temporary referential ambiguity following the adjective *red*, reflected in looks to both adjective-compatible objects. We also found robust Stroop effects, suggesting that Incongruent trials reliably induced conflict. However, we found no evidence that cognitive control affected the resolution of this ambiguity: Manipulation of cognitive

Table 2. Bayesian model results.

Parameter	CI lower	CI upper	Bayes factor	Post. Prob.	% HDI in ROPE
Scene Type	0.73	1.04	0	0	0%
Prior Stroop Trial Type	−0.08	0.22	90.11	0.99	69%
Stroop × Scene Type Interaction	−0.07	0.21	414.8	1	57%

Notes: Credible interval upper and lower boundaries are given for each effect tested in the original cluster permutation analysis. To reject the null hypothesis, 100% of the HDI should be outside of ROPE (0% in ROPE). These results provide further support for a null interaction between Stroop and Scene Type, in line with the weak effect found in the permutation test. Post. Prob. = posterior probability; HDI = highest density interval; ROPE = region of practical equivalence.

control status, accomplished using the Stroop task, did not modulate the pattern of eye fixations on the Competitor (e.g. the heart) versus the Target (e.g. the pear) during the subsequent language task. Namely, we did not observe cross-task adaptation of cognitive control (Hsu & Novick, 2016; Ovans et al., 2022a; Thothathiri et al., 2018; see also Hsu et al., 2021; Ness et al., *in press*).

Sensitivity to referential competition

Our results demonstrate the sensitivity of a novel remote visual world paradigm, in which participation occurs over the internet and eye gaze direction is coded by hand from webcam video recordings, rather than an automatic camera-based eye-tracking system. We found that the time-course of eye movements captured by this paradigm closely matched the time-course found in lab settings; looks to the semantically incompatible object (e.g. the heart, which is incompatible with “eat”) started to increase approximately 250 ms after the onset of the adjective (e.g. “red”), and fall around 1200 ms. Crucially, our findings clearly show that the method is capable of detecting subtle referential competition effects and delineating the time-course of these effects, which replicate in-lab observations (Kukona et al., 2014; Nozari et al., 2016).

Although in-lab eye-tracking with automatic, computerised estimation of eye position is the gold standard tool for experimental research involving eye movements, our study demonstrates that the level of eye-tracking precision provided by laboratory eye trackers may not always be necessary when using a visual world paradigm (Prystauka et al., 2024; Slim & Hartsuiker, 2022; Snedeker & Trueswell, 2004; but see Ovans et al., 2021). The key dependent measures generally involve distinguishing among looks to a small number of objects that are clearly separated within a scene. These sorts of differences in looking behaviour are relatively easily coded by human inspection of webcam video. Our paradigm does not replace automatic eye tracking, but appears to provide a highly sensitive alternative under specific conditions, and may be valuable for research in a variety of contexts in which in-person participation is not possible. Future research is needed to determine whether other subtle psycholinguistic effects at the word and sentence levels can also be replicated in the remote environment.

No effects of cognitive control on referential competition

We manipulated the state of cognitive control during sentence listening by interleaving sentence processing

trials with trials of the Stroop task. Previous findings have demonstrated that Incongruent Stroop trials will engage cognitive control more than Congruent trials, with impacts that carry over into the following sentence trial when the linguistic input requires the resolution of conflicting interpretations (Hsu & Novick, 2016; Ovans et al., 2022a; Thothathiri et al., 2018; see also Hsu et al., 2021; Ness et al., *in press*).

Our manipulation of cognitive control did not impact the referential competition between the Competitor and the Target. Looks to the distractor object (heart) were no different following Incongruent and Congruent Stroop trials. A post-hoc Bayesian analysis concluded that there was strong evidence in favour of the null hypothesis over an alternative hypothesis that cognitive control processes influenced referential ambiguity resolution.

The failure of the cognitive control manipulation to impact sentence processing is unlikely to stem from a lack of effectiveness of our cognitive control task when deployed within the remote visual world paradigm. Participant responses on the Stroop trials showed robust congruency effects, in both reaction times and accuracy, which demonstrates that our method is capable of modulating cognitive control engagement, as it has in previous cross-task studies conducted in the lab (Hsu & Novick, 2016; Ovans et al., 2022a; Thothathiri et al., 2018).

Our explanation of this result pattern is that non-conflict ambiguities of the sort observed here do not necessarily engage cognitive control. Although the Competitor clearly competes with the Target, as indicated by the eye gaze patterns, its overall support in the linguistic input is weak since it does not satisfy the semantic constraints of the verb “eat”. In this situation of competition, cognitive control is not needed to resolve the temporary referential ambiguity. Instead, the availability of multiple sources of linguistic evidence here allows rapid resolution of the ambiguity in favour of the more plausible pear-as-referent interpretation. Thus, evidence of *competition* is not necessarily a predictor of *conflict* and the need for cognitive-control engagement.

Our results are different from – *but are not at odds with* – previous findings showing that increased cognitive control does affect the resolution of representational conflicts such as those requiring syntactic revision, or those with strong probabilistic syntactic and semantic cues that point toward opposing interpretations (Hsu et al., 2021; Hsu & Novick, 2016; Ovans et al., 2022a; Thothathiri et al., 2018). In these situations, the longer-to-resolve conflict between two evidentially supported but incompatible interpretations of a sentence is detected by a monitoring system, which recruits cognitive control to bias processing in favour of the more plausible analysis to resolve the conflict (Ness et al., *in*

press). Thus, our findings, taken together with previous findings, help establish the boundary conditions of our proposal that cognitive control aids language comprehension, by mapping the situations where cognitive control is not involved, contrasting with the previous findings showing where cognitive control is involved.

In fact, there is suggestive evidence from previous work that cognitive control may not be involved during referential ambiguity resolution (January et al., 2009), consistent with our view, and with our observations. January et al. (2009) used fMRI to test whether, within individuals' brains, the same prefrontal regions that are involved in resolving Stroop conflict are also involved in resolving linguistic conflict during sentence comprehension. They showed that when syntactically ambiguous sentences generated conflict between two interpretations (e.g. "Clean the pig with the leaf", in which the modifier and instrument interpretations of "with the leaf" were both contextually supported by the visual scene), these activated areas of prefrontal cortex that overlapped significantly, within individuals, with the areas activated by Stroop (see also Hsu et al., 2017). However, when sentences were referentially ambiguous about which object was intended, until the final noun arrived (e.g. "Clean the pig that has the leaf", when the visual scene contained a pig holding a leaf and a pig holding a rock), this activated brain areas that *did not* overlap with the areas activated by Stroop. Together, this indicates that the general cognitive-control mechanism that is recruited to resolve syntactic conflict is not also recruited to resolve referential ambiguity.

Some possible alternative explanations to consider

There are potential alternative explanations of our results, that warrant consideration. We will discuss two specific types of alternative accounts. Although we will argue that both accounts are inadequate to explain our results, we acknowledge that additional future work is needed before a complete understanding of the phenomena observed here is viable.

Possible limitations in sensitivity: Given the novelty of our remote visual world paradigm, it is conceivable that our findings reflect limitations in the sensitivity of this paradigm to detect the effects of cognitive control engagement on language processing. However, we reiterate that our method was able to manipulate cognitive-control engagement successfully, as reflected in robust Stroop effects, and our approach to measuring eye-movements provided clear sensitivity to subtle referential competition effects in the same time

window that has previously been observed in laboratory experiments with computerised eye-tracking (Kukona et al., 2014; Nozari et al., 2016). Thus, our methodology appears to have the sensitivity it needs to detect cognitive control effects on referential ambiguity resolution.

We also note that as-yet unpublished work by our group has successfully demonstrated that the remote visual world paradigm is sensitive to cross-task adaptation of cognitive control, when sentence processing requires the resolution of representational conflicts (Ovans et al., 2022c). For example, five-year-olds heard globally ambiguous sentences like "VERB the elephant with the carrot", featuring either instrument-biased (e.g. "Poke") or modifier-biased verbs (e.g. "Choose"). In both cases, two interpretations were simultaneously plausible, and thus in conflict: "with the carrot" could describe which elephant in the scene to choose/poke (the one holding a small carrot, not the one wearing a bowtie), or the instrument the child should use to perform choosing or poking (a standalone large carrot). Before the sentences, Flanker trials manipulated children's level of cognitive control, and their eye movements towards referents in the scene were then measured to gauge ongoing parsing preferences using the same remote visual world paradigm we report here. When sentences contained instrument-biased verbs, children's looks to instruments *increased* after incongruent compared to congruent Flanker trials. When sentences contained modifier-biased verbs, looks to instruments *decreased* after incongruent compared to congruent Flanker trials. These findings suggest that heightened cognitive control, engaged by incongruent Flankers, influenced children's reliance on reliable cues to guide parsing decisions (see Ness et al., *in press*). Moreover, they indicate that the remote visual world paradigm is sufficiently sensitive to reveal significant effects of cross-task adaptation of cognitive control.

Thus, although we cannot rule out the possibility that the remote visual world paradigm is limited in some ways, our work so far suggests that it is fully capable of detecting the subtle effects necessary to observe cognitive control impacts on language processing. Alongside our significant findings of Stroop effects and subtle referential competition effects in the current study, we assert that methodological limitations do not adequately account for the absence of cross-task effects we report.

Potentially contradictory findings: Contrary to our conclusions, one study concluded that cognitive control does play an important role in resolving the sort of referential ambiguity that we observed here (Nozari et al., 2016). The evidence supporting this conclusion was a

significant correlation between individual participants' conflict effect on a Flanker task (Incongruent RT minus Congruent RT), used as a measure of cognitive control ability, and the size of the referential competition effect (proportion of looks to the Competitor (heart) minus the proportion of looks to the non-Competitor (igloo)). This finding could indicate that individual differences in cognitive control ability predict comprehenders' ability to resolve referential ambiguities.

Although these findings are certainly intriguing, they may not directly contradict our theoretical conclusions. This is because averaged individual congruency effects on tasks like Flanker and Stroop does not provide reliable or stable measures of individual cognitive control ability (Bender et al., 2016; Feldman & Freitas, 2016; Hedge et al., 2018; James et al., 2018; Ward et al., 2001), as the status of cognitive control varies dynamically from moment to moment. With this caveat in mind, we performed a post-hoc test analogous to that of Nozari et al. (2016), which examined whether the magnitude of the Stroop RT Congruency effect for individual participants, averaged across the experiment, was correlated with their average competition effect (heart vs. igloo). We found no relationship between the two difference scores ($t(61) = -0.22$, $p = .827$, see Appendix C).

We note that cross-task adaptation of cognitive control, as found in previous research, relies on *within-person* effects. Incongruent Stroop or Flanker trials consistently create conflict, which boosts cognitive control compared to congruent trials. These effects are reliable because the experimental manipulation has induced brief changes in an individual's state of cognitive control. That is, it is possible to experimentally manipulate the state of cognitive control within a person at any moment, even though the state is not stable across time in a way that allows for the measurement of trait-level ability. Therefore, studies that manipulate the state in this way, rather than assuming consistent performance over time, address the concerns raised above.

Logistical conclusions about hand-coded gaze direction

We hand-coded gaze direction from webcam video recordings, rather than using an automatic gaze tracking software. Our results confirm that hand-coding of gaze direction can detect subtle referential competition effects, even when eye movements are recorded via webcam, as has been reported recently for automatic gaze direction (Prystauka et al., 2024). Our hand-coded approach is more labour-intensive than automatic gaze

tracking, so it is important to explain why we took such an approach.

One reason was that, in pilot work, we were frequently unable to run the required calibration function in Webgazer.js, which collects samples of gaze to known locations from each participant in order to constrain a model that estimates the direction of arbitrary gazes. At least one other study has reported difficulty with calibration in a significant subset of participants, who were then redirected to other experiments (Slim & Hartsuiker, 2023). One way to overcome this challenge is to decrease the precision of the calibration threshold to allow the calibration procedure to complete; but this comes at the cost of lower accuracy in gaze tracking. Our difficulty running Webgazer's calibration function may be rooted in its computational demands and may also be related to variability in lighting conditions and camera quality in the participant's environment.

A second rationale for our approach was that it included a co-registration of the sentence playback with the video record of participants' gaze, as discussed above (see Method). This co-registration was necessitated by a lag between the function call for sentence playback and the actual onset of playback, which varied within and between participants. These variable delays in stimulus timing, unless adjusted for, will lead to inaccuracies in the alignment of sentence stimuli and the record of looks within a scene. Neither PCIBex nor Webgazer.js currently have a method for such corrections (Slim & Hartsuiker, 2023), but some paid platforms provide additional metrics on the actual onset of audio playback (Prystauka et al., 2024).

Overall, remote automatic gaze analysis is promising, but is also associated with technical challenges in current implementations. Our hand-coded approach provides an accurate and practical alternative until those challenges are resolved.

Conclusions

Our findings indicate that cognitive control may not play a critical role in the resolution of referential ambiguity, which is common during language comprehension. Building on prior work that has shown cognitive control's impact on facilitating the resolution of conflict during sentence processing, this finding helps map the boundary conditions of the role of cognitive control during language comprehension. Our general conclusion from the overall pattern of findings is that although language contains many instances of *competition* between alternative representations of the input, only some of them lead to linguistic *conflict* and require the assistance of cognitive control. Taken

together, delineating the language processing conditions that do and also do not engage cognitive control allows us to advance closer toward a more complete model of sentence comprehension (Ness et al., *in press*).

Methodologically, our findings contribute to a growing body of psycholinguistics research using webcam-based eye-tracking within the visual world paradigm, allowing subjects to participate remotely rather than in a lab equipped with a costly eye-tracking system. A major strength of this approach is that it increases the accessibility and diversity of the participant pool, enabling the recruitment of individuals who are outside the convenience sample of Western, Educated, Industrial, Rich and Democratic (WEIRD) contexts (Henrich et al., 2010), or are unable to come into the lab. Relatedly, larger sample sizes can be obtained, enabling research into new questions such as the investigation of individual differences, which generally require large sample sizes. In sum, expanded development of internet-based experiments could enable valuable and more naturalistic types of investigations that are not possible in laboratory settings.

Acknowledgements

We thank Jenna Ringold, Hanna Levy, Delenne Phan, and Joshua Yoo, for assistance with data coding. We are also grateful to Zoe Ovans, who provided guidance on data collection and coding.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Science Foundation (NSF) [grant #2020932] to Jared Novick and National Science Foundation (NSF) [grant #2020490] to Albert Kim.

Data availability statement

The data reported here are available at <https://osf.io/ngz5q/>.

References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Bender, A. D., Filmer, H. L., Garner, K. G., Naughtin, C. K., & Dux, P. E. (2016). On the relationship between response selection and response inhibition: An individual differences approach. *Attention, Perception, & Psychophysics*, 78(8), 2420–2432. <https://doi.org/10.3758/s13414-016-1158-8>
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652.
- Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Degen, J., Kursat, L., & Leigh, D. D. (2021). Seeing is believing: Testing an explicit linking assumption for visual world eye-tracking in psycholinguistics. In T. Fitch, C. Lamm, H. Leder, & K. Tessmar-Raible (Eds.), *Proceedings of the 43rd Annual Conference of the Cognitive Science Society* (Vol. 43, pp. 1500–1506). Austin, TX: Cognitive Science Society.
- Dempsey, J., Liu, Q., & Christianson, K. (2020). Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(10), 1906–1921. <https://doi.org/10.1037/xlm0000881>
- Feldman, J. L., & Freitas, A. L. (2016). An investigation of the reliability and self-regulatory correlates of conflict adaptation. *Experimental Psychology*, 63(4), 237–247. <https://doi.org/10.1027/1618-3169/a000328>
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164–203. [https://doi.org/10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7)
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3), 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302), Article 7302. <https://doi.org/10.1038/466029a>
- Hsu, N. S., Jaeggi, S. M., & Novick, J. M. (2017). A common neural hub resolves syntactic and non-syntactic conflict through cooperation with task-specific networks. *Brain and Language*, 166, 63–77. <https://doi.org/10.1016/j.bandl.2016.12.006>
- Hsu, N. S., Kuchinsky, S. E., & Novick, J. M. (2021). Direct impact of cognitive control on sentence processing and comprehension. *Language, Cognition and Neuroscience*, 36, 211–239. <https://doi.org/10.1080/23273798.2020.1836379>
- Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science*, 27(4), 572–582. <https://doi.org/10.1177/0956797615625223>
- Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, 55, 3461–3493. <https://doi.org/10.3758/s13428-022-01969-3>
- James, A. N., Fraundorf, S. H., Lee, E. K., & Watson, D. G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language*, 102(June), 155–181. <https://doi.org/10.1016/j.jml.2018.05.006>
- January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, 21(12), 2434–2444. <https://doi.org/10.1162/jocn.2008.21179>
- Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (Edition 2). Academic Press.

- Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, 1(2), 270–280. <https://doi.org/10.1177/2515245918771304>
- Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 326–347. <https://doi.org/10.1037/a0034903>
- Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 133(2), 335–342. <https://doi.org/10.1016/j.cognition.2014.05.017>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Ness, T., Langlois, V. J., Kim, A. E., & Novick, J. M. (in press). The state of cognitive control in language processing. *Perspectives on Psychological Science*.
- Novick, J. M., Thompson-Schill, S. L., & Trueswell, J. C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition*, 107(3), 850–903. <https://doi.org/10.1016/j.cognition.2007.12.011>
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3), 263–281. <https://doi.org/10.3758/CABN.5.3.263>
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, 4(10), 906–924. <https://doi.org/10.1111/j.1749-818X.2010.00244.x>
- Nozari, N., Trueswell, J. C., & Thompson-Schill, S. L. (2016). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychonomic Bulletin & Review*, 23(6), 1942–1953. <https://doi.org/10.3758/s13423-016-1068-8>
- Ovans, Z., Hsu, N. S., Bell-Souder, D., Gilley, P., Novick, J. M., & Kim, A. E. (2022a). Cognitive control states influence real-time sentence processing as reflected in the P600 ERP. *Language, Cognition and Neuroscience*, 37(0), 939–947. <https://doi.org/10.1080/23273798.2022.2026422>
- Ovans, Z., Huang, Y., & Novick, J. M. (2022b, March). Virtual-world eye-tracking: Replicating sentence processing effects remotely. Poster presented at the 34th Annual Human Sentence Processing Conference, Santa Cruz, CA.
- Ovans, Z., Huang, Y., & Novick, J. M. (2022c, November). Developmental parsing and cognitive control. Talk presented at the Psychonomic Society 63rd Annual Meeting, Boston, MA.
- Ovans, Z., Novick, J. M., & Huang, Y. (2021). Virtual-World eye-tracking: The efficacy of replicating word processing effects remotely. Short talk presented at the 34th Annual CUNY Conference on Human Sentence Processing, Philadelphia, PA.
- Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence (IJCAI'16)* (pp. 3839–3845). AAAI Press.
- Patra, A., Kirkwood, J., Middleton, E. L., & Thothathiri, M. (2023). Variation in how cognitive control modulates sentence processing. *Royal Society Open Science*, 10(4), 211969. <https://doi.org/10.1098/rsos.211969>
- Prystauka, Y., Altmann, G. T. M., & Rothman, J. (2024). Online eye tracking and real-time sentence processing: On opportunities and efficacy for capturing psycholinguistic effects of different magnitudes and diversity. *Behavior Research Methods*, 3504–3522. <https://doi.org/10.3758/s13428-023-02176-4>
- Slim, M. S., & Hartsuiker, R. J. (2023). Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIBex and WebGazer.js. *Behavior Research Methods*, 55(7), 3786–3804. <https://doi.org/10.3758/s13428-022-01989-z>
- Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49(3), 238–299. <https://doi.org/10.1016/j.cogpsych.2004.03.001>
- Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45(4), 447–481. [https://doi.org/10.1016/S0010-0285\(02\)00503-0](https://doi.org/10.1016/S0010-0285(02)00503-0)
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. <https://doi.org/10.1126/science.7777863>
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, 94(26), 14792–14797. <https://doi.org/10.1073/pnas.94.26.14792>
- Thothathiri, M., Asaro, C. T., Hsu, N. S., & Novick, J. M. (2018). Who did what? A causal role for cognitive control in thematic role assignment during sentence comprehension. *Cognition*, 178, 162–177. <https://doi.org/10.1016/j.cognition.2018.05.014>
- Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2), 89–134. [https://doi.org/10.1016/S0010-0277\(99\)00032-3](https://doi.org/10.1016/S0010-0277(99)00032-3)
- Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infra-red and webcam eye tracking in the visual world paradigm. *Glossa Psycholinguistics*, 1(1). <https://doi.org/10.5070/G6011131>
- Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, 60(3), 158–189. <https://doi.org/10.1016/j.cogpsych.2009.12.001>
- Ward, G., Roberts, M. J., & Phillips, L. H. (2001). Task-Switching costs, Stroop-costs, and executive control: A correlational study. *The Quarterly Journal of Experimental Psychology Section A*, 54(2), 491–511. <https://doi.org/10.1080/713755967>
- Zehr, J., & Schwarz, F. (2018). PennController for internet based experiments (IBEX). <https://doi.org/10.17605/OSF.IO/MD832>