



HAL
open science

The Effects of Event Knowledge and Semantics in the Guidance of Fixations on Scenes

Sophie Heer, Marek Pedziwiatr, Peter Bex, Antoine Coutrot, Isabelle Mareschal

► **To cite this version:**

Sophie Heer, Marek Pedziwiatr, Peter Bex, Antoine Coutrot, Isabelle Mareschal. The Effects of Event Knowledge and Semantics in the Guidance of Fixations on Scenes. European Conference on Visual Perception, Aug 2022, Nijmegen, Netherlands. . hal-03766568

HAL Id: hal-03766568

<https://hal.science/hal-03766568>

Submitted on 1 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Introduction

A person viewing a real-world event typically accesses two sources of non-visual information about it: knowledge of the preceding events and a general knowledge about the objects commonly found in a given environment. We investigated how these two types of knowledge jointly guide human gaze.

Methods

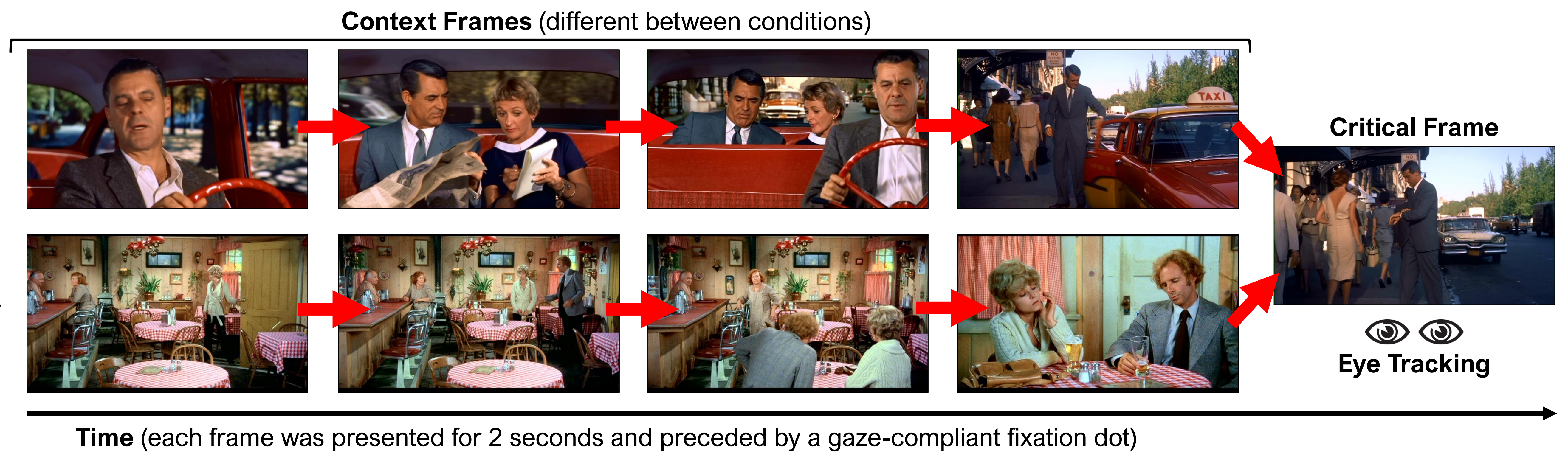


Figure 1. Participants ($N = 48$) freely viewed 80 sequences of film frames (retrieved from https://the.hitchcock.zone/wiki/1000_Frames_of_Hitchcock). Each sequence contained 3-8 'context' frames followed by a 'critical' frame. Context frames in a sequence were related to each other and showed a consistent unfolding of events over time. In the Continuous condition, the critical frame was a natural continuation of these events. In the Discontinuous condition, the critical frame was unrelated to the context frames.

Object Fixations



Figure 2. Analysis was restricted to fixations on objects within critical frames.

Image Segmentation and Labelling



Figure 3. Objects within critical frames were (a) manually segmented and (b) labelled using MATLAB's Image Labeler app.

Semantic Similarity Calculation

a

man	1	0.49	0.21	0.27	0.3	0.32	0.34	0.26	0.26
people	0.49	1	0.39	0.25	0.37	0.28	0.35	0.29	0.4
buildings	0.21	0.39	1	0.43	0.49	0.33	0.36	0.19	0.39
sidewalk	0.27	0.25	0.43	1	0.44	0.32	0.48	0.38	0.29
trees	0.3	0.37	0.49	0.44	1	0.46	0.38	0.15	0.29
sky	0.32	0.28	0.33	0.32	0.46	1	0.33	0.26	0.25
road	0.34	0.35	0.36	0.48	0.38	0.33	1	0.47	0.52
taxi	0.26	0.29	0.19	0.38	0.15	0.26	0.47	1	0.47
cars	0.26	0.4	0.39	0.29	0.29	0.25	0.52	0.47	1

b

man	0.27
taxi	0.28
sky	0.28
buildings	0.31
people	0.31
sidewalk	0.32
cars	0.32
trees	0.32
road	0.36

Figure 4. (a) The GloVe word embedding (Pennington, Socher & Manning, 2014) was used to calculate semantic similarity scores between an object label and all other labels within a frame. (b) Then, we took the average of these scores. This resulted in a semantic similarity score for each object within each critical frame.

General Linear Mixed-Effects Model

Object Fixations ~ 1 + Object Size * Semantic Similarity + Condition * Semantic Similarity + (1 + Semantic Similarity | Frame)

Results

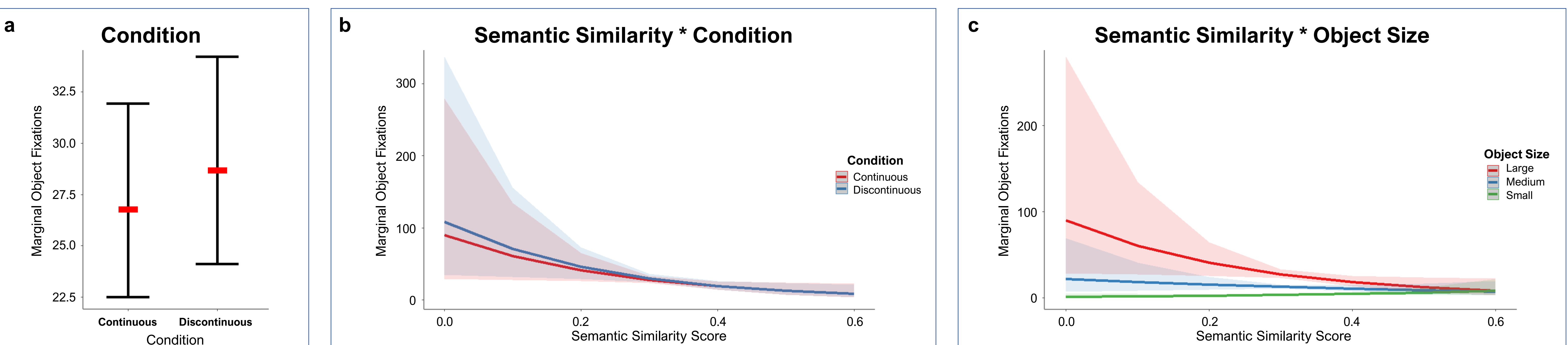


Figure 5. The marginal number of object fixations and their 95% confidence intervals by (a) condition, (b) the interaction between semantic similarity and condition and (c) the interaction between semantic similarity and object size. (a) Objects presented in the Discontinuous condition received more fixations compared to the Continuous condition ($p < .001$). (b) As objects' semantic similarity scores decreased, the number of fixations on those objects increased ($p = .026$). This effect was weaker in the Continuous condition ($p = .014$). (c) This effect was reduced for medium ($p < .001$) and small objects ($p < .001$) relative to large objects.

Discussion

Here, we show that knowledge about object-scene relationships influences how people look at scenes. Specifically, semantically inconsistent objects attract more fixations. However, this effect is modulated by object size. Furthermore, we find that the degree to which this knowledge about object-scene relationships is deployed in the guidance of eye movements is modulated by the possession of information about recent past events: inconsistent objects exert weaker influence on eye movements when they are presented in a continuous sequence of events. Taken together, we show how two types of prior knowledge, general knowledge about objects in the world and knowledge of recent past events, interact when guiding human gaze.