Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Extraction of Event Roles From Visual Scenes is Rapid, Automatic, and Interacts with Higher-Level Visual Processing

Permalink

https://escholarship.org/uc/item/32p747vc

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors

Hafri, Alon
Trueswell, John C.
Strickland, Brent

Publication Date

2016

Peer reviewed

Extraction of Event Roles From Visual Scenes is Rapid, Automatic, and Interacts with Higher-Level Visual Processing

Alon Hafri (ahafri@sas.upenn.edu)
Department of Psychology, University of Pennsylvania, 3720 Walnut Street
Philadelphia, PA 19104 USA

John C. Trueswell (trueswel@psych.upenn.edu)
Department of Psychology, University of Pennsylvania, 3720 Walnut Street
Philadelphia, PA 19104 USA

Brent Strickland (stricklandbrent@gmail.com)
Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL Research University
Institut Jean Nicod, (ENS, EHESS, CNRS)
75005 Paris, France

Abstract

A crucial component of event recognition is understanding the roles that people and objects take: did the boy hit the girl, or did the girl hit the boy? We often make these categorizations from visual input, but even when our attention is otherwise occupied, do we automatically analyze the world in terms of event structure? In two experiments, participants made speeded gender judgments for a continuous sequence of male-female interaction scenes. Even though gender was orthogonal to event roles (whether the Agent was male or Female, or vice-versa), a switching cost was observed when the target character's role reversed from trial to trial, regardless of whether the actors, events, or side of the target character differed. Crucially, this effect held even when nothing in the task required attention to the relationship between actors. Our results suggest that extraction of event structure in visual scenes is a rapid and automatic process.

Keywords: event roles; thematic roles; event perception; visual perception; switching costs

Introduction

A fundamental way we interpret the world is not just in terms of objects, but also events. Indeed, categorizing the causal relationships between people and objects is important for guiding our social behavior: Was it the boy that hit the girl, or the girl that hit the boy? The semantic relationships that exist between participants in events are called their roles (Fillmore, 1968; Gruber, 1965). For example, in *John broke the door,* John is the Agent (actor) and the door is the Patient (undergoer). The semantic properties of these roles and others (e.g., source and goal) appear to be consistent across a wide range of events, with some theorists arguing that the Agent and Patient roles subsume all others (Dowty, 1991). Relatedly, Strickland (2016) has argued that the Agent/Patient distinction is one form of "core knowledge," that is reflected across languages, pre-verbal infant cognition, and the adult visual system. While there is ongoing debate about the precise nature of the Agent/Patient dichotomy, it is clear that these representations are a crucial component of recognizing what is happening in the world.

Given the importance of event role information, then, it would be beneficial for the process of identifying roles to be automatic, continuously working in the background, since at any given moment we may be attending to other perceptual information, e.g., identifying objects or spatial properties of the scene. Recent evidence suggests that people are able to discriminate Agents from Patients from input lasting less than 75 ms, and that elements of body posture are important heuristics for these categorizations (Hafri, Papafragou, & Trueswell, 2013; see also Dobel, Gumnior, Bölte, & Zwitserlood, 2007; Wilson, Papafragou, Bunger, & Trueswell, 2011). However, it is not yet known whether this discrimination is only active when the task requires it or is a fully automatic process.

To investigate this issue, we employ a paradigm that has been used profitably in the past to investigate questions of automaticity: the switching cost paradigm (e.g., Pecher, Zeelenberg, Barsalou, 2003; Spence, Nicholls, & Driver, 2001). The logic of this paradigm is as follows: if a certain process (here role recognition) is automatically engaged, then when the role of an actor switches before participants judge an orthogonal actor property (here gender), it should result in a switching cost; i.e., a lag in reaction time.[1] If such a pattern is observed, it would provide strong evidence that analysis of event structure from the visual world is a rapid, automatic process that helps organize our understanding of the social and causal world, even when we are not explicitly focused on this information.

Experiment 1

In the first experiment, participants were asked to identify the

---

[1] Though previous work has referred to these reaction time effects as costs, here we cannot differentiate between switching costs vs. repetition benefits, because there is no meaningful baseline for comparison (i.e., in our stimuli, the roles always either repeated or they did not). In any case, whether the effects are a benefit or cost does not qualitatively change our conclusions.

side of the male or female actor in a sequence of scenes, and we looked for evidence of a switching cost when event roles changed from one scene to the next. To maximize our chances of finding such an effect if it exists, we included a secondary catch task on a subset of trials, in which participants were probed about what event they just observed (asking whether they just saw, e.g., *kicking* or *punching*). This secondary task made no mention of event roles, so participants were not required to extract role information to perform either task.

Methods

Participants

Twenty-four members of the University of Pennsylvania community participated and received either class credit or $10 for their participation.

Materials and apparatus

Forty color photographic images depicting 10 two-participant Agent-Patient events were used in the current experiment, taken from a previous study (Hafri et al., 2013) that investigated extraction of event categories and roles from briefly displayed and masked images (< 75 ms). These 10 showed the highest agreement for role assignment among participants when they were probed after brief displays. The events used were: *brushing, chasing, feeding, filming, kicking, looking, punching, pushing, scratching, tapping*.

Each event was depicted from a side view, and involved a male and a female actor. Six different actor pairs appeared in the 10 events, with each actor pair appearing in front of a different indoor or outdoor background. Each event was associated with only one of the actor pairs. For each event, there were four versions: the gender of the Agent was male or female, and the side of the Agent was left or right.[2] All events were normed for name agreement in the previous study. Each image was 640 × 480 pixels and subtended 19° × 15° at approximately 54 cm. Example images appear in Figure 1.

Stimuli were displayed on a 19" Dell 1908FP LCD monitor at a refresh rate of 60 Hz. Responses were collected using a PST E-Prime button box (mean latency 17.2 ms, *SD* 0.92 ms). The experiment was run in Matlab using the Psychophysics Toolbox (Brainard, 1997).

List design

Given that detecting switching costs depends on measuring the influence of one stimulus on another, we chose to implement a list design that controls for first-order carry-over effects (continuous carryover sequences; Nonyane & Theobald, 2007). These sequences are similar to randomized block and Latin square designs, with the added benefit of the following attributes: each item precedes and follows every other, including itself, exactly once; effects of the current and preceding stimuli are orthogonal to position effects in the list; and each item occurs exactly once every block, where a block represents a permutation of all items. Thus here, we use sequences of $n = 40$, resulting in 1601 ($n^2 + 1$) trials split among 40 blocks. Unique lists were generated for every participant.

Among these standard image trials, we randomly dispersed catch (Event Test) trials, in which participants were given a 2AFC test about what action just appeared in the previous trial (e.g., *tickling* vs. *scratching*). One label was correct, and the other was a foil randomly selected from the set of nine other possible actions. The purpose of these trials was to focus participants' attention on the event without explicitly testing them on event roles. There were 58 of these trials, with 1 to 3 per 40-trial block.

Figure 1: Example images (clockwise from top left) for *punching*, *brushing*, *feeding*, and *kicking*.

Procedure

Participants were told that they would view photographs of people performing actions in continuous sequences. They were instructed to press the button corresponding to the side of the screen that the male character was on as quickly and accurately as possible. They were also told that after a small set of the trials, they would be tested on what action just appeared in the previous trial. Task (male or female search) was between-subject.

Twelve practice trials preceded the main experiment (plus two catch trials), all involving the actor pairs performing joint or symmetrical actions (e.g., *writing, shaking hands, crying*). Image trials consisted of the following sequence: A "Ready?" screen for 350 ms, a central fixation crosshair for 250 ms, a blank screen for 150 ms, and the test image, at which point the participant searched for and pressed the side of the male (or female) actor as quickly as possible. Catch trials involved a similar sequence, but with text "What action did you just see?" and two event probes below (e.g., *tickling* or *scratching*). Image trials timed out if no response was given within 2000 ms, and catch trials within 3500 ms. Average

---

[2] We found no differences between male-Agent vs. female-Agent images in terms of the effects reported below.

duration of the experiment was 41 min and included a break every 40 image trials.

Results

Coding

Since we were investigating effects of one stimulus on the next, exclusion criteria were the following (decided in advance of analysis): response errors and those following error trials (8.8% in total), RTs faster than 200 ms (44 trials in total), timeouts (17 trials in total), trials after breaks (40 per participant), and trials after catch trials (58 per participant). An additional 63 trials were excluded due to errors in list creation. Finally, for the remaining data, trials were excluded when the RT was 2.5 standard deviations above or below each participant's mean (3.2% of trials in total), following accepted data trimming procedures. With the above criteria, a mean of 17% (*SD* 4.0%) of trials in total were excluded per subject. Average RT for the included data was 383 ms (*SD* 34 ms).

Analysis

Accuracy on catch trials was significantly above chance across subjects (mean = .85, $SD = .10$, $t(23) = 40.0$, $p < .001$, $d = 3.37$).

Individual trial reaction times from the primary task (i.e., judging gender side) were analyzed using linear mixed effects modeling with the maximal subject and item random effects structure that converged (Barr, Levy, Scheepers, & Tily, 2013). Models with and without the following factors were compared: repeated Actor Pair (repActors); repeated Side (repSide), i.e., whether the male and female actors were on the same side as in the previous trial; and repeated Role (repRole). This last factor reflects the effect of interest: how a change in the role of the person being asked to respond about affects RT (e.g., if the male remains the Agent, or switches to being the Patient). Significance of factors was tested by comparing likelihood-ratio values for models that included factors to models without them.[3]

Findings

First, and most importantly, an event role switching cost was observed. In particular, as shown in Table 1, participants were on average 6 ms slower when the role of the target character changed from one trial to the next.

This effect, though quite small, was significant: The best-fitting mixed effects model included main effects and interactions of repActors and repSide, and a main effect of repRole (the role switching cost), over a model that did not include repRole, $\chi^2(1) = 29.3$, $p < .001$. Models with additional interaction terms were not a significantly better fit, either for repActors × repRole ($\chi^2(1) = 2.54$, $p = .11$), or repSide × repRole ($\chi^2(1) = 0.57$, $p = .45$).[4]

Besides the effect of primary interest (role switching cost), post-hoc analyses re that participants were faster when the actor pair repeated, and on trials where the actor pair did not repeat, participants were *slower* when the target side repeated. Though speculative, these effects may be accounted for by two mechanisms: visual priming due to similarity of actor pair appearance (faster RTs for repeated actors), and an incorrect expectation that the response should always switch if there is significant visual change (slower RTs for different side and actor pair). See Table 1 for a summary of all effects.

Table 1: Mean RTs by condition for Experiment 1, across Subjects. Standard errors in parentheses.

	Reaction time (ms)		
Condition	Repeated	Different	Switching cost
Role	380 (6.86)	386 (6.86)	6 (0.97)**
Actors	371 (6.18)	385 (7.19)	14 (1.78)**
Side	390 (7.73)	377 (6.60)	-13 (3.03)*
Side, Repeated Actors	371 (6.22)	371 (6.52)	0 (3.09)
Side, Different Actors	393 (7.99)	378 (6.72)	-15 (3.17)**

** = Significant effect ($p < .05$) in F_1 and F_2 ANOVAs, and multilevel modeling.

* = Significant effect ($p < .05$) in either F_1 or F_2 ANOVAs, and multilevel modeling. See text for detailed statistics.

Does Agent saliency drive the effect? The switching cost could conceivably be driven by the cost of switching from Agent to Patient or vice-versa. Indeed, an Agent primacy or saliency effect has been observed both in the linguistics and vision literature (Cohn & Paczynski, 2013; Dobel et al., 2007; Dryer, 2013; Wilson et al., 2011). If so, we should observe two additional effects in our data: (1) faster RTs on Agent trials as compared to Patient trials; and (2) an asymmetry between an Agent→Patient switch and Patient→Agent switch. However, neither of these was borne out in model comparisons (all p's > .38), and parameter estimates, though not significant, indicated *greater* RTs for Agent judgments. These analyses suggest that Agent saliency cannot account for the role switching cost effect.

Summary of results. As predicted, there was a reliable role switching cost, i.e. slower RTs when switching from a judgment on the Agent on one trial to the Patient on the next, or vice-versa. Importantly, the switching cost did not interact with other factors, such as side or actor pair, and did not appear to be driven by Agent saliency.

Magnitude of the role switching cost. Although the magnitude of this effect was small (about 6 ms), it is

---

[3] RepActors (repeated actor pair) always entailed a repeated event since each event was depicted by only one actor pair, so all analyses reported include only repActors. Additionally, reaction times were transformed into inverse RTs by using -1000/RT as the response variable. Mean raw RTs are displayed in Tables 1 and 2 to illustrate basic effects. All models included nuisance regressors for trial number and preceding trial inverse RT to account for general temporal dependencies.

[4] Though the nature of the design results in unbalanced numbers of observations in each condition, very similar findings emerged in separate Subject (F_1) and Item (F_2) ANOVAs on mean inverse RTs for all analyses. This is true for both Exp 1 and Exp 2.

comparable to previously observed switching costs, relative to mean RTs for task (e.g., Pecher et al., 2003, obtained a cost of 29 ms relative to mean RTs of 1139 ms, compared with our 6 ms vs. 383 ms mean RTs). And while it may be surprising that such a small effect would be statistically significant, it is important to keep in mind that unlike a typical cognitive experiment, each observer provided on average 1329 data points (more than many cognitive studies), resulting in very stable performance estimates per subject and per item (e.g., note the low Standard Errors across Subjects in Table 1). Furthermore, as an indication of its robustness, 21/24 participants and 10/10 items showed a numerical difference in line with the role switching cost.

Experiment 2

Our results thus far provide compelling evidence that when observing scenes, people extract event structure, even when it is not a specific component of the task. However, it may still be the case that our secondary catch (Event Test) task, despite not being explicitly about role identification, inadvertently focused participants' attention on the event structure itself. That is, in order to perform the task of event category extraction, they may have defaulted to a strategy of analyzing event roles. Experiment 2 addresses this issue. We conducted the same experiment on a new set of participants, but removed the catch trials, and made absolutely no mention of events, actions, or roles in our instructions. If this effect is really a fact about the visual perception of scenes, then we expect to observe it even under these conditions.

Methods

Participants

An additional 24 members from the University of Pennsylvania community participated and received class credit for their participation.

Materials and procedure

All materials, apparatus, and procedure were identical to Experiment 1, except for the following changes: First, no catch (Event Test) trials were included. Second, instructions were modified to omit mention of the catch trial task, and importantly, the beginning of the instructions were rewritten to omit mention of actions ("You will be shown photographs of people in different scenes…"). Task (male or female search) was again between-subject. Average duration of the experiment was 38 min.

Results

Coding and Analysis

Data coding procedures were the same as in Experiment 1. We excluded trials with response errors and those following error trials (8.0% in total), RTs faster than 200 ms (62 trials in total), and timeouts (7 trials in total). Trials after breaks (40 per participant) and trials from the list creation error (216 in total) were also excluded. Outliers 2.5 standard deviations above or below each participant's mean were removed (mean 3.1%). With these criteria, a mean of 13% (*SD* 4.9%) of trials per subject were excluded. Average RT for the included data was 387 ms (*SD* 48 ms). Individual trial reaction times from the primary task (i.e., judging gender side) were analyzed using linear mixed effects modeling, as in Exp 1.

Findings

As in Experiment 1, a role switching cost was observed. In Table 2, we see that participants were on average 3 ms slower when the role of the target character changed from one trial to the next. Furthermore, the role switching cost interacted with repeated actor pairs, such that the role switching cost when the actor pair repeated was marginally greater than when the actor pair did not, $t_1(23) = 1.77$, $p = .09$, $d = .36$.

These effects, although small, were significant: The best-fitting mixed effects model included main effects and interactions of repActors and repSide, and a main effect of repRole and interaction of repRole × repActors. The fit of the model was significantly better than the same model without the additional interaction of repRole × repActors, $\chi^2(1) = 4.89$, $p = .03$; and significantly better than a model that did not include repRole at all, $\chi^2(2) = 15.5$, $p < .001$. Additionally, a model that also included an interaction of repRole and repSide was not a significantly better fit, $\chi^2(1) = .004$, $p = .95$.

As in Exp 1, post-hoc analyses indicated that participants were faster when the actor pair repeated, and on trials where the actor pair did not repeat, participants were *slower* when the Side repeated. See Table 2 for details.

Table 2: Mean RTs by condition for Experiment 2, across Subjects. Standard errors in parentheses.

	Reaction time (ms)		
Condition	Repeated	Different	Switching cost
Role	385 (9.62)	388 (9.79)	3 (0.77)*
Actors	371 (8.11)	390 (10.0)	19 (2.68)**
Side	394 (9.43)	380 (10.3)	-14 (3.65)*
Side, Repeated Actors	368 (6.75)	374 (9.74)	6 (5.62)
Side, Different Actors	398 (9.93)	382 (10.4)	-16 (3.51)**
Role, Repeated Actors	368 (8.12)	374 (8.29)	6 (2.73)*
Role, Different Actors	388 (9.91)	391 (10.2)	3 (0.89)*

** = Significant effect ($p < .05$) in F_1 and F_2 ANOVAs, and multilevel modeling.
* = Significant effect ($p < .05$) in either F_1 or F_2 ANOVAs, and multilevel modeling. See text for detailed statistics.

Effect of Agent saliency. In contrast to Experiment 1, there was evidence for an asymmetry in switching cost for Agent→Patient trials vs. Patient→Agent trials in model comparisons, $\chi^2(1) = 3.96$, $p = .05$. However, this was in the *opposite* direction of that predicted by Agent saliency: Switching from judging the Agent to the Patient was slightly *faster* than vice-versa. We also found some evidence for a difference between Agent and Patient judgments: Specifically, reaction times were faster for Agent judgments than for Patient judgments, but only when the actor pair repeated (368 ms vs. 372 ms; $\chi^2(2) = 6.66$, $p = .04$). Crucially, whether people were making an Agent or Patient judgment did not interact with our effect of interest, the role switching cost, $\chi^2(1) = 2.04$, $p = .15$.

Summary of results. As predicted, we again observed a reliable role switching cost, i.e. slower RTs when the role of the target character switched, even when participants were not probed about the event. The effect again did not appear to be driven by Agent saliency.

Comparison of Experiments 1 and 2. From Tables 1 and 2, it appears that the magnitude of the role switching cost in Exp 1 is greater than in Exp 2. Formal comparison of the role switching cost across experiments revealed that the difference is significant, $t_1(45.3) = 2.14$, $p = .038$, $t_2(9) = 3.35$, $p = .009$. Indeed, more participants and items showed the numerical difference in Exp 1 than in Exp 2 (21/24 vs. 17/24 participants, and 10/10 vs. 7/10 items). Nevertheless, items drove role switching cost consistently across experiment: the effect for individual stimuli was correlated across experiment, $r = 0.37$, $t(38) = 2.43$, $p = .02$. All of these findings further attest to the stability of the measures of central tendency (i.e., subject and item means) – likely due to the large number of observations per cell.

Discussion

The two experiments presented here demonstrate that the structure of an event, i.e. who did what to whom, is continuously involved in visual processing, even when attention is directed toward other features (here, gender) that do not require extracting any information about event roles. In particular, we observed a small but reliable switching cost associated with verifying the side of the male (or female) actor in photographic images when the event roles switch from one image to the next in sequence. This effect held even when nothing in the instructions or task made reference to events or human interactions, or required making judgments about the event (Exp 2). Indeed, in debriefing, no subject explicitly guessed our hypothesis that there could be a switching cost if the person being judged had a different role than in the previous trial.

The switching cost effect appears to be robust across a range of factors: it held (1) across events, (2) across actors, and (3) across the side of the male and female characters. Furthermore, we also showed that this effect is not driven by the saliency or priority of Agents: there was no interaction between the role switching cost effect and whether participants were making Agent or Patient judgments, and there was no difference when switching from Agent to Patient judgment, compared to vice-versa (if anything, switching Agent to Patient was slightly *faster* in Exp 2).

We do not believe that the effect is due to the degree of mismatch between Agent- and Patient-like body poses from trial to trial, which could in principle drive the effect. Indeed, the fact that the role switching cost is largely invariant to the particular side (and orientation) of the target actor argues against a location-specific body-pose-matching explanation. Nevertheless, we cannot rule out location-general pose-matching. Certainly body pose is a strong and reliable heuristic to event role (Hafri et al., 2013), so higher-level body pose recognition may be the first route for identifying event roles in initial processing. We are not opposed to such an explanation, and indeed there seems to be a correspondence between Dowty's (1991) proto-role entailments and visually salient aspects of body posture, including orientation of the head and body (Dowty's "volitional involvement"), and outstretched extremities (Dowty's "ability to causally effect change").

Given that this effect is not explainable by a simple location-specific pose-matching hypothesis, what is its origin? One possibility is the following: In the gender identification task, participants search for the target character to plan their response, attending to the male (or female) character as necessary (e.g., in *girl-kicks-boy*, attention is placed on the boy). If the role of the target character changes (e.g., the next image is *boy-taps-girl*), then there is a mismatch between the previous role of the male actor (Patient) and the current role (Agent). This explanation assumes that when people identify a particular person-based attribute (e.g., gender), others "come along for the ride," and that mismatches in these attributes from trial to trial manifest in increased reaction time at the decision making stage. Our assertion here is that event role identification is automatic, and is thus consistent with the claim that Agent/Patient event role information is part of human "core knowledge" (Strickland, 2016).

Although it was not a primary focus of these studies, might our results contribute to the ongoing debate about the degree to which roles are event-general or event-specific? In particular, some have argued for an innate and limited set of semantic relations (e.g., Pinker, 1989), while others have argued that event-general properties of roles develop from observation of commonalities among event-specific roles (e.g., Tomasello, 2000). In Exp 2, we did find evidence that the role switching cost was greater *within* event (repeated actors), but we note that in both experiments, the effect still held *between* events. Thus whatever its ontogenetic origins, Agent and Patient roles are at least partly event-general.

One interesting question that arises is the extent to which this is a general property of event scene analysis, or is specific to human interactions. That is, in event scenes that involve interactions with or among inanimate objects (e.g., *A woman opens a door* or *A ball hits a rock*), are roles assigned using similar processes? If we view language as a reflection of the semantic structures available to the human mind, it is of course possible for objects to fill the same argument slots as humans do in an utterance. Yet even then, there are expectations of what can fill which roles: animacy of sentential constituents constrains on-line sentence interpretation (contrast *The defendant examined by the lawyer…* with *The evidence examined…*; Saffran, Schwartz, & Linebarger, 1998; Trueswell, Tanenhaus, & Garnsey, 1994). It may be that on the path from vision to complex, structured thought, animacy constrains role assignment in a similar fashion. If we apply the evidence from the sentence processing literature to the visual realm, then the prediction is that more extreme switching costs would emerge when switching to an inanimate Agent/animate Patient event (I/A) from a two-animate-entity interaction (A/A). An alternate possibility, and one predicted by the animacy hierarchy within neural systems (Connolly et al., 2012; Scholl & Gao, 2013), is that switching costs may be observed regardless of animacy, but only between like-animacy pairs (so for A/A→A/A and I/I→I/I, but not for A/I→I/A or I/A→A/I).

A most exciting possibility is that our paradigm may be used to test the relationship among hypothesized event roles, or the degree to which role representations are distinct across different kinds of events, such as those involving implicit causality (e.g., *frighten*).

Conclusions

To summarize, a significant switching cost for event roles was observed and replicated across experiments. To put this effect in context, we believe that a separation of its practical and theoretical consequences is warranted. In terms of its absolute magnitude (on the order of 5 milliseconds), the role switching cost probably has few *practical* consequences for scene perception. But its magnitude is not relevant for our theoretical point: that an automatic mechanism for perceiving event structure must be operating, with a pervasive enough influence on visual perception and decision making that its presence is detectable in reaction times of participants engaged in orthogonal tasks. What our results suggest is that the human visual system is continuously engaged in extracting meaningful "high level" representations of the world, not just for the objects or space around us, but also for the causal relationships that are taking place between people.

Acknowledgements

Thanks to the actors, to Russell Epstein for lab space, and to Estee Ellis and Stamati Liapis for assistance in data collection. This research was supported by ANR-10-IDEX-0001-02 PSL*, ANR-10-LABX-0087 IEC, UPenn graduate research funds, and NSF Graduate Research Fellowship and NSF IGERT Traineeship (to A.H.).

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of agents: the processing of semantic roles in visual narrative. *Cognitive Psychology*, *67*(3), 73–97.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., … & Haxby, J. V. (2012). The representation of biological classes in the human brain. *The Journal of Neuroscience*, *32*(8), 2608–18.

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica*, *125*(2), 129–43.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619.

Dryer, M.S. (2013). Order of Subject, Object and Verb. In M.S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available at http://wals.info/chapter/81, Accessed 2016-01-25.)

Fillmore, C. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory*. Holt, Rinehart, & Winston.

Gruber, J.S. (1965) Studies in lexical relations. PhD Dissertation, MIT, Mass.

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology. General*, *142*(3), 880–905.

Nonyane, B. A. S., & Theobald, C. M. (2007). Design sequences for sensory studies: achieving balance for carry-over and position effects. *British Journal of Mathematical and Statistical Psychology*, *60*(Pt 2), 339–49.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different modality properties for concepts produces swtiching costs. *Psychological Science*, *14*(2), 119–124.

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure.* Cambridge, MA: MIT Press.

Saffran, E. M., Schwartz, M. F., & Linebarger, M. C. (1998). Semantic influences on thematic role assignment: Evidence from normals and aphasics. *Brain and Language*, 62(2), 255–97.

Scholl, B. J., & Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment? In M.D. Rutherford & V.A. Kuhlmeier (Eds.), *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention*. MIT Press.

Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, 63(2), 330–336.

Strickland, B. (2016). Language reflects "core" cognition: A new hypothesis about the origins of cross-linguistic regularities. *Cognitive Science*, 1-32.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74(3), 209–253.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*.

Wilson, F., Papafragou, A., Bunger, A., & Trueswell, J. (2011). Rapid Extraction of Event Participants in Caused Motion Events. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.*