

Sentence Processing is “Good Enough”: Evidence from Sentence-Video Matching

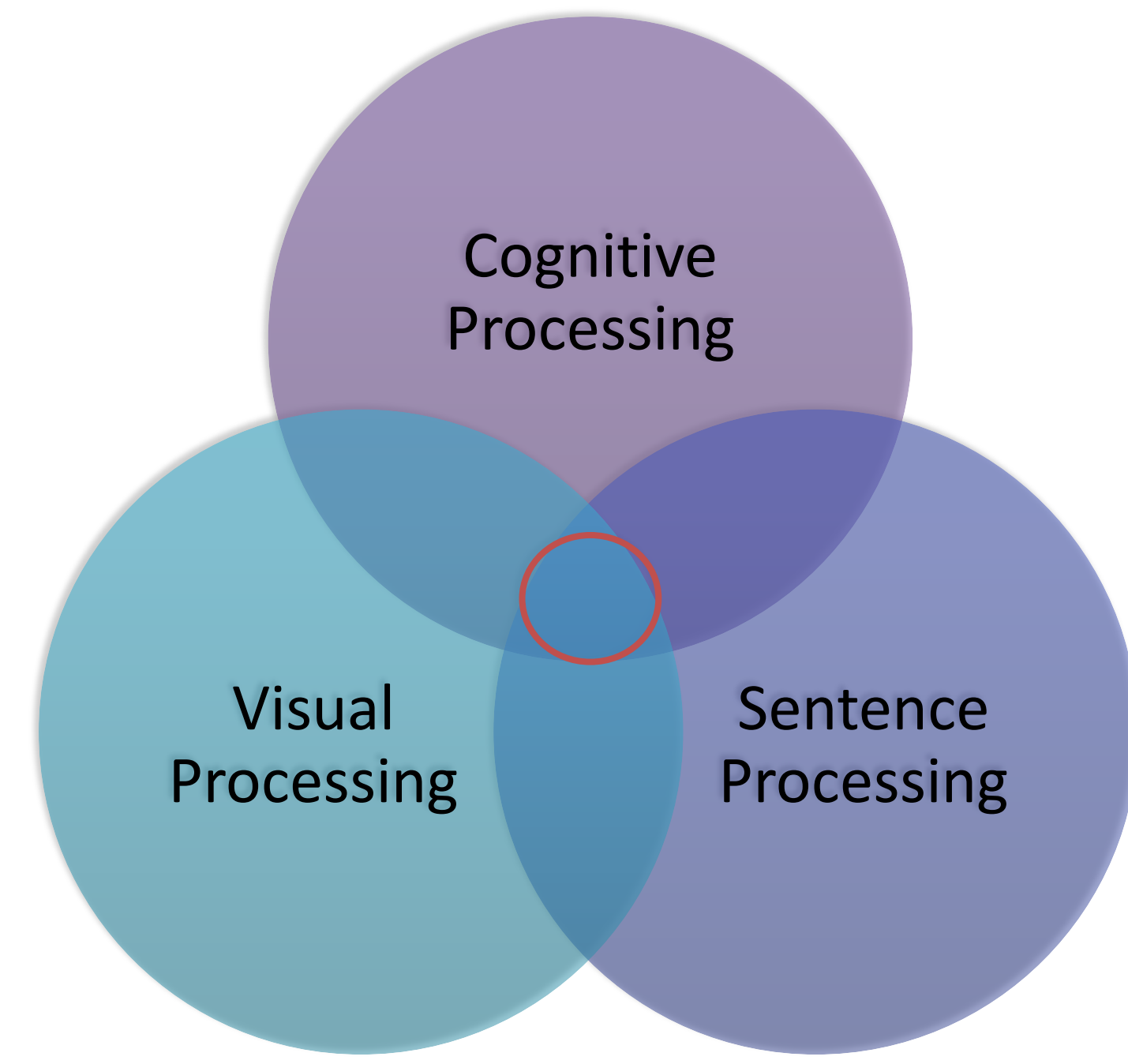
Gaurav Kharkwal^{1,2}, Karin Stromswold^{1,2}

kharkwal@rci.rutgers.edu karin@rucss.rutgers.edu

1. Department of Psychology, Rutgers University – New Brunswick
2. Center for Cognitive Science, Rutgers University – New Brunswick



RUTGERS



We investigated the level of detail encoded in linguistic representations of visual events. In two experiments, participants watched captioned videos and decided whether the captions accurately described the videos. Results indicate the level of detail encoded in the visual and linguistic representations is task-dependent.

Experiment 1- Actives Only

Task: Does the caption match the video?

Subjects: 25 English-speaking college-aged adults.

Independent Variables:

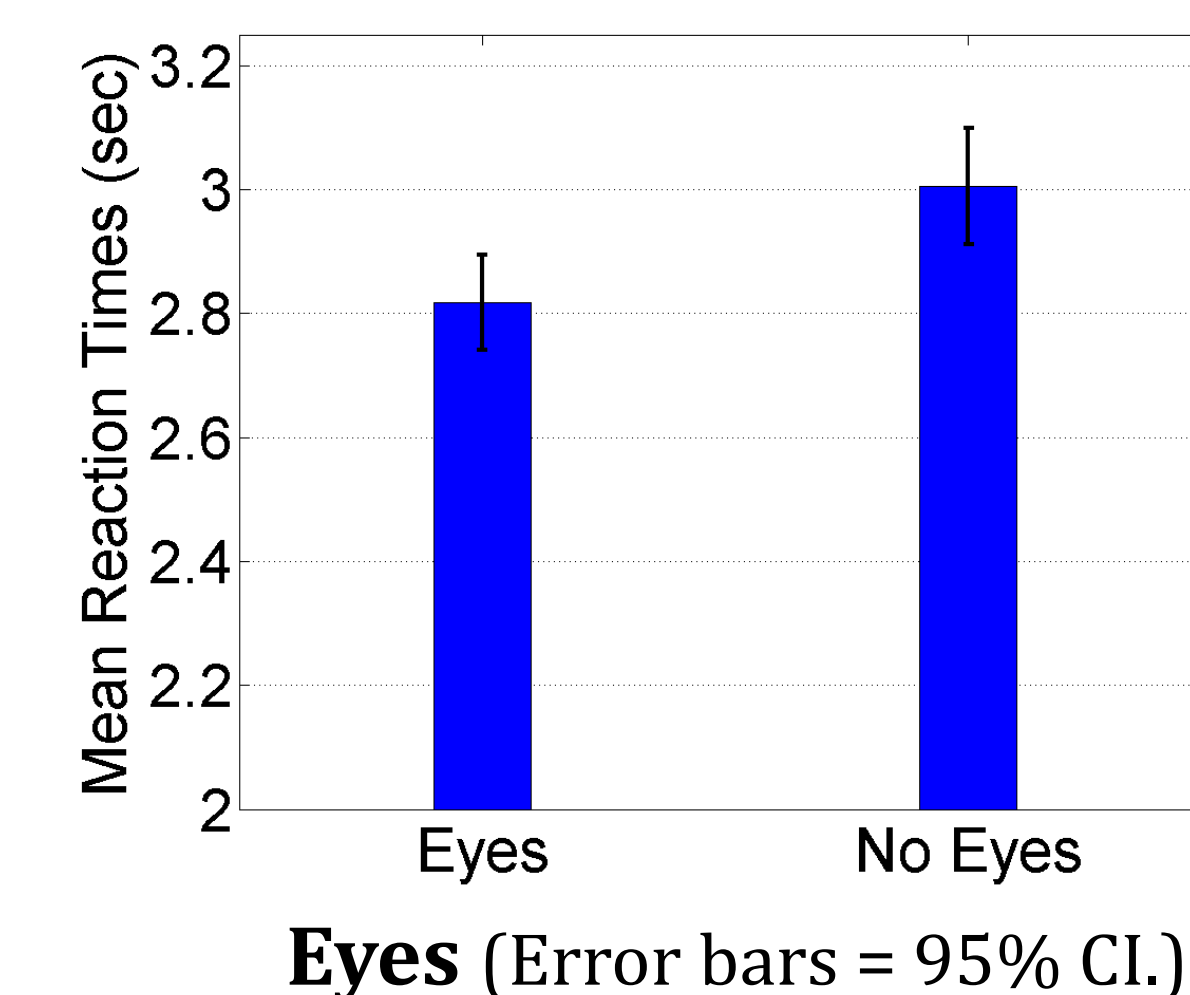
- Voice*:
 - Active only (e.g. *The circle is chasing the square.*)
- Verb:
 - Chase } “Source to Goal” ○ ⇒ □
 - Follow } “Goal to Source” ○ ⇒ □
 - Flee } “Goal to Source” ○ ⇒ □
 - Lead } “Goal to Source” ○ ⇒ □
- Eyes (acentric dots): ○ : □ : □ :
- Caption Veracity
 - True
 - False (NPs interchanged)

Dependent Variables:

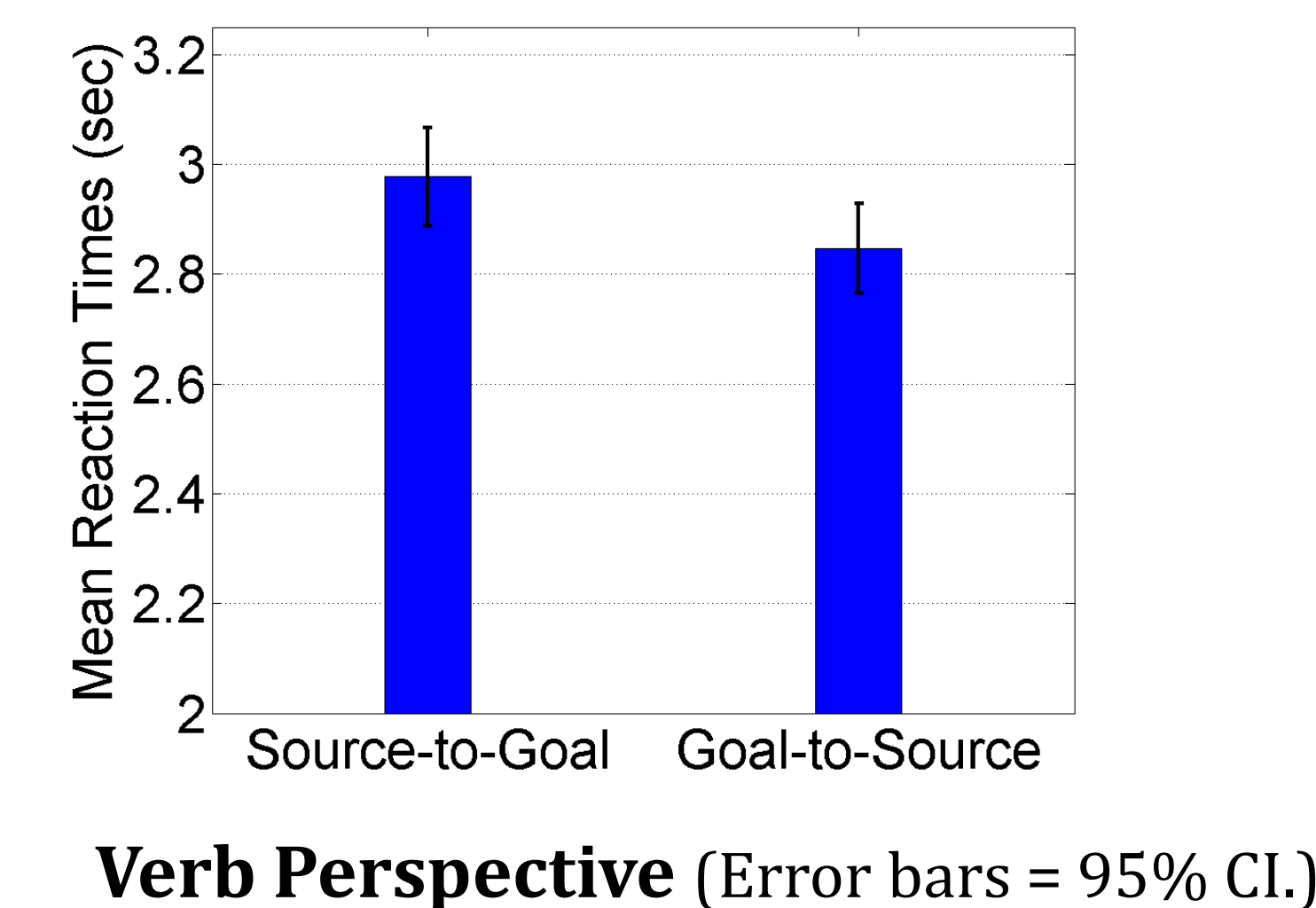
- Reaction Time, [Error Rates]

Results

- Mean accuracy: 94.91%
- RTs for correct trials only were analyzed.
- Main effect of Eyes ($p < .01$)



- Main effect of Verb Perspective ($p < .05$)



Acknowledgements

This work was supported by the National Science Foundation (BCS-0446850, BCS-0124095, & IGERT DGE-0549115). We thank Jacob Feldman, Sabrina Angelini, Tina Hou-Imerman, Nora Isacoff, Choonkyu Lee, Heather Yaden, and Nikhita Karki for their suggestions and advice.

Experiment 2- Actives and Passives

Task: Same as Exp. 1.

Subjects: 24 English-speaking college-aged adults.

Independent Variables:

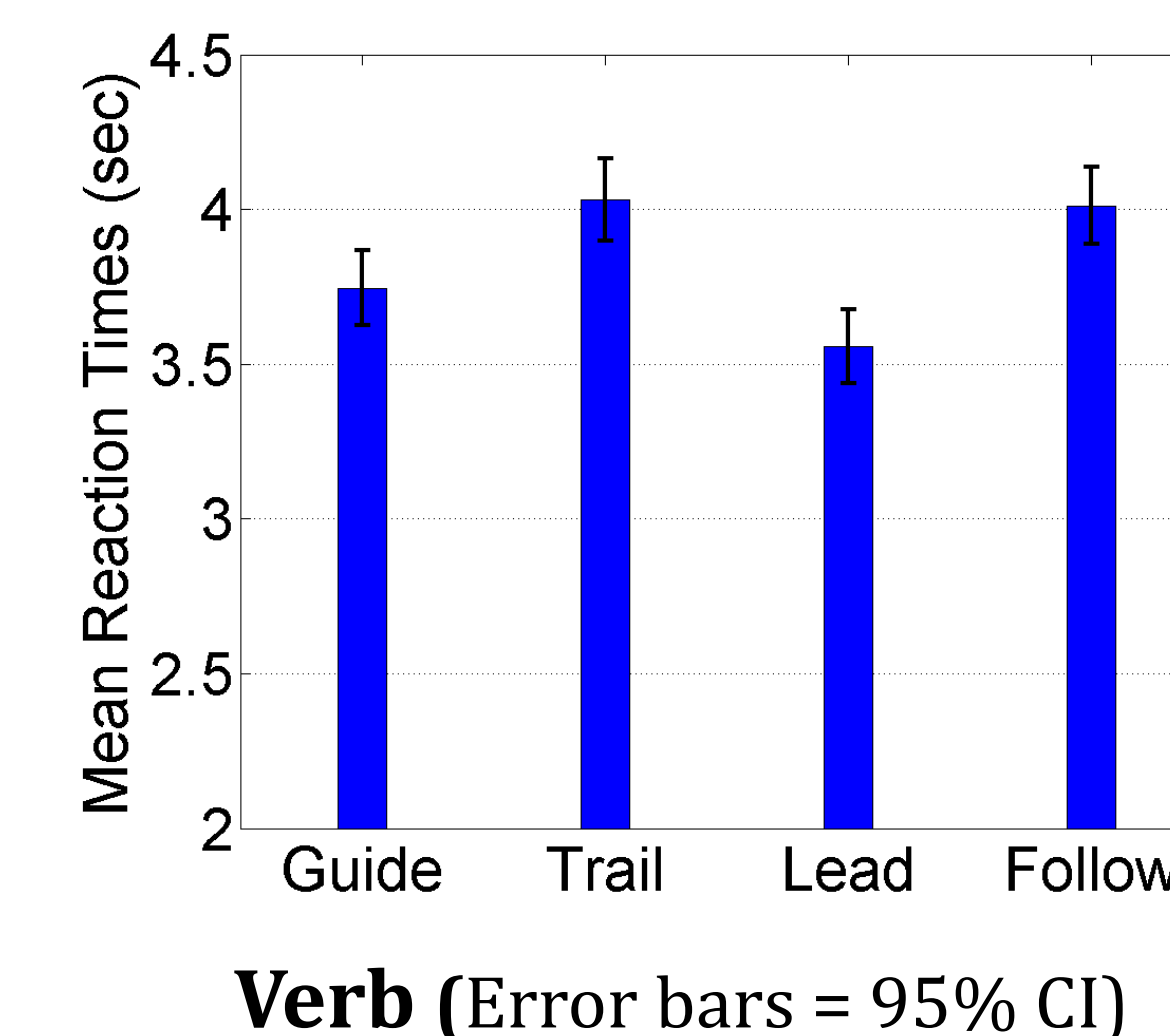
- Voice:
 - Active (e.g. *The circle is chasing the square.*)
 - Passive (e.g. *The square is chased by the circle.*)
- Verb:
 - Trail } “Source to Goal” ○ ⇒ □
 - Follow } “Goal to Source” ○ ⇒ □
 - Guide } “Goal to Source” ○ ⇒ □
 - Lead } “Goal to Source” ○ ⇒ □
- Eyes & Caption Veracity:
 - same as Exp. 1

Dependent Variables:

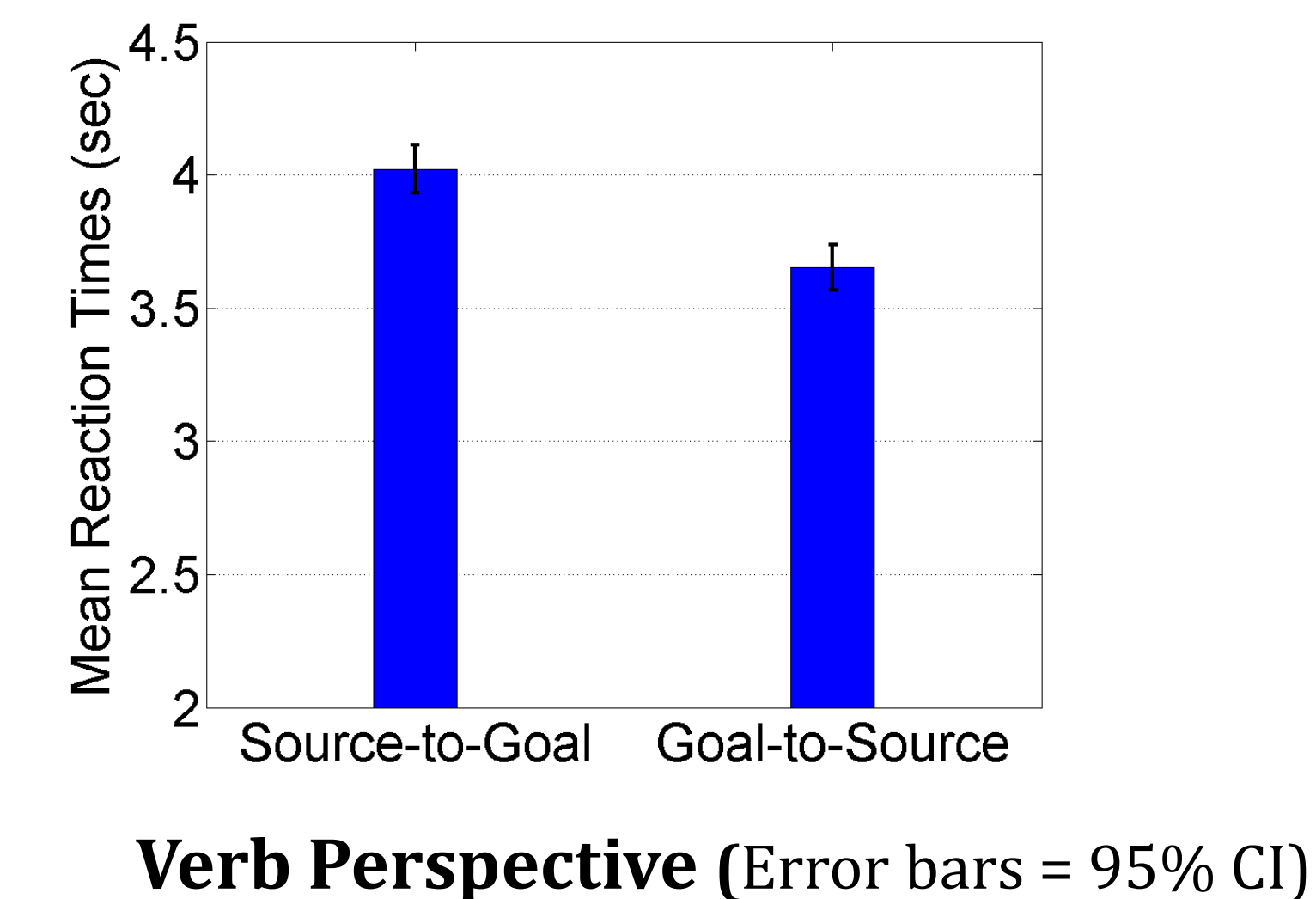
- Reaction Time, [Error Rates]

Results

- Mean accuracy: 95.75%
- RTs for correct trials only were analyzed
- Main effect of Verb ($p < .001$)

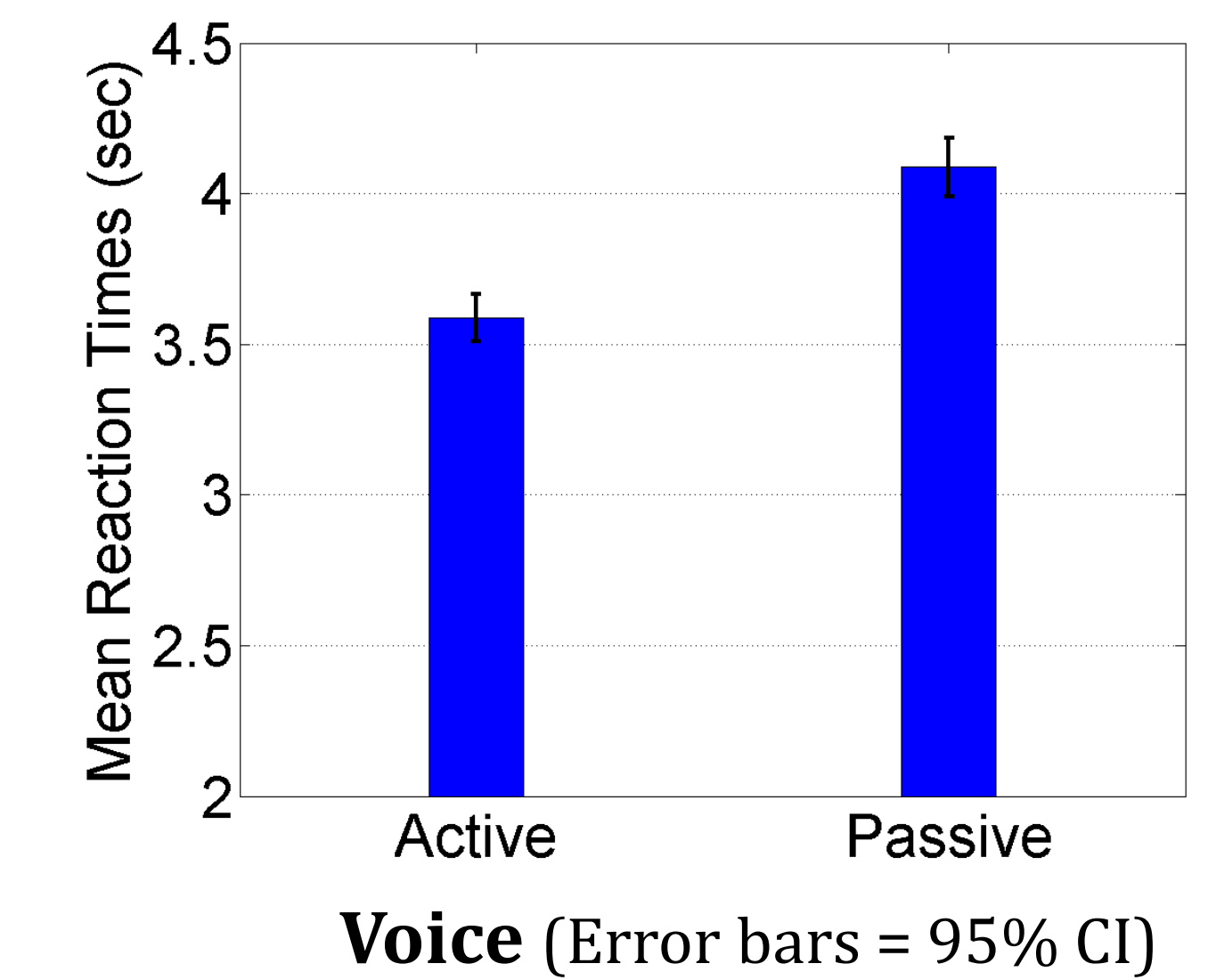


- Main effect of Verb “Perspective” ($p < .001$)

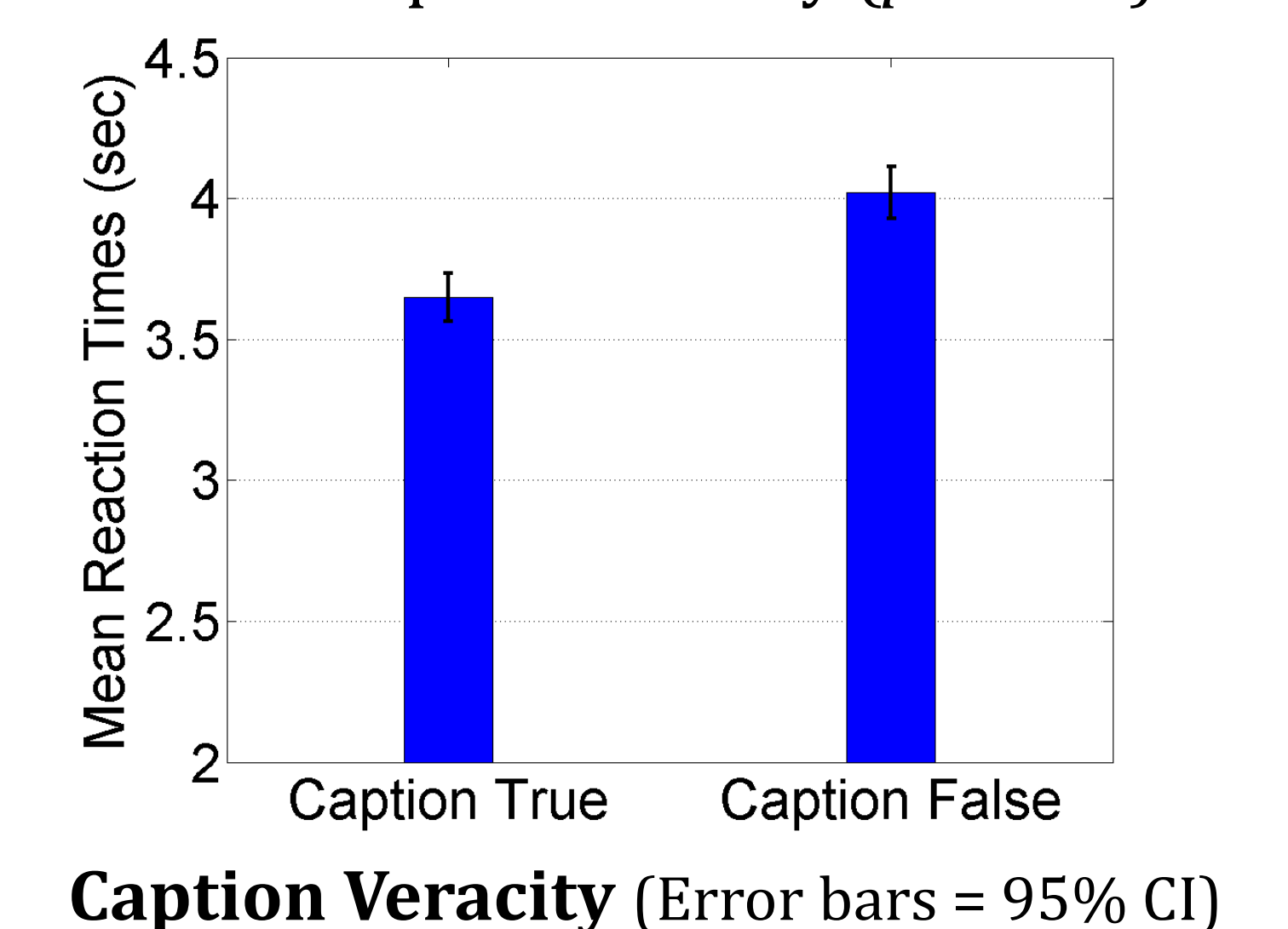


Results (contd.)

- Main effect of Voice ($p < .001$)



- Main effect of Caption Veracity ($p < .001$)



Conclusion

Expt. 1

- Linguistic task simpler
 - > heuristic (N V N) parser sufficient^{1,2}
- Visual processing is limiting factor

Expt. 2

- Linguistic task harder
 - > heuristic (N V N) parser insufficient
 - > detailed parse required
- Language processing is limiting factor

Hypothesis

- Horse-race model
 - > slower system modulates effects
- “Good enough” processing of language^{3,4}

References

1. Bever, T. G. (1970). *The cognitive basis for linguistic structures*, 279-362. New York: Wiley & Sons, Inc.
2. Townsend, D. and Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.
3. Ferreira, F., Ferraro, V., and Bailey, K. G. D. (2002). *Good-enough representations in language comprehension*. Current Directions in Psychological Science, 11:11-15.
4. Ferreira, F., and Henderson, N. D. (2007). *The ‘good enough’ approach to language comprehension*. Language and Linguistics Compass, 1:71-83.