



## Original Articles

# Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing

Alon Hafri<sup>a,\*</sup>, John C. Trueswell<sup>a</sup>, Brent Strickland<sup>b</sup>

<sup>a</sup> Department of Psychology, University of Pennsylvania, 425 S. University Avenue, Philadelphia, PA 19104, USA

<sup>b</sup> Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL Research University, Institut Jean Nicod, (ENS, EHESS, CNRS), 75005 Paris, France



## ARTICLE INFO

## Keywords:

Thematic roles  
Argument structure  
Scene perception  
Event perception  
Action perception  
Core cognition

## ABSTRACT

A crucial component of event recognition is understanding event roles, i.e. who acted on whom: *boy hitting girl* is different from *girl hitting boy*. We often categorize Agents (i.e. the actor) and Patients (i.e. the one acted upon) from visual input, but do we rapidly and spontaneously encode such roles even when our attention is otherwise occupied? In three experiments, participants observed a continuous sequence of two-person scenes and had to search for a target actor in each (the male/female or red/blue-shirted actor) by indicating with a button press whether the target appeared on the left or the right. Critically, although role was orthogonal to gender and shirt color, and was never explicitly mentioned, participants responded more slowly when the target's role switched from trial to trial (e.g., the male went from being the Patient to the Agent). In a final experiment, we demonstrated that this effect cannot be fully explained by differences in posture associated with Agents and Patients. Our results suggest that extraction of event structure from visual scenes is rapid and spontaneous.

## 1. Introduction

In order to successfully navigate a perceptually chaotic world and share our understanding of it with others, we must not only extract the identity of people and objects, but also the roles that they play in events: *Boy-hitting-girl* is very different from *girl-hitting-boy* even though the event category (i.e. *hitting*) and actors involved are the same. In the former, the boy is the Agent (the actor) and the girl the Patient (the one acted upon), while in the latter, their roles are reversed. The fundamental importance of such “thematic roles” has long been emphasized in linguistics: Theories of thematic roles were initially developed to account for the consistent semantic properties of grammatical arguments (e.g., *Subjects* and *Objects*) across linguistic descriptions of events (Croft, 2012; Dowty, 1991; Fillmore, 1968; Gruber, 1965; Kako, 2006; Levin & Rappaport-Hovav, 2005) but now they are also a component of some theories of conceptual representation (Jackendoff, 1990; Langacker, 1987; Talmy, 2000), development (Baillargeon et al., 2012; Leslie, 1995; Muentener & Carey, 2010; Yin & Csibra, 2015), and perception (Leslie & Keeble, 1987; Strickland, 2016) more generally.

## 1.1. Event role extraction

While there is ongoing debate within linguistics about the precise number and nature of thematic roles in language, here we are interested

in whether the mind, independently from explicit language production and comprehension tasks, rapidly and spontaneously extracts role information from perceptual input. Our work takes inspiration from a wealth of previous literature that has demonstrated rapid and bottom-up encoding of semantic content from visual scenes. These studies have revealed that categories of both objects (Biederman, Bickler, Teitelbaum, & Klatsky, 1988; Biederman, Mezzanotte, & Rabinowitz, 1982; Thorpe, Fize, & Marlot, 1996) and places (Oliva & Torralba, 2001; Potter, 1976) can be recognized from brief displays (sometimes as little as 13 ms); that the computation itself is rapid – occurring within 100–200 ms (VanRullen & Thorpe, 2001); and that the computation is relatively automatic (Greene & Fei-Fei, 2014).

In previous work we have shown that, just as with object and place categories, event category and event role information is in principle available in a bottom-up fashion from very brief displays (Hafri, Papafragou, & Trueswell, 2013; see also Dobel, Diesendruck, & Bölte, 2007; Glanemann, Zwitserlood, Bölte, & Dobel, 2016; Wilson, Papafragou, Bungler, & Trueswell, 2011). However, it is not yet known whether encoding of event information is rapid: all tasks in previous studies (to our knowledge) explicitly required participants to make a post-stimulus judgment about what was happening in the scene. Thus, the computation itself (although based on a briefly presented visual stimulus) could conceivably have continued for several seconds, up until response to the post-stimulus probe. Additionally, the

\* Corresponding author.

E-mail address: [ahafri@sas.upenn.edu](mailto:ahafri@sas.upenn.edu) (A. Hafri).

computation might have occurred only because of the explicit demands of the task, rather than being spontaneous.

### 1.2. Spontaneity and generality of role encoding

Here, we define a spontaneous process as any process that is executed independently of an explicit goal. Such a process could be automatic, in the sense that it is mandatory given certain input characteristics (Fodor, 1983), but it could also be spontaneous but not automatic in the sense that, under some conditions and with some cognitive effort, the process could be prevented from being executed (Shiffrin & Schneider, 1977). In the present work, we test for spontaneity of event role encoding.

Given the particular importance of event roles to event understanding, the spontaneity of such a process would be beneficial as we engage the social world, since at any given moment we may be performing other perceptual tasks, e.g., identifying objects or spatial properties of the scene. It would also prove useful to the young language learner tasked with mapping utterances to the events that they refer to (a problem discussed in detail in Gleitman, 1990; Pinker, 1989).

In both of these situations (social and linguistic), the utility of role information arises from its relative generality, i.e., the identification of commonality between the actors engaged in different events, such as *kicking* and *pushing* (with the degree of commonality perhaps dependent on abstract properties shared between the participants in these events, such as volition or cause; Dowty, 1991; Jackendoff, 1990; Pinker, 1989; Talmy, 2000; White, Reisinger, Rudinger, Rawlins, & Durme, 2017). However, research on action recognition using psychophysical and neuroscientific methods has largely focused on how the perceptual system differentiates between action categories (e.g., *kicking*, *pushing*, *opening*) and generalizes within action category (Hafri, Trueswell, & Epstein, 2017; Jastorff, Begliomini, Fabbri-Destro, Rizzolatti, & Orban, 2010; Oosterhof, Tipper, & Downing, 2012; Tucciarelli, Turella, Oosterhof, Weisz, & Lingnau, 2015; Wurm & Lingnau, 2015). This research has not yet addressed how we come to recognize the distinct roles that multiple actors play in visual scenes, or how (and whether) our perceptual system generalizes across the agents of different actions.

Investigating the perception of events in visual scenes provides an ideal avenue to test hypotheses about the generality of event roles. One hypothesis is that awareness of event-general properties of event roles (e.g., volition or cause) arise through explicit and deliberate observation of commonalities among event-specific roles (e.g., *kicker*, *kickee*) outside of the domain of perception (Tomasello, 2000). However, to the degree that we can find evidence that perception itself rapidly and spontaneously furnishes such event-general role information, the notion of event-specific roles as drivers of event understanding from scenes becomes less plausible. We hypothesize that in initial scene viewing, the perceptual system rapidly categorizes event participants into two broad categories – “Agent-like” and “Patient-like” (denoted Agent and Patient from here on for simplicity; Dowty, 1991; Strickland, 2016) – even if these assignments are later revised or refined in continued perceptual or cognitive processing of the event (see Section 6.1 for elaboration on these issues).

### 1.3. The current study: an event role switch cost?

The goal of the current work is to establish the degree to which the visual system gives the observer event roles “for free”, as part of routine observation of the world. We aim to show the following: (1) that the visual system encodes event roles spontaneously from visual input, even when attention is otherwise occupied (i.e. even when the observer is not explicitly asked to recognize events but rather is engaged in some orthogonal task); (2) that the computation of role itself is rapid; (3) that this encoding of event roles is at least partly event-general; and (4) that any evidence we find for encoding of event roles cannot be fully accounted for by simple visual correlates of event roles alone, such as posture.

To achieve this goal, we employed a “switch cost” paradigm (Oosterwijk et al., 2012; Pecher, Zeelenberg, & Barsalou, 2003; Spence, Nicholls, & Driver, 2001). In several experiments, participants observed a continuous sequence of two-person scenes and had to rapidly identify the side of a target actor in each (Experiments 1a and 1b: male or female actor; Experiments 2 and 3: blue- or red-shirted actor). With our design, event role identities provide no meaningful information for the primary task of gender or color identification, so observers need not attend to such irrelevant information. Nevertheless, we hypothesized that when people attend to the target actor to plan a response, then if event roles are spontaneously encoded, they should “come along for the ride.” Thus, we should be able to observe an influence of this role encoding on responses even though event roles are irrelevant to the primary task.

More specifically, we reasoned that if role assignment is spontaneously engaged, then when the role of the target actor switched from trial to trial, it would result in a cost, i.e., a relative lag in reaction time, even though subjects were tasked with identifying a property orthogonal to roles (here, gender or shirt color). If such a pattern were observed, it would provide compelling evidence that analysis of event structure from visual scenes is a rapid, spontaneous process that is engaged even when we are attending to other perceptual information. Furthermore, by using simple tasks based on visual information known to be rapidly available (including gender; Mouchetant-Rostaing, Giard, Bentin, Aguera, & Pernier, 2000), we expected that observers would respond quickly, allowing us to test the rapidity of extraction of event role information.

## 2. Experiment 1a

Participants observed a series of simple still images displaying an interaction between a male and a female, and were simply asked to say whether the male/female was on the left or right of the screen. We predicted that although the task fails to actively encourage role encoding (and may even discourage it), participants would nevertheless be slower on trials in which the event role of the target actor differed from his or her role in the previous trial, i.e., a “role switch cost”.<sup>1</sup>

### 2.1. Method

#### 2.1.1. Participants

Twenty-four members of the University of Pennsylvania community participated and received either class credit or \$10. Because we were collecting a large number of trials within-participant (see Section 2.1.3 below), we predicted that this number of participants would be sufficient to observe the role switch cost, if it were to exist. All participants in this experiment and in the other experiments reported below gave informed consent, following procedures approved by the university’s institutional review board.

#### 2.1.2. Materials

The stimuli were 40 color photographic images depicting 10 two-participant event categories taken from a previous study that investigated extraction of event categories and roles from briefly displayed and masked images (Hafri et al., 2013). The event categories used were *brushing*, *chasing*, *feeding*, *filming*, *kicking*, *looking*, *punching*, *pushing*, *scratching*, *tapping*. These categories were chosen because they showed the highest agreement among subjects for role assignment from brief display (i.e., male as Agent or Patient) in the previous study. All stimuli were normed for event category and role agreement in the previous study.

<sup>1</sup> We cannot differentiate between switch costs vs. repetition benefits (priming) because there is no baseline for comparison, but in keeping with the terminology in previous investigations using this paradigm (e.g., Pecher et al., 2003), we use the term switch costs. Whether the effects are a benefit or cost does not qualitatively change our conclusions.

Download English Version:

<https://daneshyari.com/en/article/7285330>

Download Persian Version:

<https://daneshyari.com/article/7285330>

[Daneshyari.com](https://daneshyari.com)