Event gist makes rapid contact with ongoing linguistic processing, even in the absence of visual preview

Junyi Chen & John Trueswell
University of Pennsylvania

Introduction: Event perception research shows that event-category and -role information can be extracted rapidly (< 300 ms) from centrally fixated two-participant action images (Hafri et al., 2013, 2018), which is based in part on low-level visual postural properties. Here, we examine whether rapidly extracted event representations make immediate contact with ongoing linguistic processes. Participants heard sentences (e.g., "The red person is kicking the blue person") while searching for the matching image, with eye movements recorded. To estimate when visually recognized event information becomes available to linguistic processing, we manipulated visual preview time before hearing the critical verb and altered human physical postures to assess event role recognition. Participants rapidly extracted the event gist and integrated it with linguistic input almost immediately. In particular, even when there was no visual preview (i.e., when images appeared at verb onset), looks to the Target exceeded chance in just 400–600 ms.

Methods: On each trial, observers fixated centrally before viewing two images, each depicting two-participant actions involving a red- and blue-shirted person. Participants clicked the image matching the spoken sentence. Depending on the study, images appeared either at verb onset (Exp 1A, Exp 2A) or just before sentence onset (Exp 1B, Exp 2B). The postural properties of Foils (Exp 1 & 2) and Targets (Exp 1) were manipulated. Eye movements were recorded with a Tobii X300, coded for screen-side (Target vs. Foil), and binned into 100-ms intervals. Data were analyzed using (sum-coded) linear models of Elogit-transformed looks, with cluster-based permutation tests identifying reliable windows of main effects and interactions. Each experiment had 32 trials over 2 blocks.

Exp 1: Eighty-nine native English speakers participated. Stimuli consisted of paired images depicting two-person interactions (red/blue-shirted participants). Eight experimental lists counterbalanced Target side (left/right) and Agent color (red/blue), with Foil Agent color always matching the Target's. The red/blue person's position remained consistent per trial (Fig1). Each trial included an audio cue (e.g., "The red person is verb-ing the blue person"). Experiment 1 showed that linguistic processing can rapidly integrate visual event information, with Target fixations exceeding chance 400-500 ms after verb onset ($p<.01$; Exp 1A) and 200 ms earlier when images were previewed ($p<.01$; Exp 1B) (Fig2). Contrary to expectations, Patient Posture had minimal effects on looking behavior (Table1).

Exp 2: Another ninety-four native-English speakers participated. Materials were the same as in Exp 1, except Target images had prototypical postures, while Foils showed role-reversed participants. Foils varied in a 2×2 design by Action Type (Same/Different from Target) and Patient Posture (Patient-like/Agent-like) (Fig3). Event gist integration occurred slightly later in Exp 2A (500-600 ms) compared to Exp 1A. Previewing the images (Exp 2B) enabled Target anticipation prior to verb onset (Fig4) because the NP "The red person…" disambiguates reference. Contrary to expectations, Agent-like Posture in the Foil image diverted attention relatively late (800–1100 ms), suggesting non-incremental processing for refining representations or response verification.

Summary: Two eyetracking experiments on linguistically guided visual search provide evidence for rapid event gist extraction and its integration with ongoing linguistic processing, with looks to the Target action exceeding chance 400–600 ms after verb onset. This aligns with prior findings (Hafri et al., 2018) on event gist extraction and suggests that rapidly recognized event information is abstract enough to interface with language. While postural information influenced search in expected ways, these effects were weak and appeared later, possibly indicating that further visual interrogation helps refine event structure and verify Target choices.

Figure 1: Example item, four conditions (Exp 1)

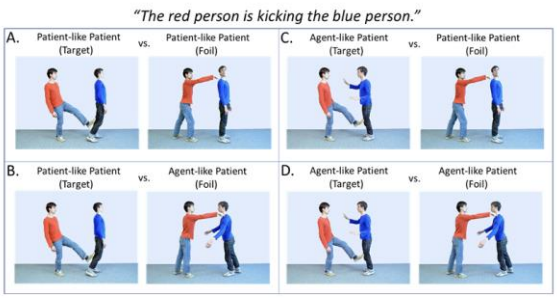

Figure 2: Proportion of looks to Target Side (Exp 1)

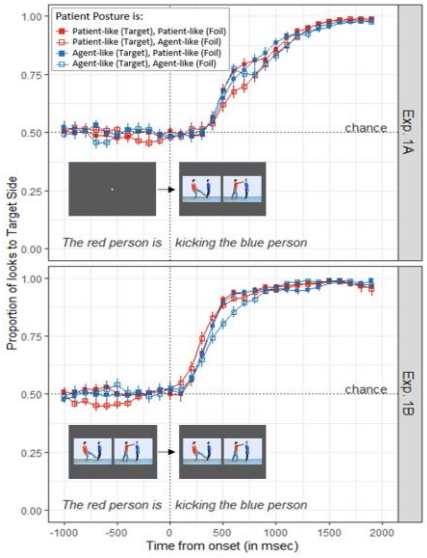

Table 1: Exp 1 Cluster-Based Permutation Tests

A.Effects (Exp1A)	Cluster in ms	Sum *t*	*p* =
Target Patient Posture	700 to 1100	-11.40	0.087
Block (1 vs. 2)	500 to 1700	-41.60	0.005
B.Effects (Exp1B)	Time of Cluster	Sum *t*	*p* =
Target Patient Posture	500 to 800	-11.34	0.058
Block (1 vs. 2)	300 to 800	-17.00	0.001
	1100 to 1500	-15.73	0.001
C.Effects (Combined Exp1A and Exp1B)	Time of Cluster	Sum *t*	*p* =
Experiment (1A vs. 1B)	0 to 1400	91.42	0.001
Target Patient Posture	500 to 1100	-17.34	0.015
Foil Patient Posture x Experiment	400 to 600	-6.59	0.001
Target Patient Post. Foil Patient Post. x Exp.	-700 to -500	6.11	0.001
	400 to 700	-9.82	0.001
Block (1 vs. 2)	400 to 1800	-56.00	0.001
Block x Exp. x Target Patient Posture	1300 to 1500	5.09	0.001
Block x Exp. x Target Patient Post. x Foil Patient Post.	-700 to -200	12.36	0.047
(No other reliable effects or interactions)			

Formula (Exp 1A&1B): Elog ~ 1 + TargPatPost*FoilPatPost*Block + (1 | Subject) + (1 | TargetVerb)
Formula (Combined): Elog ~ 1 + Exp*TargPatPost*FoilPatPost*Block + (1 | Subject) + (1 | TargetVerb)
Applied to 100 bins from -1000 to 2000. Significant clusters identified with R package jlmerclusterperm

Figure 3: Example item, four conditions (Exp 2)

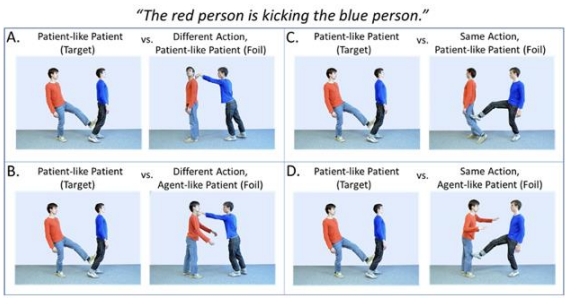

Figure 4: Proportion of looks to Target Side (Exp 2)

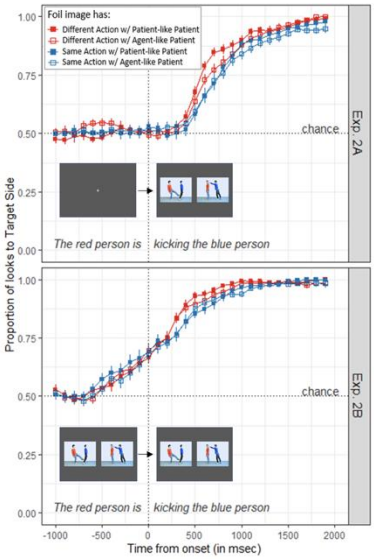

Table 2: Exp 2 Cluster-Based Permutation Tests

A.Effects (Exp2A)	Cluster in ms	Sum *t*	*p* =
Foil Action Type	500 to 900	15.45	0.052
Block (1 vs. 2)	400 to 1800	-45.37	0.001
B.Effects (Exp2B)	Time of Cluster	Sum *t*	*p* =
Foil Action Type	300 to 1200	26.58	0.001
Block (1 vs. 2)	-900 to -700	-5.13	0.001
	-100 to 1200	-60.86	0.001
C.Effects (Combined Exp2A and Exp2B)	Time of Cluster	Sum *t*	*p* =
Experiment (2A vs. 2B)	-500 to 1500	140.61	0.001
Foil Action Type	300 to 1300	34.27	0.001
Foil Patient Posture	800 to 1100	-7.19	0.001
Foil Action Type x Foil Patient Type	500 to 700	-5.91	0.001
Block (1 vs. 2)	-200 to 1400	-70.8	0.001
Foil Action Type x Experiment	600 to 700	-3.75	0.001
	1700 to 1900	-6.61	0.001
(No other reliable effects or interactions)			

Formula (Exp 2A&2B): Elog ~ 1 + FoilAction*FoilPosture*Block + (1 | Subject) + (1|TargetVerb)
Formula (Combined): Elog ~ 1 + Exp*FoilAction*FoilPatient*Block + (1| Subject) + (1|TargetVerb)
Applied to 100 bins from -1000 to 2000. Significant clusters identified with jlmerclusterperm

References

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: 1recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General, 142*(3), 880.

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition, 175*, 36-52.