

## Does contrastive attention guidance facilitate action recall? - An eye-tracking study.

Amit Singh, Katharina J. Rohlfing  
Paderborn University, Germany

Teaming up with AI systems, humans could be guided for action performance. However, little is known about what kind of guidance is facilitating human performance. Studies in social science suggest that when humans use explanations, they are often contrastive (Miller, 2021) Contrastive explanations promote easier communication, fine grained understanding, and reduce cognitive load (Lipton, 1990). Although there are many studies investigating the linguistic representational aspects of contrastive explanation there are very few investigating it in a domain of motion event understanding. The following study fills this gap by bringing together contrastive explanation and studies on motion event. A motion event consists of a source, a figure, and a goal (Talmy, 2000), which provides a possibility to investigate the effect of different forms of verbal guidance on the action understanding. We reasoned that if contrastive verbal explanation is easier to understand generally, it might also lead to better understanding of action sequences. Previous research in psycholinguistics suggests a goal prominence in motion event, whereby participants retain a robust memory of goal than the source or the path (Papafragou, 2010). Informed by the research background, in this study we investigated whether, a) contrastive verbal guidance facilitates the motion event recall as in line with the contrastive explanation literature, and b) to what extent the goal prominence is affected by such guidance.

**Stimulus and Method:** Participants were presented video stimulus ( $N = 20$ ) in which a ball was moved in relation to three landmark objects (Fig.1). Crucially, each action sequence was performed with a verbal guidance illustrating the path of the motion. The guidance was designed along two variables accounting for the contrastiveness in terms of verbal utterance (assertion or negation) and performed movements (i.e., motion path). For the verbal utterances, we used an Assertive(A) in contrast to Negative(N) verbal instructions. We chose negation for the reasons that a) negation has been shown to activate the alternate representation (Kaup, 2006), which can be contrasted against its positive counterpart, and b) negation guides, what not to do in addition to what has to be done and might reduce the goal bias which is prevalent otherwise (e.g., not down/up). For the performed movements, the ball followed either a contrastive (C) e.g., (up-down or down-up) vs non-contrastive (NC) (up-up or down-down) configurations. The recall was immediately assessed after each trial where participants turned 180 deg. from the eye-tracker and performed motion sequence without time constraint. To measure the goal bias, we calculated the fixation proportion on the final object path (alternate path) in the late window timeline (fig. 1), when the ball moved in the opposite direction (actual path).

**Results:** The results ( $N = 29$ ) show the effect of contrast and verbal guidance on event recall and goal bias (fixation). To visualise activation of goal; the mean fixation on alternate Area of Interest (AOI) was calculated (fig.1 c(ii), nc(i)). The recall was coded 0 for incorrect or 1 for correct responses in pre- and post-windows for each correct path performed. A mixed effect logistic model was fitted with voice and path condition as fixed effect and random intercept for subjects. Figure 2(b) shows the recall for C and NC conditions yielding a main effect of path, such that recall for  $C > NC$ . Taking NV (No Voice) as the baseline, there was an overall better recall for AA voice condition in C and NC path conditions which was not surprising. Participants could better recall the paths when it was accompanied by assertive voice guidance. Crucially there was a significant interaction between path and voice such that for NA voice condition the recall was higher in C than the NC path condition. As predicted, we find that contrastive verbal guidance is facilitative when combined with a contrastive path and otherwise for NC the verbal contrast was detrimental. Moreover, the recall for NC path was higher in NN voice condition which might be due to the repetition of negation which was absent in C paths. For goal activation, we performed GCA analysis (Mirman et al., 2008) with 3<sup>rd</sup> order polynomial on logit transformed fixation data with voice as fixed effect. Fig. 2(a) shows the fixation pattern to the goal object. There was a main effect of voice, such that fixation for  $NV > AA > NN > NA > AN$ , suggesting that negation significantly reduced the fixation on alternate goal path during post window motion which was otherwise maximum when there was no guidance in NV baseline.

**Keywords:** Contrastive Explanations, Action Recall, Motion Event, Negation, Attention.

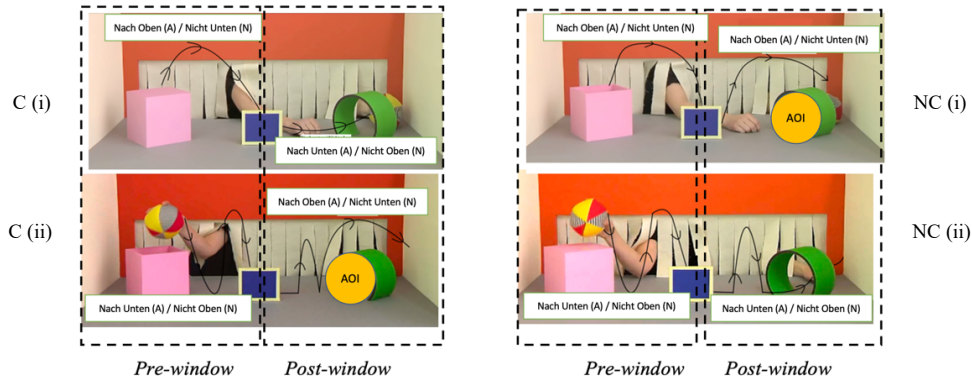


Figure 1. Path conditions, C, NC, C, NC (from top left to bottom right). Each movement window was preceded by either an Assertive(A) e.g., “nach oben” [towards up] or Negative(N) e.g., “nicht oben” [not up] voice predicate to create different degrees of contrasts including No Voice (NV) as baseline (AA, NA, AN, NN, NV).

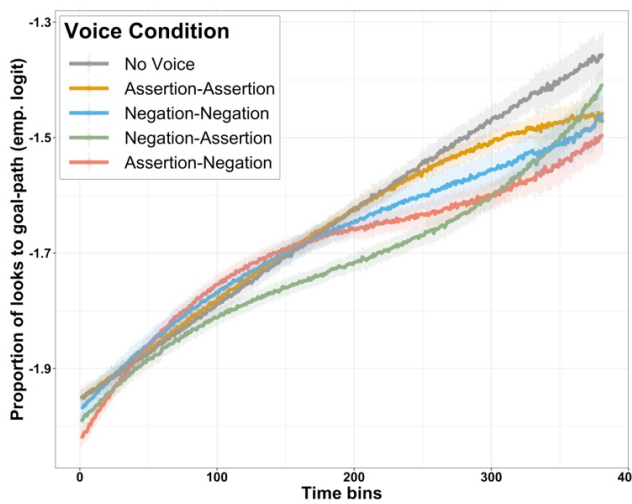


Figure 2 (a). Mean proportion of fixation on goal alternate path. Results are shown for analysis time-window of post-window region highlighted in fig 1.

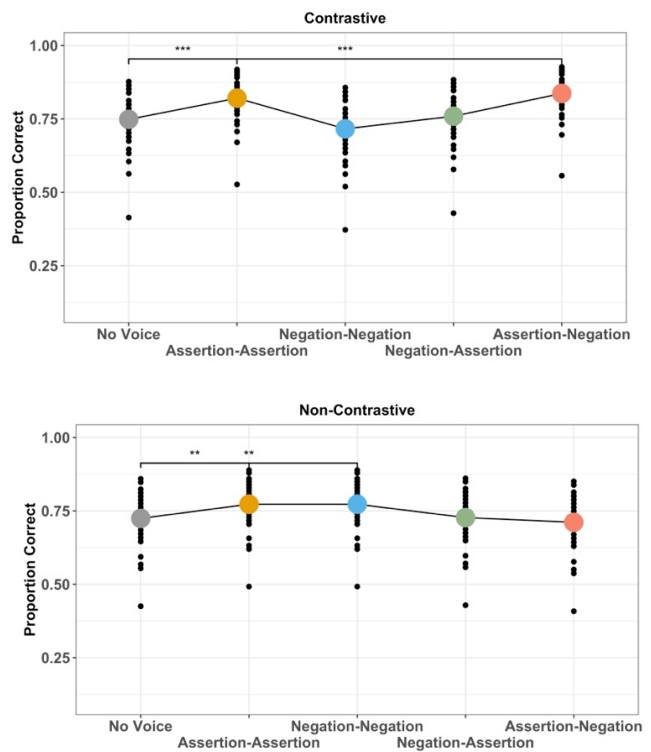


Figure 2 (b). Proportion of correct recall for C and NC path conditions. Pairwise comparison with No Voice as baseline. (Codes: \*\*\* =  $p \leq 0.001$ , \*\* =  $p \leq 0.01$ ).

## References

- Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, 38(7), 1033–1050.
- Lipton, P. (1990). Contrastive explanation. *Royal Institute of Philosophy Supplement*, 27, 247-266.
- Miller, T. (2021). Contrastive explanation: A structural-model approach. *The Knowledge Engineering Review*, 36, (E14).
- Mirman, D., Dixon, J.A., & Magnuson, J.S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475-494.
- Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science*, 34, 1064–1092.
- Talmy, L. (2000). *Toward a cognitive semantics*, vol. 1. MIT Press.

Supplementary materials: <https://osf.io/ZFCBK/>