

Investigating the conditions under which movie viewers' event models influence their attention

by

Prasanth Chandran

M.S., Indian Institute of Science Education and Research Pune, 2017

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Abstract

Film, television, and video play a major role in daily information consumption, making it important to understand how we comprehend dynamic scenes and how this shapes visual attention. Although event models—mental representations of ongoing events—are known to guide attention during reading and static scene viewing, their influence on attentional selection in films is minimal or absent, a pattern known as the Tyranny of Film. This project provides an existence proof that event models can influence selective visual attention in commercial film clips. Based on the Scene Perception & Event Comprehension Theory and Biased Competition Theory, I propose two conditions under which event models should guide attention: 1) The target must be prioritized differently because of clear differences in viewers' event models (Differential Event Model Criterion); and 2) The target must have low bottom-up default gaze probability, making it unlikely to attract attention automatically (Low Default Gaze Probability Criterion). I validated the first criterion using linguistic and moment-to-moment spatial importance measures, and the second using DeepGaze IIE predictions, identifying eight film clips that met both criteria. In a within-subjects Eye-Tracking study ($N = 44$), participants viewed each clip without context and then with context. Targets were fixated four times more often in the Context condition, whereas fixation rates in the No-context condition were at chance. Moreover, *a priori* event-based importance predicted attention more strongly than *a priori* default-gaze predictions when participants were aware of the context. These findings demonstrate that when both criteria are satisfied, event models can influence attentional selection during film viewing, overcoming the Tyranny of Film and offering strong support for SPECT.

Investigating the conditions under which movie viewers' event models influence their attention

by

Prasanth Chandran

M.S., Indian Institute of Science Education and Research Pune, 2017

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Approved by:

Major Professor
Dr. Lester Loschky

Copyright

© Prasanth Chandran 2026.

Abstract

Film, television, and video play a major role in daily information consumption, making it important to understand how we comprehend dynamic scenes and how this shapes visual attention. Although event models—mental representations of ongoing events—are known to guide attention during reading and static scene viewing, their influence on attentional selection in films is minimal or absent, a pattern known as the Tyranny of Film. This project provides an existence proof that event models can influence selective visual attention in commercial film clips. Based on the Scene Perception & Event Comprehension Theory and Biased Competition Theory, I propose two conditions under which event models should guide attention: 1) The target must be prioritized differently because of clear differences in viewers' event models (Differential Event Model Criterion); and 2) The target must have low bottom-up default gaze probability, making it unlikely to attract attention automatically (Low Default Gaze Probability Criterion). I validated the first criterion using linguistic and moment-to-moment spatial importance measures, and the second using DeepGaze IIE predictions, identifying eight film clips that met both criteria. In a within-subjects Eye-Tracking study ($N = 44$), participants viewed each clip without context and then with context. Targets were fixated four times more often in the Context condition, whereas fixation rates in the No-context condition were at chance. Moreover, *a priori* event-based importance predicted attention more strongly than *a priori* default-gaze predictions when participants were aware of the context. These findings demonstrate that when both criteria are satisfied, event models can influence attentional selection during film viewing, overcoming the Tyranny of Film and offering strong support for SPECT.

Table of Contents

List of Figures	x
List of Tables	xii
Acknowledgements	xiii
Chapter 1 - Investigating the conditions under which movie viewers' event models influence their attention	1
How do we understand visual narratives?	5
What determines where we will attend next in films?	7
Bottom-up factors	8
Top-down factors	8
Priority maps and Biased Competition Theory	9
Chapter 2 - Developing a Systematic Approach to Studying the Effects of Event Models on Attentional Selection in Film Viewing	14
The Differential Event Model Criterion (Criteria 1)	16
Low Default Gaze Probability Criterion (Criteria 2)	17
Study Plan	23
Chapter 3 - Preliminary Studies and Analyses to Verify the Two Criteria	25
Verification of the Event Model Criterion	25
Preliminary Study 1 (Prediction and Comprehension Task)	25
Method.	26
Procedure.	26
Analysis.....	27
Results.....	32
Preliminary Study 2 (Importance Judgment Task)	34
Research Question.	34
Event Model Importance Hypothesis.....	34
Null Hypothesis.	35
Stimuli.....	35
Method.	35
Analyses.....	36

Target Areas of Interests Analyses.	36
Frame Order.	37
Results.....	38
Proportion of Time Spent within AOIs.....	38
Verification of Low Default Gaze Probability Criterion	43
DeepGaze IIE Study	44
Chapter 4 - Eye-Tracking Study	51
Attentional Selection Hypotheses	51
Event Model Hypothesis	51
Tyranny of Film Hypothesis	52
Methods	52
Participants.....	52
Sample Size rationale.....	53
Materials	56
Data Recording	56
Study Design.....	57
Procedure	58
Data Cleaning measures.....	60
Data Variables.....	60
Proportion of fixations within AOIs	60
A priori Predictions from Preliminary Studies	61
Stimulus-based Default Gaze prediction ($P(D)$).....	61
Event-based Importance prediction ($P(I)$).	62
Analysis Plan	62
Proportion of fixations within AOIs	62
Relative Weights of stimulus-based default gaze probability and event-based importance.	63
Results.....	65
Proportion of fixations within AOIs	65
Relative Weights of stimulus-based default gaze probability and event-based importance.	70
Chapter 5 - Discussion	74
Broader Application of the Current Method of Breaking the Tyranny of Film	77

Event model and default gaze congruence spectrum of videos	78
Role of familiarity	80
Role of individual differential event elements in attentional selection	80
Implications for filmmaking and viewers' attentional modulation	81
Chapter 6 - Conclusion	84
References	85
Appendix A - Videos and their targets of attention	95
Burning	95
Cache	96
Code Unknown	97
Community	99
Crimson Gold Version 1 & Version 2	100
Dogtooth	101
Drive	102
Mr. Robot.....	103
Ocean's Thirteen.....	104
Revanche.....	105
Spoonloos Version 1 & Version 2.....	106
The Fire Within Version 1 & Version 2	107
The Passenger	108
The Secret in Their Eyes.....	109
Appendix B - Schema for Steps in Creating a Context Manipulation.....	110
Appendix C - Preliminary Experiment 1 Parameter Estimates	111
Appendix D - Selecting Effective Videos.....	112
Appendix E - Radar Charts for "ineffective" clips	113
Appendix F - Video Specifications.....	115
Appendix G - Complex Random Intercept Selection and Nested Regression for Mouse Hit	
Probability Models.....	116
Appendix H - Graphs for Frame Order and Gaze Probability	118
Appendix I - Power Analysis Effect Size Comparisons	132
Appendix J - Video set order	133

Appendix K - Selecting Random Effect Structure for Fixation Hit Probability	135
Appendix L - Nested Model Comparison for Hypothesis Testing for Fixation Hit Probability	136
Appendix M - Frame-by-frame changes in the raw average probability values of fixation hit probability and the two a priori predictions.....	137
Appendix N - Nested Model Comparison to Test the Relative Influence of Stimulus and Event Model.....	146

List of Figures

Figure 1	2
Figure 2	3
Figure 3	6
Figure 4	11
Figure 5	18
Figure 6	20
Figure 7	22
Figure 8	26
Figure 9	31
Figure 10	33
Figure 11	37
Figure 12	41
Figure 13	42
Figure 14	46
Figure 15	48
Figure 16	57
Figure 17	67
Figure 18	69
Figure 19	72
Figure 20	112
Figure 21	113
Figure 22	113
Figure 23	118
Figure 24	119
Figure 25	120
Figure 26	121
Figure 27	122
Figure 28	123
Figure 29	124

Figure 30	125
Figure 31	126
Figure 32	127
Figure 33	128
Figure 34	129
Figure 35	130
Figure 36	131
Figure 37	137
Figure 38	138
Figure 39	139
Figure 40	140
Figure 41	141
Figure 42	142
Figure 43	143
Figure 44	144
Figure 45	145

List of Tables

Table 1	28
Table 2	40
Table 3	43
Table 4	49
Table 5	66
Table 6	71
Table 7	111
Table 8	115
Table 9	116
Table 10	117
Table 11	132
Table 12	135
Table 13	136
Table 14	146

Acknowledgements

I want to express my heartfelt gratitude to my advisor, Dr. Lester Loschky, for his unwavering support and guidance throughout my journey at Kansas State University. His welcoming attitude towards my ideas, combined with his extensive knowledge and meticulous attention to detail, truly helped me bring my ideas to life. I have learned so much and grown in my approach to research and writing under his mentorship. I am also deeply grateful to my committee members: Dr. Heather Bailey (Kansas State University), who has been a tremendous support from the very start of my dissertation; Dr. Michael Young, who has greatly helped me sharpen my statistical skills; Dr. Joseph P. Magliano (Georgia State University), who has been there from the beginning of my research project and assisted with content analyses of my stimuli; Dr. Tim Smith (University of Arts London), whose inspiring guidance has always helped me find the right perspective; and finally, Dr. Matt Wineswski, who kindly accepted my late invitation to join the committee and has been incredibly supportive in helping me reach this milestone. I also want to thank Dr. Smith, Dr. Magliano, and Dr. Bailey for always treating me like their own student whenever we met at conferences, taking the initiative to support my work, and connecting me with other experts in the field.

Additionally, I am grateful to my many current and former graduate and undergraduate research collaborators who provided valuable feedback, assisted with participant recruitment and data analysis, and offered emotional support and encouragement along the way. My undergraduate collaborators include Bailey Dreyfuss, Camille Dortland, Claire Holt, Eva Naab, Karla Salto, Laurel Conner, Maddy Pilarz, Hannah Burns, Alyssa Smith, Ashton Lofing, Lindsey Wilson, Brayden Wallace, Julian Nuwamanya, Isabella Hubbell, Cashel Fitzgibbons, John

Pagen, Sam Schultz, and Will Smith. My graduate collaborators are Dr. John Hutson, Dr. Maverick Smith, Dr. Taylor Vargo-Simonson, Dr. Kari Payne, and Amir Nikbhakht.

A significant part of the analyses in this dissertation wouldn't have been possible without the incredible support of Dr. Kari Payne, who was instrumental in setting up and managing the deep learning algorithms used in my work. Additionally, all of these analyses were made possible by the wonderful Beocat Research Cluster at Kansas State University, which is partly supported by the Cognitive and Neurobiological Approaches to Plasticity (CNAP) Center of Biomedical Research Excellence (COBRE) of the National Institutes of Health under grant number P20GM113109.

Finally, I want to express my heartfelt gratitude to my family—my mother, sister, and father—for their unwavering support, love, and encouragement. I especially want to thank my mother, who I believe has had even more sleepless nights than I have, worrying about my dissertation.

Chapter 1 - Investigating the conditions under which movie viewers' event models influence their attention

Dynamic audiovisual media, such as films, TV shows, YouTube videos, and more, are ubiquitous in our daily lives. In 2021, US adults spent an average of 167 minutes per day watching TV and 149 minutes per day watching digital video, for a total of roughly 5 hours per day just watching videos (Stoll, 2024). In sum, we spend a lot of time watching and processing videos. Hence, developing and refining theories of how we make sense of media and the role of eye movements in that process is essential. Video stimuli are characterized by high spatiotemporal information density, typically presenting over 24 frames per second, which can exceed the capacity of the human visual system. To manage these limited cognitive resources, the brain must selectively prioritize specific streams of sensory data for detailed processing. This is achieved through overt attentional selection, in which eye movements fixate on a smaller region of the scene. Because this focus on a selective region (foveal information) necessitates the exclusion of what's outside that region (peripheral information), our attentional shifts act as a filter that fundamentally shapes our moment-to-moment understanding of the dynamic scene (see, e.g., Smith, 2012b).

Subsequently, does your understanding of what happened in a video determine where you will look next? Generally, after watching a video, viewers develop a similar understanding of the narrative despite the complexity of the audiovisual stimuli and significant individual differences among viewers. One potential reason is that, over the last 100+ years, filmmakers have intuitively developed highly systematic filmmaking techniques to very effectively direct all viewers' attention to the same places simultaneously (Bordwell et al., 2024). However, these intuitions have rarely been empirically validated (but see Smith, 2012a; Cutting, 2021).

Similarly, we can also have different narrative experiences despite watching the same video. Consider the following scenario: your significant other has been watching the movie *Ocean's Thirteen* (Soderbergh, 2007) on your couch. You join them on the sofa only towards the end of the film and see the last 10 seconds of its final scene, as shown in Figure 1. Let us assume you have not seen this movie before for this example. You see that Brad Pitt's character is at an airport, walking toward the camera with a little smile. After watching the scene in Figure 1, you would most likely overtly attend to Brad Pitt in the foreground and predict that he will either take a flight or leave the airport. If you were to make this prediction, even though you would not be wrong, your prediction would lack critical information and be incomplete.

Figure 1

Frame from Ocean's Thirteen



On the contrary, your roommate, who has been watching the movie from the beginning, may have a different prediction and be aware of related critical information. Hence, you both are looking at the same TV screen, but each understands it differently. In this example, we will say that you have “jumped-in-the-middle” (i.e., joined late) (Loschky et al., 2015). In my experimental procedure, you would be considered to be in the *No-context* condition (as you do

not know the full context of the scene). In contrast, your roommate, who has been watching the film from the beginning, would be considered as being in the *Context* condition. The shared portion of the video you both watch will be referred to as the *Common Viewing Period*.

Figure 2

Frame from a film clip from Ocean's Thirteen showing what happened before the previous clip.



Let us now briefly show you what your roommate has already learned while watching from the beginning. Here is a highly condensed description of the context: Brad Pitt's character is a con man who likes stealing from rich companies to give money to poor people. He has designed a “special” coin that makes slot machines hit the jackpot. Just before this scene, he handed out such a coin to a stranger in the background (as shown in Figure 2), so the stranger could win a fortune. The stranger is near a slot machine. Hence, as your roommate watches the scene, they will most likely notice the stranger putting a coin into the slot machine in the background, while Brad Pitt smiles with satisfaction and walks towards the screen. Your roommate must be predicting that the stranger will win the jackpot any moment now. Hence, you and your roommate will have different higher-level understandings of the story while watching the same *Common Viewing Period* clip, as reflected in your predictions of what will happen next.

Our key question is whether your and your roommate's eye movements show corresponding differences, or will they be the same? The answer to this question has important implications for theories of attention and comprehension, as well as real-world applications in mass media and education.

Based on the literature on eye movements while reading (see, for review, Rayner, 1998), and while viewing static scene images (see, for review, Williams & Castelhana, 2019; other e.g., Hutson et al., 2022; Hutson et al., 2018), differences in event understanding are found to be associated with differences in attentional selection. One would predict that the dynamic scenes in film will show the same. However, this prediction assumes that film comprehension and its attentional control are similar to those of textual or comics-based narratives. As in the scenario above, one can recreate the jumped-in-the-middle paradigm in a lab setting by having one group of participants watch the scene from the “middle” without any context, and another group views the entire scene after reading a highly condensed context. The few studies that have used the jumped-in-the-middle paradigm with videos to study such differences in event models have successfully created significant differences in understanding. However, they all showed minimal or no influence on viewers' attentional selection (Hutson et al., 2017; Loschky et al., 2015). One such study (Hutson et al., 2017) used the opening scene of Orson Welles' film, *Touch of Evil* (Welles & Zugsmith, 1958), and, from now on, I would refer to it as the *Touch of Evil* dynamic study. Viewers in the No-context condition jumped in the middle of the clip and saw a couple walking through city streets, leading to a border crossing. Viewers in the Context condition watched the scene from the beginning, and knew that a bomb was placed inside a crucial car at the border crossing, giving them a richer event model. We would expect viewers in the two conditions to look at different locations on the screen depending on their condition (i.e., the car

with the bomb is “important” to people in the Context condition but not to those in the No-context condition). However, contrary to expectations, viewers' eye movements were very similar across the two viewing conditions. This phenomenon, in which differences in eye movements did not accompany differences in understanding, was called *the Tyranny of Film* (Davis et al., 2021; Huff et al., 2014; Hutson et al., 2017; Loschky et al., 2015). Is the Tyranny of film unique to dynamic stimuli or Hollywood-style edited clips? Does that indicate that moment-to-moment comprehension in dynamic stimuli will not affect attentional selection under any circumstances? To address these questions, let's explore comprehension and attention in dynamic visual narratives.

How do we understand visual narratives?

Visual narratives are composed of smaller events, just as any other narrative. Based on studies of event comprehension, viewers' understanding of events is theorized to be represented in terms of event models, mental models that capture the viewer's understanding of the event. According to the event indexing theory, these event models are made up of various building blocks known as Event indices (Zwaan, Langston & Graesser, 1995; Zwaan, Magliano & Graesser, 1995), such as characters, actions, goals, space, time, causation, and emotions (Magliano et al., 2012). For example, in the Ocean's Thirteen film clip, *the stranger in the background* is one of the characters in your roommate's event model.

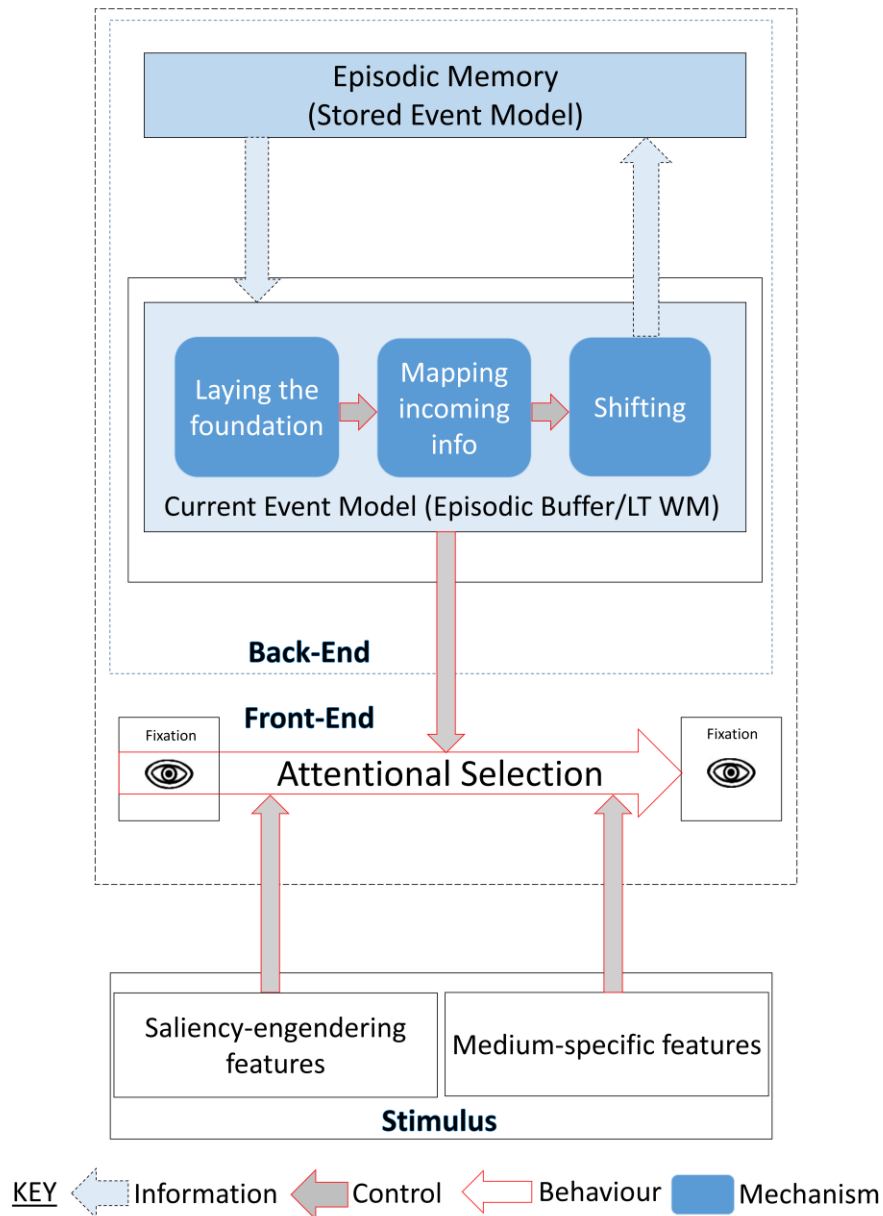
However, how does one's moment-to-moment understanding of the event develop during online video viewing? The visual system extracts information (such as the event indices) from the video, which is then used to create the current event model. Based on the *Structure Building Framework* (Gernsbacher, 2013), which is incorporated in the *Scene Perception & Event Cognition Theory* (SPECT)(Loschky et al., 2018, 2020, 2025), the creation of a new current

event model will be initiated with the presentation of a new video in the viewer's Episodic Buffer/Long-Term Working memory.

Figure 3

Simplified Scene Perception & Event Comprehension Theory (SPECT) to show influences on attentional selection.

Scene Perception & Event Comprehension Theory (SPECT) – Attentional Selection



The process starts by laying the foundation for the event model, including basic information such as one or more characters, their actions, and spatio-temporal information. Viewers will then map new incoming information to the current event model if it is coherent with the existing foundation. Finally, suppose the incoming information is no longer coherent with the current event model. In that case, the current event model is moved to long-term episodic memory, where it becomes a stored event model (Radvansky, 2012), in a process known as shifting, and a new current event model is created by laying a new foundation (see Figure 3). SPECT assumes that the stored event model in episodic memory facilitates all three aspects of the current event model, primarily helping to generate predictions about upcoming events. Facilitation by the stored event model is evident in studies that used the jumped-in-the-middle paradigm and found “richer”¹ and more accurate predictions in the Context condition (e.g., Hutson et al., 2017; Loschky et al., 2015).

What determines where we will attend next in films?

I will focus only on how the previously stored event model can influence attentional selection in dynamic visual scenes, as well as on moment-to-moment changes in the current event model. SPECT predicts that attentional selection is affected by bottom-up factors, primarily including stimulus features (visual saliency) and medium-specific features, or by top-down factors, including the current event model and long-term memory (Loschky et al., 2018, 2020, 2025).

¹ “richer” event model or predictions essentially means that the event model or, the prediction is very detailed. Hence, “rich” with information.

Bottom-up factors

Visual saliency can capture the viewer's attention irrespective of the stimulus's medium and the scene's content. Some of the significant visual saliency features that are known to capture attention are: *motion*, i.e., moving items attract attention (Carmi & Itti, 2006; Mital et al., 2011; Theeuwes et al., 1999); *blurred items* are less salient and rarely attended (Enns & MacDonald, 2013; Peterson 2018); and Animate objects such as people and animals capture attention (Fletcher-Watson, 2008; Humphrey & Underwood, 2010), especially faces (for example, Langton et al., 2008). Medium-specific features can also grab attention. One such feature is that, in still images and edited videos, viewers tend to attend to the screen center, also known as the center bias (Dorr et al., 2010; Kümmerer et al., 2017; Tatler et al., 2011).

Note that a current state-of-the-art gaze prediction algorithm, DeepGaze IIE, is attracted by both low-level visual feature contrast (the essence of visual saliency—Itti & Koch, 2000) and high-level visual objects, such as human faces and written text (Linardos, Kümmerer, Press, & Bethge, 2021). This better reflects the default probability of what human viewers fixate than what has hitherto been called “saliency.” Nevertheless, the factors influencing such gaze predictions are far more of a “black box” than the more analytically transparent Itti-Koch type of saliency. As discussed in Stimulus-based Default Gaze Probability Criterion and Verification of Default Gaze Probability Criterion, I will use DeepGaze IIE’s default gaze prediction as representing all stimulus-based (bottom-up) influences on attentional selection.

Top-down factors

Overall, there are many well known and well-studied top-down influences on attentional selection, for example the role of viewer's task on attentional selection, such as when they were asked to infer certain specific information from the image other than the story (DeAngelus &

Pelz, 2009; Yarbus, 1967); when asked to take a perspective while watching a scene, for example, looking at photos of homes from the perspective of either a burglar or a home buyer (Anderson & Pichert, 1978; Kaakinen et al., 2011); or watching a scene while taking the perspective of either a detective or an interior decorator (Lahnakoski et al., 2014); or when asked to draw a map of the scene after watching it (Hutson et al., 2017); or identify the scene depicted in a video (Smith & Mital, 2013). Apart from that, items in working memory are also shown to be prioritized for visual attentional selection (Olivers et al., 2006; Soto et al., 2005). However, my current study will focus on the effect of the event model on attentional selection. As you can see from the simplified version of SPECT in Figure 3, SPECT predicts the event model to influence attentional selection. However, as mentioned above, the viewer's event model while watching films/dynamic stimuli has been shown to have minimal or no influence on attentional selection, a.k.a., the Tyranny of Film (Davis et al., 2021; Huff et al., 2014; Hutson et al., 2017; Loschky et al., 2015). Finally, it is unclear whether retrieving the stored event model from Long-term Episodic Memory affects the attentional control guided by the current event model.

Priority maps and Biased Competition Theory

As I have discussed so far, many factors, both top-down and bottom-up, influence attentional selection. It is therefore vital to understand that when multiple such factors compete for attentional selection, what determines where the viewer will end up looking, since you can only overtly attend to one thing at a time. One theory that explains how these attentional selection mechanisms work together is the *Biased Competition Theory (BCT)*. BCT proposes that visual attention is driven by winner-takes-all neural competition among object representations, in which bottom-up visual information is weighted and filtered at various levels of processing by top-down modulations (Adeli & Zelinsky, 2018; Desimone & Duncan, 1995; Ptak, 2012; Tsotsos

et al., 1995). Hence, BCT involves the interaction between components I briefly discussed above: 1) bottom -up competition and 2) top-down bias.

As visual information from a scene enters the brain, it is processed across multiple retinotopic “feature maps” (Figure 4) that represent basic visual elements, such as color, luminance, motion, and orientation (Itti & Koch, 2000). These feature maps enter a neural competition to select a “winner” (eg, red object in a group of blue ones), creating a bottom-up saliency map. There will also be higher-level stimulus feature maps, such as faces and written text, which are known to reflexively capture viewers’ attention (Cerf, Frady & Koch, 2009). Hence, I could refer to the bottom-saliency map containing the winners from both low-level and high-level stimulus features as *default gaze probability maps*. If there is no top-down biasing, then attention should go to the winner solely based on these default gaze probability maps.

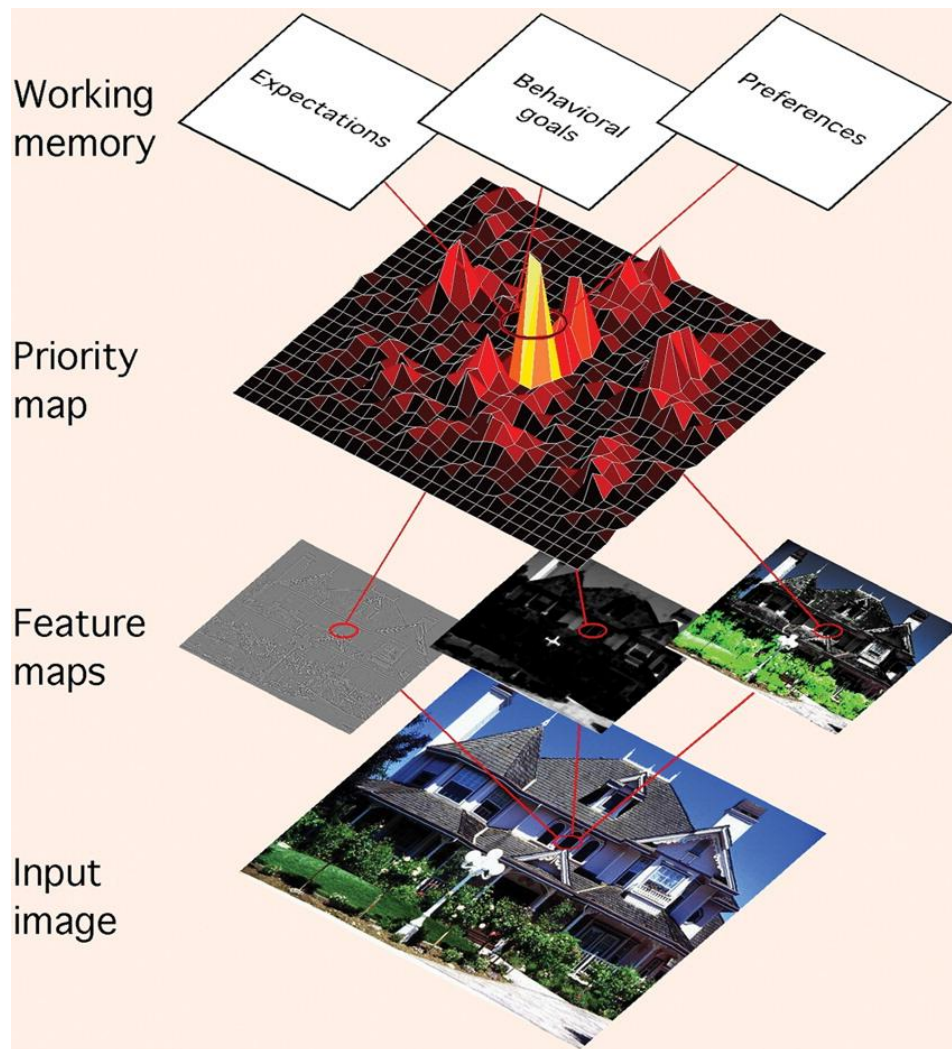
However, Biased Competition Theory suggests that feedback from top-down processing, such as goal setting and task assignments, could bias the default gaze probability maps through feedback connections (Adeli & Zelinsky, 2018; Hamker, 2004; Tsotsos et al., 1995), weighting and filtering bottom-up information to prioritize task-relevant information. These biases amplify task-relevant information, transforming the purely stimulus-based default gaze probability map into a *priority map* that dictates the final attentional selection. Priority maps are 2D retinotopic² maps generated in the Superior Colliculus (Adeli et al., 2017) that represent the spatial mapping of attentional prioritization across different regions of the input image (Zelinsky & Bisley, 2015). Greater attentional priority in these maps is reflected in higher activity at that retinal location (see Figure 4). Hence, according to BCT, goal- or task-biasing has the strongest influence on

² Associated with various locations on the retina, which will also correspond to various locations in the input image. Importantly, the spatial relations among these retinal locations, such as left-right, and up-down, are preserved in various visual processing brain areas that code for a variety of visual features (e.g., color, orientation, motion, etc.).

attentional selection. In the absence of such goal-directed maps (for example, when viewers are engaging in a free-viewing³ task), strong visual stimulus features such as motion heavily influence attentional selection.

Figure 4

Attentional priority maps and how they are influenced by stimulus-based Feature maps and other Working memory-based top-down factors (Ptak, 2012; p. 503).



Note. The input image is preattentively decomposed into multiple independent feature maps in the early visual cortex (e.g., spatial frequency, intensity, and color as shown above from left to

³ In Free-viewing tasks, participants are asked to just watch the scene without engaging in any secondary-task. They are essentially free to look at any part of the screen.

right, respectively). The priority map combines inputs from these feature maps and top-down signals based on their expectations, behavioral goals, and preferences. The location where the summation of these inputs from feature maps and top-down signals peaks is prioritized for attention.

The Scene Perception & Event Comprehension Theory (SPECT) incorporates BCT to explain how a viewer's moment-to-moment narrative comprehension dictates where they look. Traditionally, BCT attributes top-down biases to explicit tasks (e.g., “find waldo”) or prior schemas. SPECT expands this by proposing that the viewer's current event model can also exert this top-down bias (Loschky et al., 2025).

Hence, one potential explanation for the Tyranny of film, based on the Biased competition Theory, is that the visual salience of motion in films is highly overpowering and heavily influences the selection of gaze location (Hutson et al., 2022). It has been hypothesized that this strong influence of motion in films prevents the event model from biasing competition and affecting attentional selection (Loschky et al., 2015; Hutson et al., 2017). Support for this hypothesis emerged when a slideshow version of the Touch of Evil opening shot, in the absence of motion, produced effects consistent with the event model on attentional selection (Hutson et al., 2022). In contrast, this effect was absent when viewers watched the standard movie clip (Hutson et al., 2017). Hence, when bottom-up saliency becomes weaker, the event model is more likely to bias the event-relevant information to win the neural competition for attentional selection. Also see Smith & Mital (2013) for a similar demonstration of the influence of motion on gaze in matched static/dynamic scenes. It is also important to note that even with the heavy influence of motion, tasks and goals have still shown the ability to introduce spatial bias for task-relevant objects or information in the scene, as discussed previously in the top-down factors

section (e.g., Lahnakoski et al., 2014; Hutson et al, 2017). Hence, in films, it is possible to introduce top-down biasing and control attentional selection. However, there are exceptions, such as Taya et al. (2012), who found no difference in eye movements when observers were asked to judge who won the points during a short clip of a Tennis match compared to when they were observing the clip.

Chapter 2 - Developing a Systematic Approach to Studying the Effects of Event Models on Attentional Selection in Film Viewing

Most studies testing the effects of the event model on attentional selection used narrative films created by highly skilled filmmakers who aim to direct your attention to the most important elements of the narrative. They generally achieve that by prioritizing bottom-up stimulus features to guide the viewer's attention rather than relying just on the viewer's event model, which may lead to more idiosyncratic viewing. Hence, typically in conventional film clips, there is a strong association between the high visual saliency of the target and its importance based on the event model, making the attentional selection of the target very automatic and effortless, which allows the viewer to continue following the story easily (Smith, 2012a; Smith, 2012b). This will likely lead the target to be reasonably described as "pop-out" for the film viewer, similar to an effortless visual search for a highly salient target (Braun, 1998; Wolfe, 2004). In accordance with that, the robust evidence for the Tyranny of Film from studies using narrative films (Davis et al., 2021; Huff et al., 2014; Hutson et al., 2017; Loschky et al., 2015) also indicates that visual saliency has a greater priority than event models on attentional selection in conventional films. Hence, to isolate the top-down effect of the event model on attentional selection, I must find examples that have a dissociation between bottom-up influences and the event model. I believe there are situations in which the viewer's event model will likely be prioritized over bottom-up influences in narrative films. Consider the scene of *Ocean's Thirteen* (Soderbergh, 2007), where the filmmaker does not keep the stranger at the slot machine in the background in focus, even though he is an important character. Moreover, even though the shot is 10 seconds long, the filmmaker also avoids cutting to a close-up of the stranger in the background to show his interaction with the slot machine in detail. And probably because this climax shot is only 10s, the

director is relying on the viewer's event model solely to guide their attention to a critical character and not utilizing conventional filmmaking techniques (e.g., pulling camera focus to the important character, editing between shots, etc.) to create high salience to attract the viewer's attention. This is likely for dramatic effect, because the filmmaker may want the viewer to figure out what will happen next based on what they know happened before, and by doing so, giving them the same level of satisfaction that Brad Pitt's character has, who is also aware and proud of the obvious result of what he just did (i.e., he has allowed the stranger to win the jackpot using the special coin he gave him).

I believe this scene offers an opportunity to demonstrate how the film viewer's event model shapes their attentional selection during viewing. Thus, an essential methodological direction for testing the effect of event comprehension on attentional selection is to develop a systematic approach to investigating how the viewer's event model influences their attentional selection in film clips. Here, I propose an approach that could help anyone find examples like the *Ocean's Thirteen* clip, in which there is a dissociation between bottom-up default gaze predictions and the viewer's event model. This approach utilizes the jumped-in-the-middle paradigm where 1) the Context condition viewer's event model for the target information is rich compared to the No-context condition, hence, potentially strong enough to win the biased competition for attentional selection; and 2) the bottom-up driven default gaze prediction for the target information is weak according to a state-of-the art model for predicting fixations, avoiding "pop-out" of the target, and hence is unlikely to be looked at without a strong event model's guidance. I discuss these two criteria in detail below.

The Differential Event Model Criterion (Criteria 1)

A strong event model must differentially prioritize the target of attention. The differences in event models that do versus do not guide attention to the target need to be measurably distinct in their event indices (Zwaan & Radvansky, 1998). Such differences in event indices can be termed *differential elements* between the two event models (one created when watching the clip without context, and the other when watching the target film clip with context). Thus, differential elements are event indices that uniquely enable participants who know the scene's prior context to make predictions. The differential elements can be identified through a content analysis of the event model to determine which elements of the scene are important in association with the differential targets of attention.

I have identified six primary differential elements based on the event indexing model (Zwaan, Langston & Graesser, 1995; Zwaan, Magliano & Graesser, 1995; Zwaan & Radvansky, 1998) and the Causal Inference Maker model (van den Broek, 1990):

1. Entities are agents/objects/highly delimited locations of interest that are the designated targets of attention. For example, the stranger in the background using the slot machine in the *Ocean's Thirteen* clip. The correspondence between the target of attention and a specific entity (animate or inanimate) is very direct, and the existence of such a correspondence might be a necessary condition for my first criterion. However, it also seems likely that other differential elements can strengthen the impact of the event model on attentional selection.
2. Goals are the purposes or motivations of an agent, typically the target. However, these can also be goals of the non-target agents if they are connected to the target of attention. For example, the goal of Brad Pitt's character in the *Ocean's Thirteen* clip is to help others, like the stranger in the background, to win a fortune.

3. Temporal Operativity indicates whether the event leading up to the target prediction is still ongoing. For example, in the *Ocean's Thirteen* clip, the stranger is still operating the slot machine throughout the 10-second clip.
4. Spatial Proximity indicates the closeness of entities in the scene to other entities that can lead to targeted predictions. For example, in *Ocean's Thirteen*, the stranger is close to the slot machine to operate it.
5. Enabling Features are any feature (other than spatial proximity and temporal operativity) in the story that enables targeted predictions. For example, in the *Ocean's Thirteen* clip, the coin is a “special” one that always hits the jackpot.
6. Emotions indicate all emotion(s) of the agent(s) relevant for the target predictions. For example, the expression of gratification on Brad Pitt's face.

When multiple differential elements emphasize the importance of the target of attention, I predict that the viewer's event model can substantially influence their attentional selection toward the target. As discussed in the *Future Research Directions* section, this is a quantifiably testable hypothesis.

Low Default Gaze Probability Criterion (Criteria 2)

The target of attention should have a relatively low default gaze probability based on stimulus features (visual saliency), so it will not reflexively pull the viewer's attention to itself, despite having a strong event model. The default gaze probability is determined by both low-level factors (e.g., motion, color, luminance contrast, or being at the screen center) and some high-level visual stimulus factors (e.g., faces, text). I checked that several important visual factors were absent, to reduce the default probability of fixating the target in each film clip, by finding targets that either were not a person/animate entity (Fletcher-Watson, 2008; Humphrey &

Underwood, 2010; Langton et al., 2008), or lacked motion (Carmi & Itti, 2006; Mital et al., 2011; Theeuwes et al., 1999), or were not in the screen center (Dorr et al., 2010; Kümmerer et al., 2017; Tatler et al., 2011), or were not in focus (Enns & MacDonald, 2013; Peterson 2018).

Figure 5

Predicted Low Default Gaze Prediction for the different locations in the scene as measured by the DeepGaze IIE model.



Note. Low Default Gaze Prediction for the different locations in the scene as measured by the DeepGaze IIE model. The warmer, redder colors indicate a higher likelihood that viewers will fixate on that location, based on stimulus features. Note that the target of attention (yellow box) has less red shading, indicating potentially a lower probability than many other regions of the scene, especially Brad Pitt's face.

As discussed above, any of these factors can reliably attract the viewer's attention to the target. Not all of the above conditions must be satisfied to achieve the second criterion, so long as the film clip satisfies some of them. For example, the target of attention in the *Ocean's Thirteen* clip, the stranger at the slot machine, is not in the screen center, not in focus, and has only limited motion (see Figure 1 and Figure 5). Indeed, a film clip could fail any or all of the above checks yet be acceptable if the target of attention was one of many such entities, all of

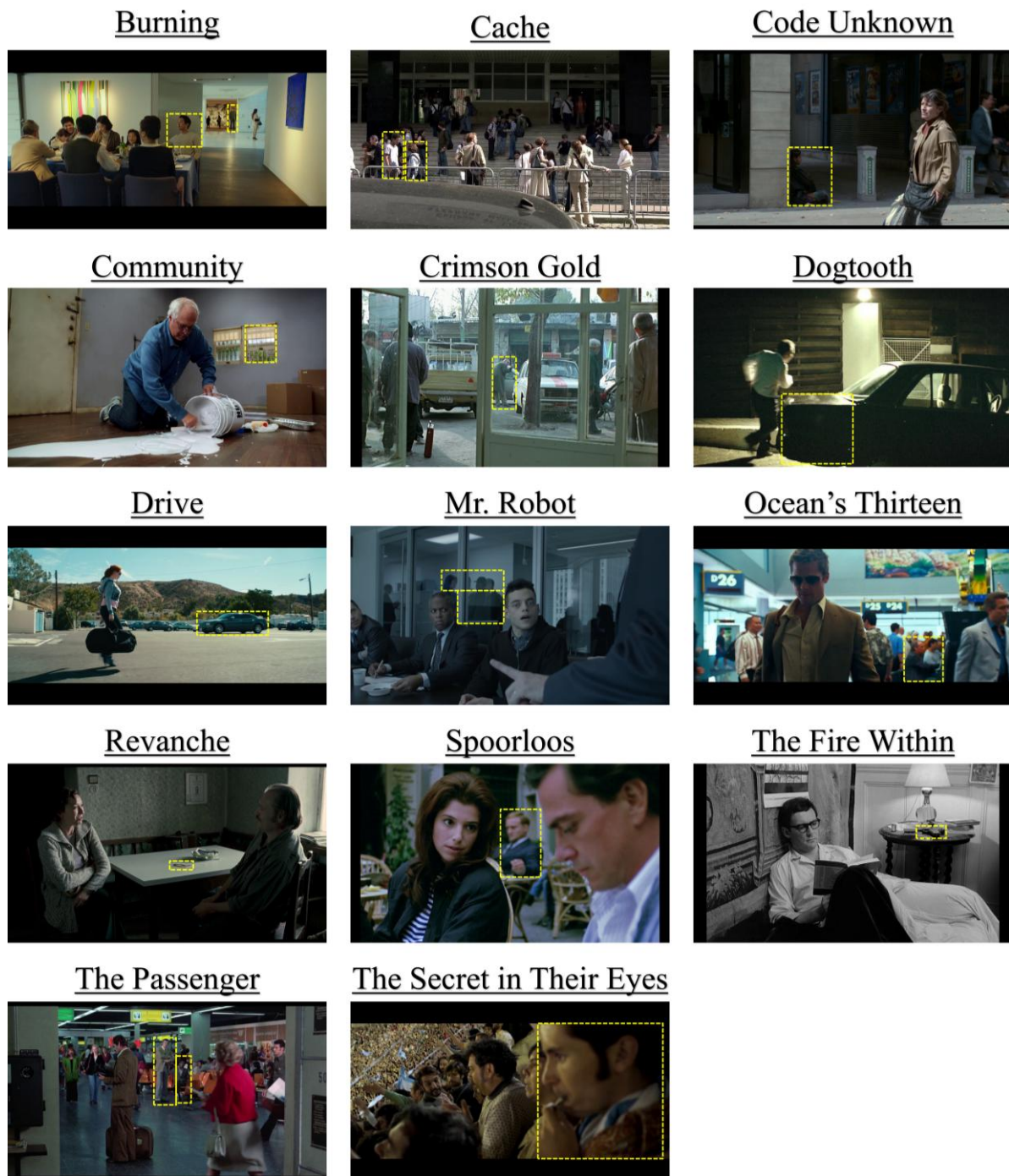
which similarly failed the same check. For example, in the *Cache* (Haneke, 2006) clip (see the middle image in the first row of Figure 6), the targets of attention are people who move, approach the screen center, and come into focus. Still, they are only two of roughly 100 such people in the scene who also fail the same checks (see Appendix A, *Cache* section for details). At this point, I do not know the relative weights of each of the above conditions. Hence, I have employed an objective statistical test to determine whether the target of attention has a relatively low default gaze probability. See Verification of Low Default Gaze Probability Criterion for details.

Due to the lack of context, the participant will have a weaker event model in the No-context condition. Hence, based on biased competition, I can expect stimulus-based features to be the primary determinants of attentional selection, creating a strong association between default gaze predictions and importance maps (see Figure 7). The default gaze probability map refers to regions of the scene that influence attentional selection purely on visual grounds. The importance map refers to the regions that influence attentional selection in the current event model. However, I do not want visual saliency to reflexively draw the viewer's attention to the target without the viewer knowing its significance in the event model. This was precisely what occurred in the *Touch of Evil* dynamic study (Hutson et al., 2017) and in other studies showing the Tyranny of Film (Huff et al., 2014; Loschky et al., 2015; Taya et al., 2012). For example, in the *Touch of Evil* dynamic study, viewers in the No-context condition, as described in the Introduction section, still looked equally at the car even when they were unaware of the bomb inside it, mainly because the car had a high default gaze probability. Hence, as stated in the Low Default Gaze Probability Criterion, the target of attention should have a relatively low default probability of fixation compared to the rest of the regions in the scene. As the *Ocean's Thirteen*

video clip (Figure 1) was selected using the *Low Default Gaze Probability Criterion*, the bottom-up features alone will be less likely to direct the viewer's attention to the target of attention.

Figure 6

Representative screenshots from all 14 films clips initially identified based on two criteria.



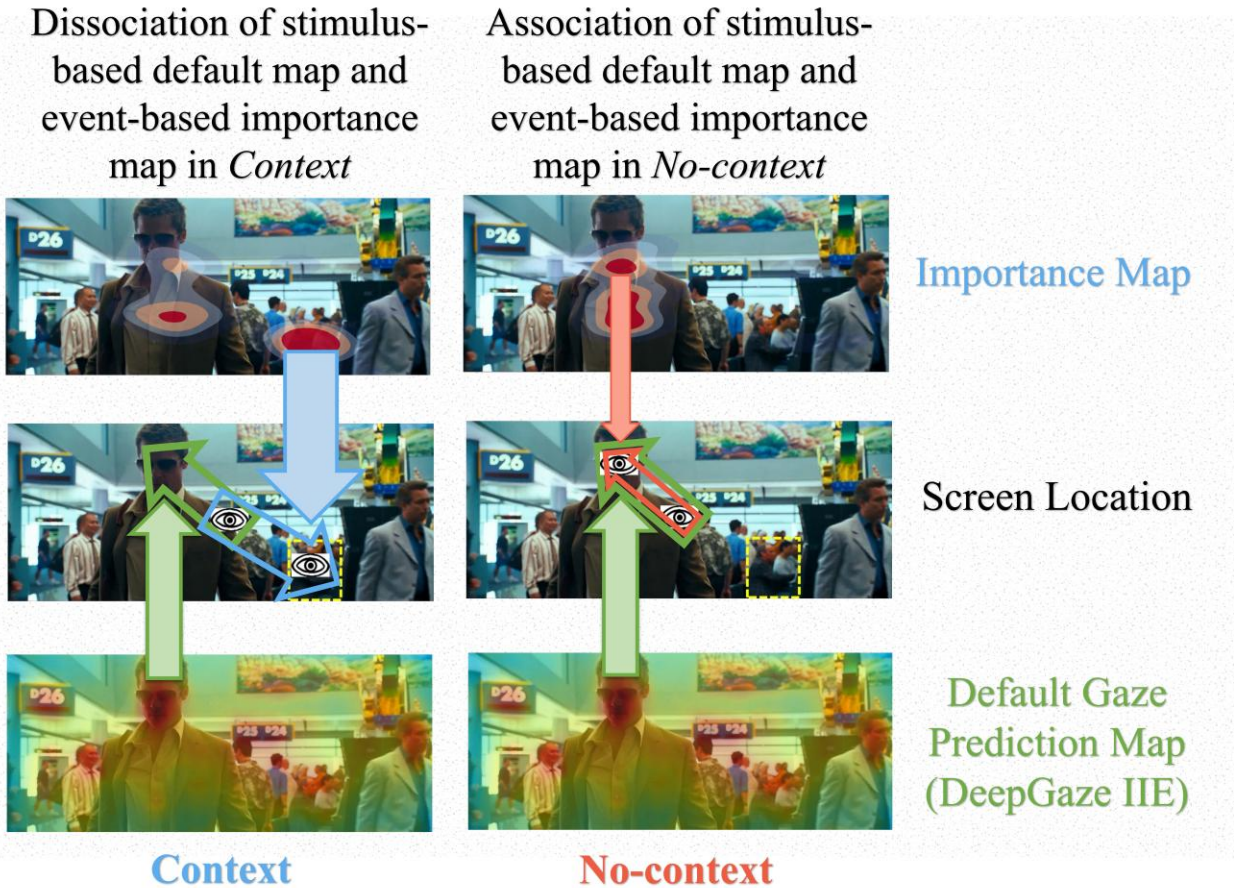
Note. Representative screenshots from the 14 films utilized in the study, arranged alphabetically from the top left. For *Crimson Gold*, *Spoorloos*, and *The Fire Within*, two versions of each clip were included to accommodate variations in the start and end points of the common viewing period, resulting in a total of 17 stimulus clips. Yellow bounding boxes indicate the predefined targets of attention (AOIs).

A state-of-the-art default gaze probability model (DeepGaze IIE) (Linardos et al., 2021) predicts that viewers are more likely to look at the main character's face or some of the bright signboards in the background than to look at the target stranger in the background (see Figure 5). On the Contrary, all the differential elements are predicted to direct viewers' attentional selection to the experimental target of attention in the Context Condition, which, if you go back to the example of *Ocean's Thirteen*, is the *stranger in the background*, interacting with the slot machine. In fact, an importance map was empirically generated from the density of mouse locations of 30 participants who were given the task of showing the important locations during the common viewing period based on their understanding (for more details, see the Preliminary Study 2 section). The importance map showed that only Context participants deemed the *stranger in the background* important, whereas No-context participants did not (see Importance Maps in Figure 7). This shows an example of big differences in event models regarding the target of attention, as stated in the Differential Event Model Criterion. When the two criteria are satisfied, the target of attention is not visually prominent enough to automatically attract the viewer's attention, yet it remains essential for understanding the story. This will create a dissociation between the default gaze prediction and importance maps. Hence, if the viewers in the Context condition look at the stranger in the *Ocean's Thirteen* example more than viewers in the No-context condition, it would imply that the event model has won the biased competition

through the dissociation between stimulus-based default gaze prediction and the viewer's event model (see Figure 7). Similarly, 17 video clips (including *Ocean's Thirteen*) (see Figure 6 & Appendix A for details on these videos and their targets of attention) were initially identified by me based on these two criteria. After verification processes (discussed below), I ultimately selected eight clips for my proposed study.

Figure 7

A diagram showing the stimulus-based default gaze prediction map for the Ocean's Thirteen clip and its event-based importance map is associated to direct viewer's current fixation to the same location in the No-context condition. While they are dissociated in the Context condition to direct viewer's current fixation to two different screen locations, there can be only one winner.



Note. The bottom row shows the Low Default Gaze Prediction heatmap generated using DeepGaze IIE, indicating locations with the highest likelihood of fixating based on stimulus features, which should be invariant across the No-context and Context conditions, given that the

common viewing period is the same. The top row shows the importance map generated from the density of 30 participants' mouse movements in Preliminary Study 2, providing empirical predictions for locations prioritized by the event model. In both heat maps, warmer, redder regions are predicted to have the highest likelihood of current fixation. The central row represents the image on the screen. For simplicity, the previous fixations are assumed to be at the center of the screen, represented by the eye icon. The green, red, and blue arrows indicate the direction and destination of the shift in attention from the previous central fixation predicted by stimulus-based features, the No-context event model, and the Context event model, respectively. The second eye icon indicates my predicted current fixation location when both criteria are satisfied.

Study Plan

The proposed study aims to systematically provide an existence proof (for example⁴, Jones et al., 2018; Larocque et al., 2001) of the effect of the viewer's event model on their attentional selection while watching commercial film clips. I aim to do so by first establishing a set of film clips that satisfy both proposed criteria. Here are the two steps I followed to verify that:

1) Verification of the Differential Event Model Criterion

I employed a two-step verification process to ensure the selected videos satisfy the differential event model criterion.

- i. In the first step, I wanted to establish that the context manipulation introduced by my selected videos can lead to differences in event understanding, as predicted. I

⁴ These are couple of examples in Cognitive Psychology that have employed existence proof in their area of research to show evidence for existence for an effect that previous studies failed to show.

confirmed this in my *Preliminary Study 1*, namely, the *Prediction and comprehension task*, and the corresponding results are shared in the Preliminary Study 1 section.

- ii. In the second step, I investigated whether these differences in event understanding also affect viewers' moment-to-moment judgments of what is essential in the scene. I confirmed that in *Preliminary Study 2*, namely, the *Importance Judgment Task*, and the corresponding results are shared in the Preliminary Study 2 section.

2) Verification of Low Default Gaze Probability Criterion

In this verification process, I established that the targets of attention during the common viewing period of my selected videos have relatively low default gaze probability compared with the majority of screen regions. I confirmed this in the *DeepGaze IIE Study*, and the corresponding results are shared in the DeepGaze IIE study section.

3) Eye-Tracking Study

During these verification processes, any videos identified as failing any of the proposed criteria, as defined by thresholds, were removed from future analyses. Based on these verification processes, I selected 8 videos and their 10 targets of attention that met both proposed criteria. I then ran the proposed *Eye-Tracking Study* using the selected 8 videos to capture differences in viewers' overt attentional selection. In this study, I investigated whether these differences in event understanding would also lead to differences in attentional selection by tracking participants' eye movements while they watched these videos.

Chapter 3 - Preliminary Studies and Analyses to Verify the Two Criteria

Before checking whether viewers' attention to the target of attention varies with their understanding of the film clips, I first need to establish that the selected videos indeed satisfy the two criteria.

Verification of the Event Model Criterion

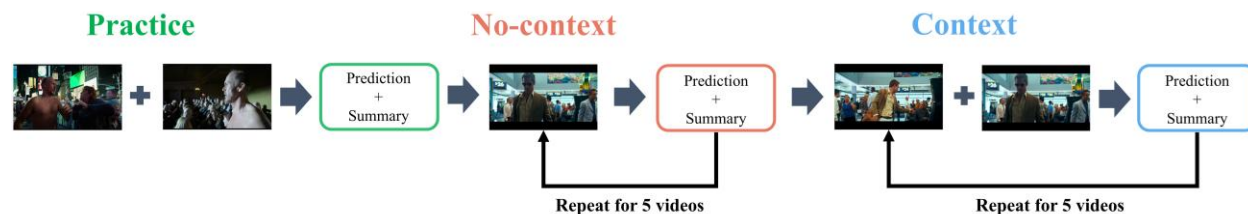
It is important to note that we can measure differences in viewers' event models in various ways. Hence, I conducted two preliminary studies to ensure that the selected videos met the event model criterion and to eliminate those that did not. Preliminary Study 1 (Predictive Inference Task) used linguistic measures of differences in viewers' event models, based on written responses to a prediction task that I coded for mentions of the target event indices. In Preliminary Study 2 (Importance Judgment Task), I used qualitatively different spatio-temporal measures of viewers' event models. I gave viewers the task of hovering their mouse over the most important elements in the video clip at each moment and clicking their mouse for particularly important moments. This method allowed me to bridge differences in viewers' event models to specific times and places in each film clip, which I can later test for differences in attentional selection using separate viewers' eye movements.

Preliminary Study 1 (Prediction and Comprehension Task)

I ran a preliminary study to determine whether videos created according to the two criteria above could lead to differences in viewers' understanding. I addressed the following research questions in Preliminary Study 1. RQ1) Do these context manipulations create differences in predictions for each differential element? RQ2) How many videos were “effective” in creating differences in prediction?

Figure 8

Experimental design for Preliminary Experiment 1 (Prediction and Comprehension Task).



Note. The viewers first watch the No-context videos one by one, and after each video, they give their predictions and summaries. After watching the five videos in the No-context condition, they watched the same videos, which are now preceded by the context manipulation clip. Predictions and summaries are collected again after each clip.

Method. Eighty-three participants received course credit after giving informed consent ($N(\text{females}) = 56$). Seventeen pairs of silent video clips⁵ from commercial films were selected based on the two criteria. Each pair of video clips has 1) a *Common Viewing Period*, and 2) a context manipulation clip with video and intermittent written descriptions (see Appendix B for details on the creation of context manipulation clips). Common viewing period clips had an average duration of 46.35 s, a maximum of 136 s (2 m 16 s), and a minimum of 10 s. Context clips had an average duration of 338.65 sec (5 m, 38 s), a maximum of 620 s (10 m, 20 s), and a minimum of 46 s.

Procedure. The experiment was carried out online on Pavlovia. On Average, 20 participants watched each video pair. I presented all videos at 24 fps and used normalized size scaling to have the video stimulus size scale naturally with each participant's window size⁶.

⁵ The commercial films originally had audio but I removed that to control for any effect of audio on attention, as I wanted to only focus on the effect of visual features and processing on attentional selection in films.

⁶ In Preliminary Experiment 1, which was an online study, I were more concerned about having the videos presented fully within the participant's screen, and not worried about having a fixed stimulus size irrespective of the monitor/window. In Preliminary Experiment 2, and the Eye-tracking Experiment, I strictly control the stimulus size to fit the pre-specified monitor size in our laboratory.

Participants first watched the common viewing period of five film clips in the No-context condition. After each film clip, I asked participants to write 1) a prediction of what would happen next, and 2) a summary of what they saw. After that, they rewatched each of the five clips in the Context condition, with each context video clip followed by its *common viewing period* video clip that they had previously seen. Again, they answered the same questions (prediction and summary) after each clip (see Figure 8).

Analysis. Two independent coders were trained using 20 pilot responses from previous research assistants who were naïve to the study hypotheses. Training continued with iterative feedback until inter-rater agreement exceeded 80%, at which point the coders began scoring the experimental responses from Study 1. Coders independently evaluated participants' predictions of the presence of six differential elements using a standardized codebook (see Table 1 for details). Each element was initially scored using a binary system (1 = present, 0 = absent).

For clips containing multiple differential elements (e.g., Entity), individual binary scores were summed to create an aggregate category score. For instance, if a stimulus clip contained two distinct target entities, a participant mentioning both would receive a total score of 2, whereas mentioning only one or neither would result in scores of 1 or 0, respectively. A critical exclusion criterion was also applied to all categories: an element was only scored as present if the prediction was grounded in the information provided during the common viewing period. Even if an element (such as an Entity) met the primary coding criteria, it was not rewarded a point if the prediction clearly ignored the common viewing period and relied solely on broader video context. This ensured that the scores reflected participants' integration of the specific events in the common viewing period rather than general narrative context knowledge, which is especially crucial in the Context condition.

Table 1

Coding Rulebook for all the differential elements with examples from the Ocean's Thirteen clip.

Differential Elements	Definition/Criteria	Coding Rules	Examples
Entity	Person/Object/Place present in the prediction	1) Is the entity (Person/Object) mentioned using the exact terminology? 2) If not, is the entity (Person/Object) mentioned by a different name, but it is evident from the descriptors or their actions? 3) If not, is there any direct action or event mentioned that directly implies the person/object (i.e., no other person/object could perform or be involved in that action or event)?	The entity is "Stranger at the slot machine": 1) A prediction of "Stranger at the slot machine hits the jackpot" receives 1 point for using the exact terminology. 2) A prediction of "Old man who received the coin hits the jackpot" receives 1 point for using a different name, but based on the descriptions and actions, it is clear that the reference is to the stranger at the slot machine. 3) A prediction of "People gather around the slot machine to congratulate the jackpot win" receives 1 point for mentioning the event that directly implies the stranger at the slot machine, as no one else could be involved in that event.
Goal	Goals of the agent included in the prediction.	1) Are any of the actions described in the predictive inference a clear implementation of any character's predicted differential goals? 2) OR, are any of the outcomes described in the predictive inference a clear success, failure, or continuation of any character's predicted differential goals? 3) OR, are there new goals mentioned in the predictive inference? If yes, is that new goal a clear update to any character's predicted differential goals?	The goal of Brad Pitt's character is "helping the poor have a chance at winning a fortune": 1) A prediction of "Brad Pitt will help more poor people win the jackpot" receives 1 point for mentioning an action that clearly implements the goal. 2) A prediction of "The coin Brad Pitt gave helped the poor old man win a fortune" receives 1 point for mentioning an outcome that clearly achieves Brad Pitt's differential goal. 3) A prediction of "Brad Pitt will find a different way to help the poor" receives 1 point for mentioning a new goal that clearly updates Brad Pitt's previous goal.
Temporal Operativity	An antecedent event that leads to the prediction and is ongoing during the Common Viewing Period	1) Is the temporally operative antecedent event described directly by the participant in their prediction, OR as something leading to the prediction? 2) If not, is the prediction a continuation or an ending of the antecedent event?	The temporally operative event is "the stranger putting the coin and playing the slot machine": 1) A prediction of "He puts the coin in the slot machine and wins big" receives 1 point for mentioning the temporally operative event as the leading incident in the prediction. 2) A prediction of "He wins the jackpot at the slot machine" receives 1 point for mentioning the end of the antecedent event.

Spatial Proximity	How the closeness of different entities in the scene leads to the prediction	1) Is the spatial proximity explicitly cited (i.e., using phrases like “because” or “due to”) as a reason for their prediction? 2) Is the influence of spatial proximity implied indirectly but clearly in their prediction? 3) If not, is the prediction about an increase or a decrease in proximity?	Spatial proximity is "the distance between Brad Pitt and the Stranger in the background being significant,": 1) A prediction of "The Stranger will look for Brad Pitt after hitting the jackpot because he is now further away" earns 1 point for directly citing spatial proximity as the reason for the prediction. 2) A prediction of "The Stranger looks for Brad Pitt after hitting the jackpot but can't find him" earns 1 point for indirectly implying that Brad Pitt is far enough from the Stranger for him not to be seen. 3) A prediction of "Brad Pitt will continue walking away from the Stranger and the jackpot machine to avoid any suspicion" earns 1 point for indicating a decrease in proximity.
Enabling Features	A feature in the story that enables prediction is not included in the temporal operativity or spatial proximity.	1) Is the enabling feature directly or explicitly cited (i.e., using phrases like “because” or “due to”) as a reason for their prediction? 2) Is the enabling feature’s influence indirectly but clearly implied in their prediction?	The enabling feature is "The coin is a special coin designed to hit the jackpot": 1) A prediction of "The Stranger wins the jackpot because the coin is designed to rig the slot machine" receives 1 point for directly citing the coin's "specialty" as a reason for the prediction. 2) A prediction of "As soon as he uses the coin in the slot machine, he will win the jackpot" receives 1 point for implying that the coin's "specialty" ensures he will certainly hit the jackpot.
Emotions	Emotion(s) of the agent(s) relevant for the prediction.	1) Are the agent's emotions relevant to the prediction mentioned using the exact words? 2) If not, are the agent's emotions mentioned using a synonym for the same emotion? 3) If not, are there any direct actions by the agents in the predictions that are a direct result of those emotions?	Emotion is "staying cool under pressure": 1) A prediction of "Brad Pitt stays cool under pressure and walks away unnoticed" receives 1 point for using the exact words to describe his differential emotion. 2) A prediction of "Brad Pitt stays calm, shows no outward expressions, and leaves the airport unnoticed" receives 1 point for using a synonym of the same emotion. 3) A prediction of "Brad Pitt casually walks out of the airport without drawing any suspicion" receives 1 point for mentioning an action that is a direct result of his differential emotion.

For the Entity category, a score was assigned if a target was mentioned by its exact name, referred to via clear descriptors, or implied through unique actions or events that only that specific entity could perform. Goal presence was identified (assigned 1) if the prediction included actions implementing a character’s goal, outcomes representing the success or failure of

that goal, or clear updates to previous objectives; otherwise, that goal is considered absent (0).

Temporal Operativity was coded if the participant explicitly cited a temporally operative antecedent event or if the prediction itself represented a direct continuation or conclusion of that event.

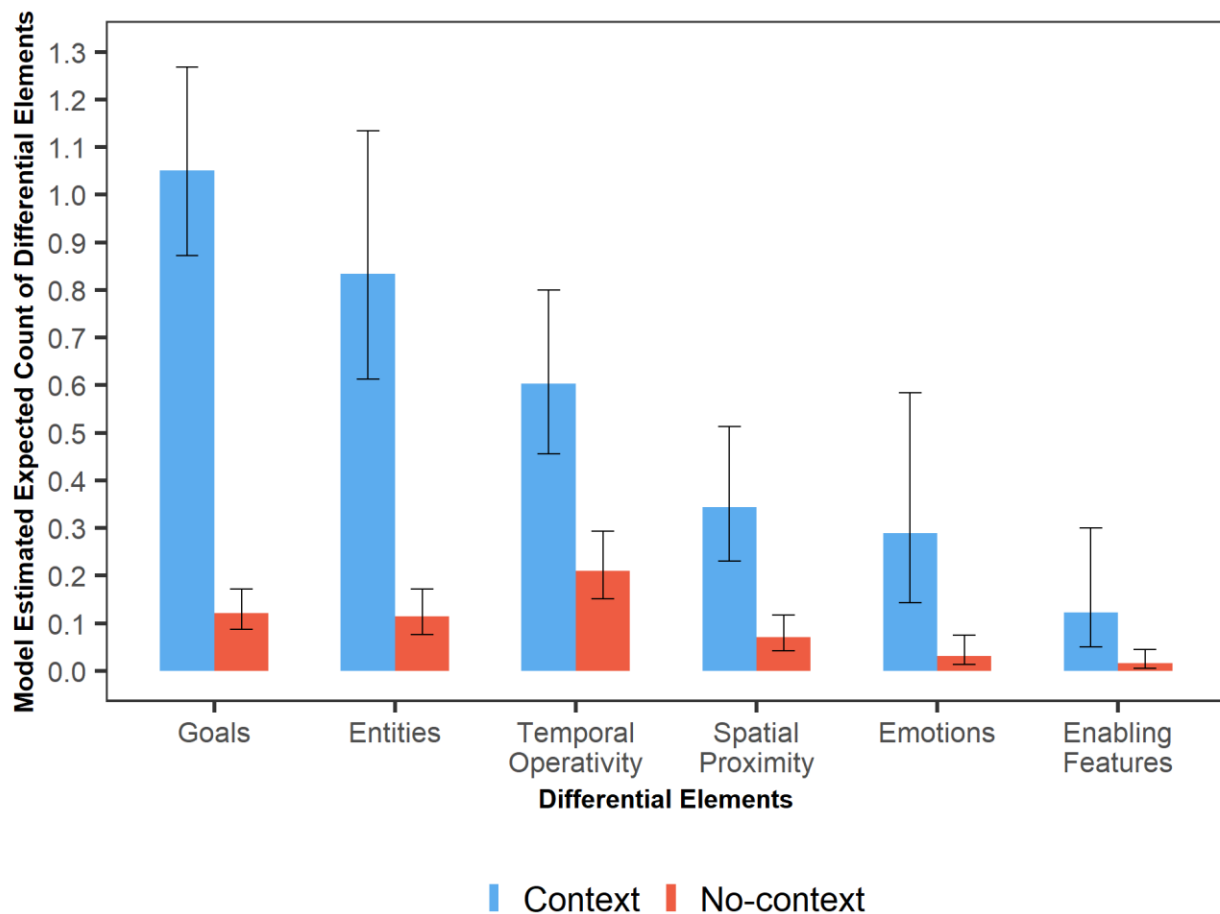
Criteria for Spatial Proximity required either an explicit citation of proximity as a causal factor or an unambiguous implication of its influence, including predictions involving a change in physical proximity. Similarly, Enabling Features were scored if they were explicitly cited as a reason for a prediction or if their influence was clearly implied within their predictive inference. Finally, Emotions were recorded if a specific emotion was accurately attributed to the correct agent using exact terminology or synonyms, or if the participant described an action resulting directly from that emotional state.

To quantify coding consistency, I assessed inter-rater reliability using Cohen's Kappa (Cohen, 1960) in the first round of coding. The Cohen's Kappa varied for each differential element ranging from good agreement (.40 -.60) to excellent agreement (>.75) (Fleiss et al., 2003) as reported here in the increasing order of Cohen's K : Temporal Operativity ($K = .44$; 89% agreement), Goal ($K = .49$; 73% agreement), Enabling Features ($K = .60$; 94% agreement), Spatial Proximity ($K = .66$; 90% agreement), Emotions ($K = .68$; 89% agreement), and Entity ($K = .83$; 92% agreement). In case of differential elements, where Cohen's Kappa is relatively low, I have reported raw agreement to show that Kappa may be underestimating reliability despite high raw agreement due to the low frequency (rarity) of those observed scores (Chicchetti & Feinstein, 1990). For example, in Temporal operativity, Cohen's Kappa is at .44, very close to the poor-agreement cutoff (<.40); however, the inter-rater agreement was 89%, exceeding the conventional 80% threshold recommended for behavioral research (Hartmann, 1977). Inter-

rater discrepancies identified during the initial coding phase were resolved through a targeted second round of coding. During this stage, coders used participants' clip summaries to provide the necessary context to disambiguate vague predictions (e.g., clarifying ambiguous pronoun references such as 'he' or 'she'). Any remaining disagreements following the second round were adjudicated through consensus meetings moderated by me. In these sessions, coders compared their respective justifications and collaborated to reach a final agreed-upon score for each instance. Final inter-rater agreement reached 100% following the moderated adjudication phase.

Figure 9

The Average number of each differential elements present in the predictions for both conditions.



Note. The Context condition has a significantly higher expected count for all the differential elements in the participant's prediction than the No-context condition for all the differential

elements. Error bars are equal to $\pm$ 95% Confidence Intervals. All values are back-transformed from the original model's log scale.

Results. RQ1: I ran six individual multilevel Poisson regressions in which the condition (No-context vs Context) predicted the number of instances of each differential element, based on the coding data (see Appendix C for parameter estimates). I also controlled for random intercepts for each video and for random intercepts and slopes that varied by condition and participant. Context manipulations led participants to identify significantly more differential elements in the Context condition than in the No-context condition (see Figure 9). None of the six models showed significant overdispersion⁷. Thus, viewers used all six differential elements to make predictions based on their event model.

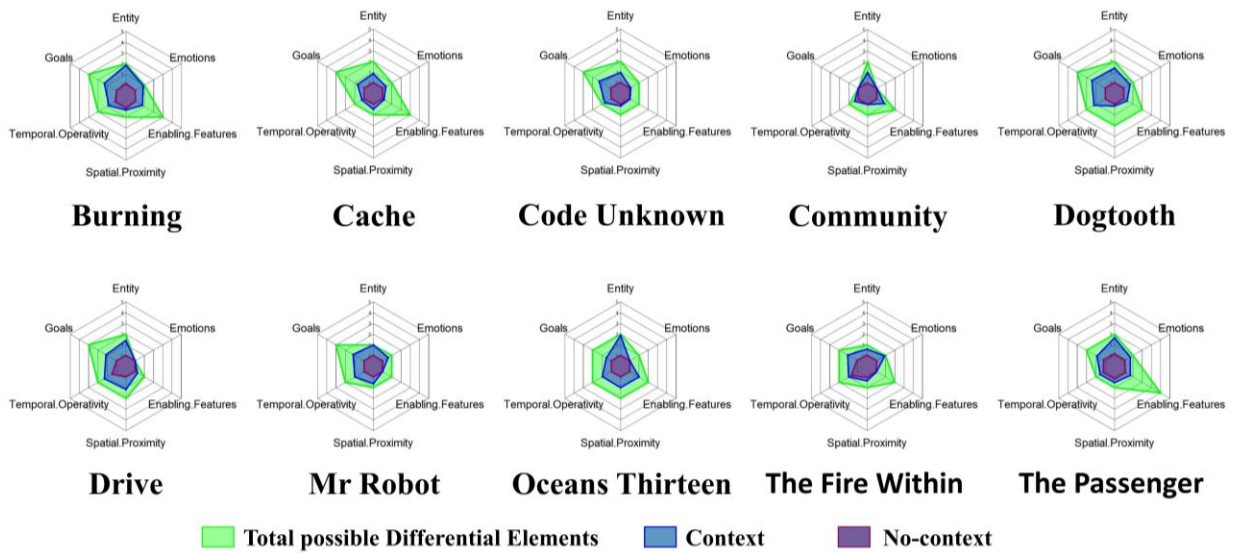
RQ2: I identified which video pairs most effectively induced differential event models among viewers. For that, I calculated *the total increase in differential elements* for each participant (the sum of the increases in the number of each differential element from their No-context prediction to their Context prediction). I then ran a linear mixed-effects model to predict the total increase in differential elements, treating Video now as a fixed effect and accounting for participant-level variability via a random intercept. To ensure robust stimulus selection, I applied a conservative statistical criterion: I selected only those videos for which the lower bound of the 95% Confidence Interval (CI) for the marginal mean was greater than 1. This threshold ensures that, after accounting for random variance, I can be 97.5% confident that the average participant experienced a total increase of at least one differential element in that video. Based on this

⁷ To check for overdispersion (see Appendix 3 for dispersion ratio calculated using *performance* (Lüdtke et al., 2019) package in R) I plotted the frequency distribution standardized deviance residuals, in which I found that there were very few residuals beyond $-2/+2$ interval for all the poisson regression models. Similarly, overdispersion was also checked based on computing the square root of the penalized residual sum of squares divided by n , the number of observations (scaled parameter) where all of the computed values were between 0.75 and 1.4, which were the recommended values (Blmeico, 2022).

criterion, 10 of 17 videos were retained for subsequent analyses (see Appendix D). As illustrated in Figure 10, the Context event models for these 10 videos were substantially richer (i.e., with a higher number of differential elements) than their No-context counterparts, spanning multiple differential elements (see Appendix E for radar charts of the excluded videos).

Figure 10

Radar charts depicting the differences in the average number of differential elements represented in viewers' predictions in Context and No-context conditions for the 10 videos selected from Preliminary Experiment 1 for future analyses.



Note. Each axis of the hexagon in the radar plot represents a differential element. Each hexagon ring represents the count of differential elements with an increment of 1, where the innermost ring represents the value 0 for all differential elements and the outer ring represents the count of 5. These plots show how participants' event models differed across these differential elements in the Context (blue region) and No-context (maroon region) conditions. The green area represents the total possible differential elements the participant could have included in their predictions, highlighting the maximum values they could have obtained for each differential element in each video. The area covered by the Context conditions for all ten videos is always greater than that

covered by the No-context conditions, indicating the extent of richness in their context event models across all six differential elements.

Preliminary Study 2 (Importance Judgment Task)

When watching a movie, the viewer's current event model determines what and where the most critical information is on the screen from moment to moment. Event indices in the viewer's current event model in working memory are highly prioritized in comprehension and memory (Radvansky & Zacks, 2014; Zwaan & Radvansky, 1998). These event indices could direct attentional selection to the relevant spatial locations on the scene represented by the target of attention. This is because the Biased Competition Theory (Adeli & Zelinsky, 2018; Desimone & Duncan, 1995; Tsotsos et al., 1995) predicts that the perceptual representations of these indices require sustained, prioritized activation to be retained in working memory (Loschky et al., 2025). Importance maps from event models may help me empirically identify the spatial regions prioritized for attentional selection, namely the priority map. Preliminary Study 2 investigates whether these differences in event comprehension also change viewers' judgments of the most *important* piece of information at each moment in the *common viewing period*.

Research Question. Do differences in event comprehension change what viewers judge as important times and locations in a film narrative?

Event Model Importance Hypothesis. The viewer's event model specifies the meaningful/relevant information in the scene; hence, differences in the viewer's event models prioritize different spatial locations as important, leading participants to create empirically generated importance maps using their mouse movements. In the Context condition, participants will consider the targets of attention to be among the most important information on the screen.

In comparison, they will fail to identify the target of attention as important in the No-context condition.

Null Hypothesis. The differences in event models will not change the importance of any information in the scene. Namely, there will be no significant differences between the important information judged in the No-context condition and the Context condition.

Stimuli. 10 pairs of silent video clips were selected using the results of Preliminary Study 1 for this study. For these 10 pairs, Common Viewing Period clips had an average duration of 44.7s, a maximum of 136s (2 m, 16 s), and a minimum of 10s. Context clips had an average duration of 319.9 s (5 m, 19.9 s), a maximum of 620 sec (10 m, 20 s), and a minimum of 46 s. These commercial films were shot in various aspect ratios ranging from 1.66:1 to 2.39:1 (see Appendix F for detailed video specifications). As the videos will be shown on a screen with a 16:9 aspect ratio, all videos were converted to a “letterbox” format by adding black borders above, below, or on the sides, depending on the original aspect ratio. This is done to avoid altering the original aspect ratio, ensuring that the pixels are not stretched and that the physical width and height remain the same. Hence, all the video files were presented in the MP4 format and had a resolution of 1920 x 1080 pixels.

Method. The procedure is similar to Preliminary Study 1 (see Figure 11). The only difference was that after the common viewing period (either in the No-context or Context condition), I asked the participant to rewatch only the common viewing period and complete the Importance Judgment Task. I presented a practice video at the beginning of the experiment to familiarize the participant with the task. Using the 10 pairs of silent video clips selected based on the Event Model Criterion and the results of the predictive inferences in Preliminary Study 1, Study 2 measured spatiotemporal differences in viewers' event models for the common viewing

period in the No-context and Context conditions. Participants ($N = 60$) repeated the same procedures as Study 1 (first viewing the No-context condition Common Viewing Period, then making a prediction; later watching the Context manipulation video + Common Viewing Period, followed by answering prediction questions again). However, after making their predictions, viewers were asked to watch the common viewing period again and use their mouse (visible on-screen as a four-sided arrow) to 1) track the most important person/place/thing at each moment, and 2) click at particularly important moments. I advised them to follow/stay on that information as long as it remained important, and to switch to new information as it became important. After each video clip was in the Context condition, I also asked participants *whether they had seen the movie at any time before this experiment*.

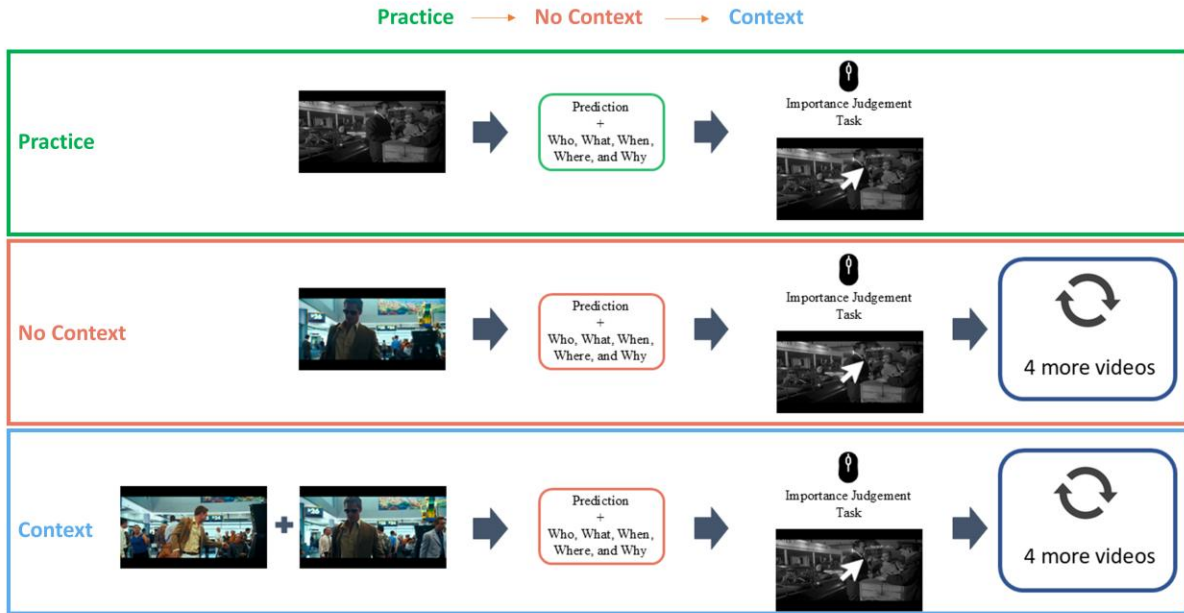
Analyses. I removed participants' responses from video clips in which they reported having seen the movie before this experiment. However, I still included their responses from the other video clips in the analyses. I removed a total of 7 responses: 3 from the *Community* video, 2 from *Dogtooth*, and 1 each from *Ocean's Thirteen* and *Mr. Robot*.

Target Areas of Interests Analyses. Dynamic rectangular Areas of Interest (AOIs) were created around the targets of attention to selectively analyze mouse movements to those AOIs in the Importance Judgment Task. I generated them semi-automatically using the free online software Computer Vision Annotation Tool (CVAT) (<https://github.com/cvat-ai/cvat.git>) (Sekachev et al., 2020), a commonly used interactive video and image annotation tool in computer vision (Mary T_Intel, 2019). I created rectangular bounding boxes around the target of attention for each video, which were propagated across subsequent frames for relatively static AOIs or were automatically tracked using CVAT's Deep Neural Networks (DNNs). I then inspected these regions of interest for each frame. I manually corrected them to ensure the

correct targets of attention were annotated and that they were entirely enclosed within the rectangular boxes. After this annotation, CVAT produces an output with the x-y coordinates of the top left corner (xmin, ymin) and bottom right corner (xmax, ymax) of the Target AOIs across all frames in pixels.

Figure 11

Experimental Design for the Importance Judgment Task.



Note. Participants go through the Practice block first, then the No-Context block, and then the Context block. In each block, they watch a video and answer the prediction questions at the end. After the predictions, they are told to perform the Importance Judgment Task using their mouse cursor while the common viewing period clip is played back. They will go through five videos in each block.

Frame Order. Target AOIs may enter and exit the screen multiple times during the common viewing period, or may not be present at certain periods within it. To account for that, I will only consider the Frames with AOIs for any AOI-based analyses. Also, instead of using the standard Frame Index (Frame number in the order of appearance for each video), I will use

Frame Order, which is the frame number in the order of appearance while indexing only the frames where the AOI is present. For example, if in the common viewing period video the AOI is absent for the first 10 frames and only appears from the 11th frame, Frame Order will start numbering from the 11th frame as 1 and will not count any subsequent frames where the AOI is absent.

Results. As I have used participants' mouse locations to identify the importance of specific regions on the screen, the importance of a given Target AOI can be measured by analyzing the total time the mouse cursor spends within that AOI across video frames in which that AOI is present.

Proportion of Time Spent within AOIs. I measured the importance of the target AOIs by the time the participant spends using the mouse cursor within them. I used *total time* rather than a measure similar⁸ to the proportion of fixations on AOIs, which is commonly used in eye-tracking studies, because participants were instructed to hover and follow the most important person/place/thing with their mouse. In eye-tracking studies, saccades between fixations within an AOI are commonly ignored because minimal visual information is extracted during saccades due to saccadic suppression (Burr, Marrone & Ross, 1994; Dorr & Bex, 2013). However, in the Importance Judgment Task, the participant's mouse location indicates an important location on the screen, unless they quickly shift between locations to signal a shift in importance. I could also have differentiated between this quick mouse movement during a shift in importance and mouse movement when following a moving object on the screen by identifying a threshold

⁸ Based on other studies, a stay/stop is a period of time during which the user does not move the mouse or scroll. A stay can be calculated using a minimum threshold of 220ms—if the stop is less than 220ms, then the mouse event will be considered as the continuation of the previous mouse action (Milisavljevic et al., 2021).

mouse speed. However, the mouse importance judgment is a unique meta-cognition task, with few direct comparisons in the literature; hence, identifying and setting such a threshold will require verification. Therefore, I have not made that distinction in the current analyses, as it would represent only a small fraction of the entire dataset.

I analyzed the proportion of time spent within Target AOIs by converting each participant's mouse location data (x, y coordinates) over time into a binomial variable indicating whether a miss or a hit occurred. A value of 0 (miss) was assigned if the cursor's location at that moment was outside the AOI, and a value of 1 (hit) was assigned if the cursor was within the AOI. After that, I ran a multilevel binomial logistic regression where mouse hit probability was the dependent variable. I was mainly interested in whether the Condition (Context vs No-context), *Frame Order* (as a proxy for time), and their interaction affected each participant's mouse hit probability. I initially included a maximal random-effect structure in the model, as recommended by Barr et al. (2013); however, the maximal model failed to converge. Hence, to control for Type I error inflation while avoiding the convergence and singularity issues frequently associated with maximal random-slopes models, I employed a Complex Random Intercept (CRI) structure (Scandola & Tidoni, 2024). The random-effects structure included a basic random intercept for the AOI grouping factor, as well as complex random intercepts representing the residual variance of AOI and Participant interacting with Condition (Context vs No-context). Finally, Frame Order was included as a random slope for each grouping term (see Appendix G for the detailed model selection process). Now, using this random-effect structure, I also ran a nested model comparison to determine which fixed effects provided the best fit for predicting mouse hit probability. I found that the best model was a multilevel binomial logistic regression with just Condition (categorical: Context, No-context) predicting the probability of

mouse hits on each frame, which was also better than the Null model ($\chi^2(1, N = 60) = 17.14, p < .001$). Adding frame order as a predictor did not improve the model fit compared to the condition only model ($\chi^2(1, N = 60) = 0.56, p = .46$). Furthermore, adding an interaction between frame order and condition also did not improve the model ($\chi^2(1, N = 60) = 2.34, p = .13$) (see Appendix G for complete model comparisons). Hence, only the condition appears to have a substantial effect on the importance of the target AOI.

Table 2

Parameter Estimates for the Multilevel Binomial Logistic Regression Predicting AOI Hit Probability.

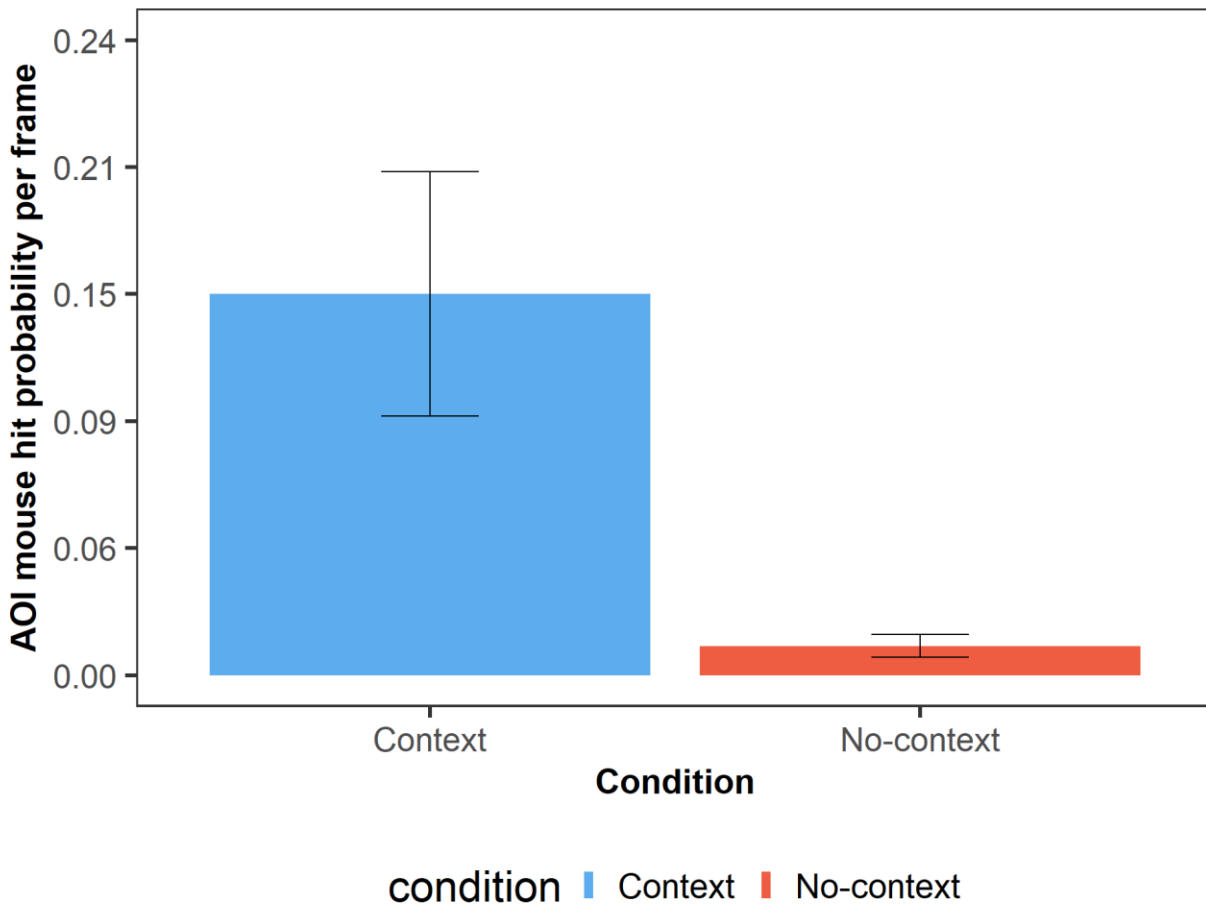
Variable	<i>B</i>	<i>SE(B)</i>	<i>Z</i>	<i>Significance(p)</i>	<i>OR</i>	95% CI for OR	
						Lower	Upper
Intercept	-3.09	0.30	-10.26	<.001	0.05		
Condition (Context)	1.35	0.23	5.78	<.001	3.89	2.46	6.17

Note: Condition was effect coded (No-context = -1; Context = 1).

The parameter estimates for the best model (i.e., the model with mouse hit probability predicted by Condition) are shown in Table 2. Based on the model, there was a significant main effect of condition predicting the probability of AOI mouse hits ($B = 1.36, z = 5.78, p < .001$). More importantly, participants were more likely to hit the target AOIs in the Context condition ($M = .150, SE = .048$) than in the No-context condition ($M = .011, SE = .004$) (see Figure 12). Thus, context manipulation influenced viewers' judgments of important screen locations. These results further confirm that my 10 video pairs meet the Differential Event Model criterion for strongly reducing the Tyranny of Film.

Figure 12

Estimated marginal means of the condition (Context vs No-context) from the multilevel binomial.



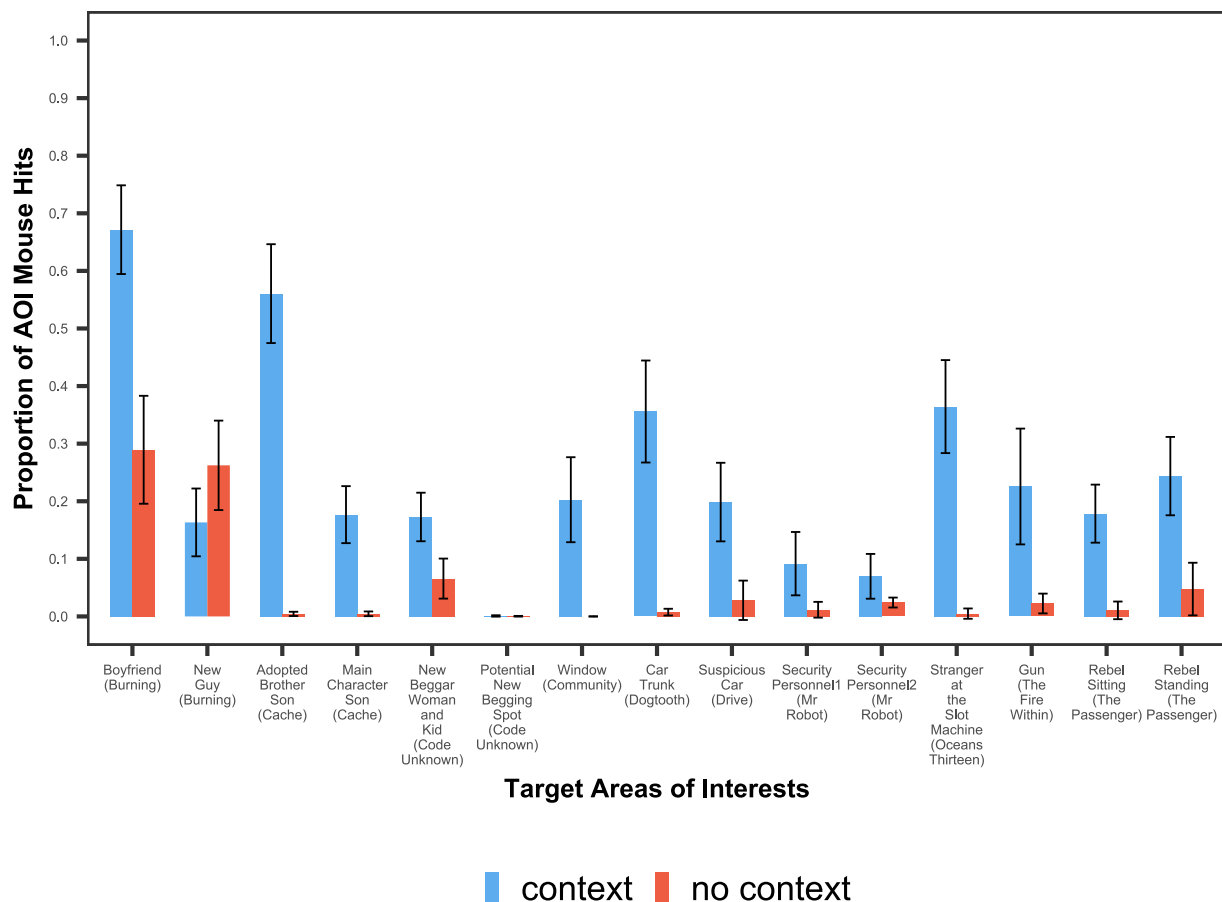
Note. Error bars are equal to +/- 95% Confidence Intervals. All values are back-transformed from the original model's logit scale.

Furthermore, to identify whether individual target AOIs in each video satisfied the event model criterion, I calculated each participant's overall proportion of AOI mouse hits for each AOI and condition. A target AOI was deemed to have satisfied the Event model criterion if the mean proportion of AOI mouse hits for the Context condition was significantly higher than the No-context condition across all 30 participants for that target AOI. I ran separate paired t-tests for each condition and target AOI to test this, setting the threshold for the AOI satisfying the Event Model Criterion to $t < 0$ (i.e., Context > No-context) and $p < .05$ (see Figure 13 & Table

3). Based on this test, 13 of 15 selected target AOIs from the 10 videos satisfied the event model criterion, while only 2 failed (see Table 3). I will not use the two failed targets of attention in the following eye-tracking study, as neither meets the first criterion.

Figure 13

Graph showing the proportion of Hits by the mouse for each Target Area of Interest, averaged across participants.



Note. The error bars are equal to +/- 95% Confidence Intervals. Those Target AOIs where the Context condition (*Blue* bars) is higher than the No-context condition (*Red* bars) are deemed to satisfy the Event Model criterion.

Table 3

Table showing paired t-test comparisons of the Proportion of AOI Mouse Hits between No-context and Context conditions for each Target AOI for 30 participants to test the Event Model criterion.

Video	Target Area of Interest	No-context		Context		df	t	p	Event Model Criterion Satisfied (t<0 & p<.05)
		M	SD	M	SD				
Burning	Boyfriend	0.29	0.26	0.67	0.22	29	-6.62	<.001	Yes
Burning	New Guy	0.26	0.22	0.16	0.16	29	2.18	.038	No
Cache	Adopted Brother Son	0.004	0.01	0.56	0.24	29	-12.60	<.001	Yes
Cache	Main Character Son	0.005	0.01	0.18	0.14	29	-6.70	<.001	Yes
Code Unknown	New Beggar Woman and Kid	0.07	0.10	0.17	0.12	29	-4.08	<.001	Yes
Code Unknown	Potential New Begging Spot	0.001	0.003	0.0003	0.002	29	-1.44	.161	No
Community	Window	0.00	0.00	0.20	0.20	29	-5.38	<.001	Yes
Dogtooth	Car Trunk	0.01	0.02	0.36	0.24	27	-7.67	<.001	Yes
Drive	Suspicious Car	0.03	0.10	0.20	0.19	29	-4.33	<.001	Yes
Mr Robot	Security Personnel1	0.01	0.04	0.09	0.15	28	-2.74	.011	Yes
Mr Robot	Security Personnel2	0.02	0.02	0.07	0.11	28	-2.19	.037	Yes
Oceans Thirteen	Stranger at the Slot Machine	0.005	0.02	0.36	0.22	28	-8.66	<.001	Yes
The Fire Within	Gun	0.02	0.05	0.23	0.28	29	-3.86	<.001	Yes
The Passenger	Rebel Sitting	0.01	0.04	0.18	0.14	29	-6.46	<.001	Yes
The Passenger	Rebel Standing	0.05	0.13	0.24	0.19	29	-4.59	<.001	Yes

Note: The p-values in **bold** are statistically significant with $p < .05$. The threshold for the AOI satisfying the Event Model Criterion is set to be both $t < 0$ (i.e., Context > No-context) and $p < .05$. The shaded cells are those that failed the Event Model criterion.

Verification of Low Default Gaze Probability Criterion

To test the Low Default Gaze Probability Criterion, I analyzed the Common Viewing Period video clips using the state-of-the-art DeepGaze IIE saliency algorithm (Linardos et al., 2021). This has allowed us to determine a priori whether the targets of attention have relatively low predicted gaze probability densities. Deepgaze IIE saliency maps will also be 2-D probability distributions indicating how likely participants are to look at a location at any given time, based on stimulus features and deep learning of eye movement data (similar to the one shown in Figure 5). Saliency maps are the standard terminology in computer vision; however, I will refer to them as Default Gaze Prediction Maps going forward for the reasons cited earlier.

DeepGaze IIE Study

The DeepGaze IIE model predicts a fixation density $P(X = (x,y)|I)$ over possible fixation location $X = (x,y)$, given a frame I . Hence, when you run the DeepGaze IIE model on a given video, it produces $\log(P(X|I))$ for every pixel location, X , represented by (x,y) coordinates in the I^{th} frame, and you will get similar values for all the frames in the video. I converted all output values for each pixel in each frame to the exponential of the DeepGaze IIE output to obtain the exact probability of fixating at each pixel (i.e., $P(X|I)$)⁹. The probability of fixating on a rectangular target AOI, given the $(x_{\min}, y_{\min})$ is the top left corner and $(x_{\max}, y_{\max})$ is the bottom right corner, should be the probability of fixating in any of the pixels within the bounds of this AOI. In other words, it should be the summed probability of pixel $(x_{\min}, y_{\min})$ or the adjacent pixel $(x_{\min}+1, y_{\min})$ or the next pixel $(x_{\min}+2, y_{\min})$, and so on, for all the pixels bound within the rectangular AOI.

$$P(AOI|I) = P(((x_{\min}, y_{\min}) \cup (x_{\min} + 1, y_{\min}) \cup \dots \cup (x_{\max} - 1, y_{\max}) \cup (x_{\max}, y_{\max}))|I)$$

When you consider the probability of the union of multiple events (A, B, C , etc.) given another event (D) in probability algebra, you use the formula: $P((A \cup B \cup C)|D) = [P(A|D) + P(B|D) + P(C|D)] - [P((A \cap B)|D) + P((A \cap C)|D) + P((B \cap C)|D)] + P((A \cap B \cap C)|D)$, where $\cup$ represents the union (or “or”) and $\cap$ represents the intersection (or “and”) between events; $P(A|D)$ means the probability of event A occurring given that event D has already happened. However, since I expect the center of the fixation to be only one pixel location at a given time, I would assume that the probability of fixating at any one pixel X_i is mutually exclusive from any other pixel X_j . If two events A and B are mutually exclusive then the intersection of those events,

⁹ Note that these probability values are small with average values on the order of 10^{-6} . This is expected since this indicates the probability of the fixation being centered at that specific pixel out of the total number of 2,073,600 pixels (i.e., 1920×1080).

$P(A \cap B) = 0$, hence, for the above equation I can assume all the intersection terms to be zero and

$$P((A \cup B \cup C)|D) = P(A|D) + P(B|D) + P(C|D) - [0 + 0 + 0] + 0 = P(A|D) + P(B|D) + P(C|D).$$

Hence, I can rewrite the equation for the probability of fixation within an AOI as follows:

$$P(AOI|I) = P((x_{min}, y_{min})|I) + P((x_{min} + 1, y_{min})|I) + \dots + P((x_{min} - 1, y_{min})|I) + P((x_{max}, y_{max})|I)$$

or

$$P(AOI|I) = \sum_{X=(x_{min}, y_{min})}^{(x_{max}, y_{max})} P((X)|I) \quad (1)$$

Hence, I calculated the default probability of fixating within an AOI for each Frame I by summing the individual probabilities of all pixels within that AOI. I also verified this calculation by summing the probabilities of all the pixels (i.e., 2,073,600) within each frame and confirmed that they all added up to 1. Hence, the current method I used to test the Low Default Gaze Probability criterion involves creating an AOI for each target of attention. Importantly, these are the same AOIs used in Preliminary Study 2. Then, using DeepGaze IIE, I calculated the overall default probability of gaze (i.e., the predicted probability of the participant's gaze landing within that AOI given a particular frame) of a given target of attention using the above-mentioned Equation 1.

Note that this second criterion is absolute, unlike the *Differential Event Model* criterion. Since the common viewing period is identical across the Context and No-context conditions, it cannot be determined by comparing the two. To test this criterion, I need to establish that the target of attention's default gaze probability does not stand out from the rest of the frame. I can estimate this by comparing the default probability value for fixating on the Target AOI to the random baseline distribution of probability values for AOIs of the same shape and size in that frame. To derive the random distribution, I placed the AOI rectangle at a random location on the

screen. I then calculated the resultant aggregate probability within the frame at that new location using the same DeepGaze IIE default gaze probability map (Equation 1). I then repeated that 1,000 times. The probability values from the 1,000 iterations would now represent an estimate of the random baseline distribution of the default gaze probability values for an AOI of the same size for that video frame. Now, I can do that for all the frames in each video.

Figure 14

Frame by Frame Default Gaze Probability Prediction for the Ocean's Thirteen common viewing period by DeepGaze IIE along with the baseline distribution measures.

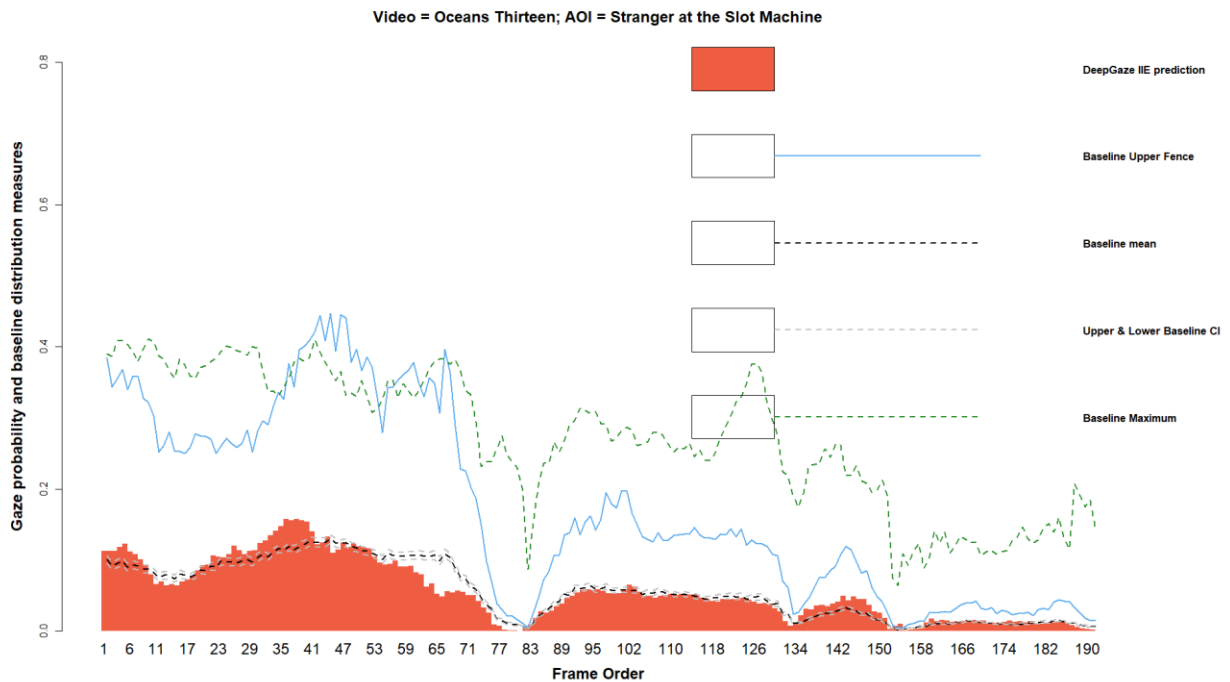


Figure 14 graphs the DeepGaze IIE predicted default gaze probability for the *Ocean's Thirteen* target of attention, the *Stranger at the Slot Machine*, and the corresponding random baseline distribution statistics for each frame of the common viewing period. This figure shows how the target AOI's default gaze probability compares to the baseline distribution for random AOIs of the same shape and size. If the corresponding target of attention AOI stands out from the rest of the distribution, then I could say that DeepGaze IIE predicts our target of attention to be conspicuous, and hence, a priori, fails the second Criterion. I tested this by identifying *the upper*

*fence*¹⁰ for outliers (3rd Quartile of the random distribution + 1.5*(Inter-Quartile Range)) (Tukey, 1977) and checking whether the default gaze probability of the target of attention was above the upper fence or not (i.e., whether it was an outlier or not). If it was clearly below the fence line, the target AOI's predicted gaze probability did not stand out from the rest and, hence, can be considered inconspicuous, satisfying criterion 2 for that frame. I can make the same comparison for all frames. If my target AOIs are below the upper fence (i.e., not outliers) throughout the majority of frames of the corresponding video, then I could accept those AOIs as having, a priori, satisfied the predicted *Low Default Gaze Probability* criterion. As shown in Figure 14, the AOI's predicted default gaze probability for this attention target is almost always well below the upper fence. You can see what happens for other videos in Appendix H. Although a frame-by-frame comparison of the default gaze prediction and upper fence could help predict moment-to-moment differences in eye movements, I need to look at their averages to identify which individual target AOIs in each video satisfy the Low Default Gaze Prediction Criteria.

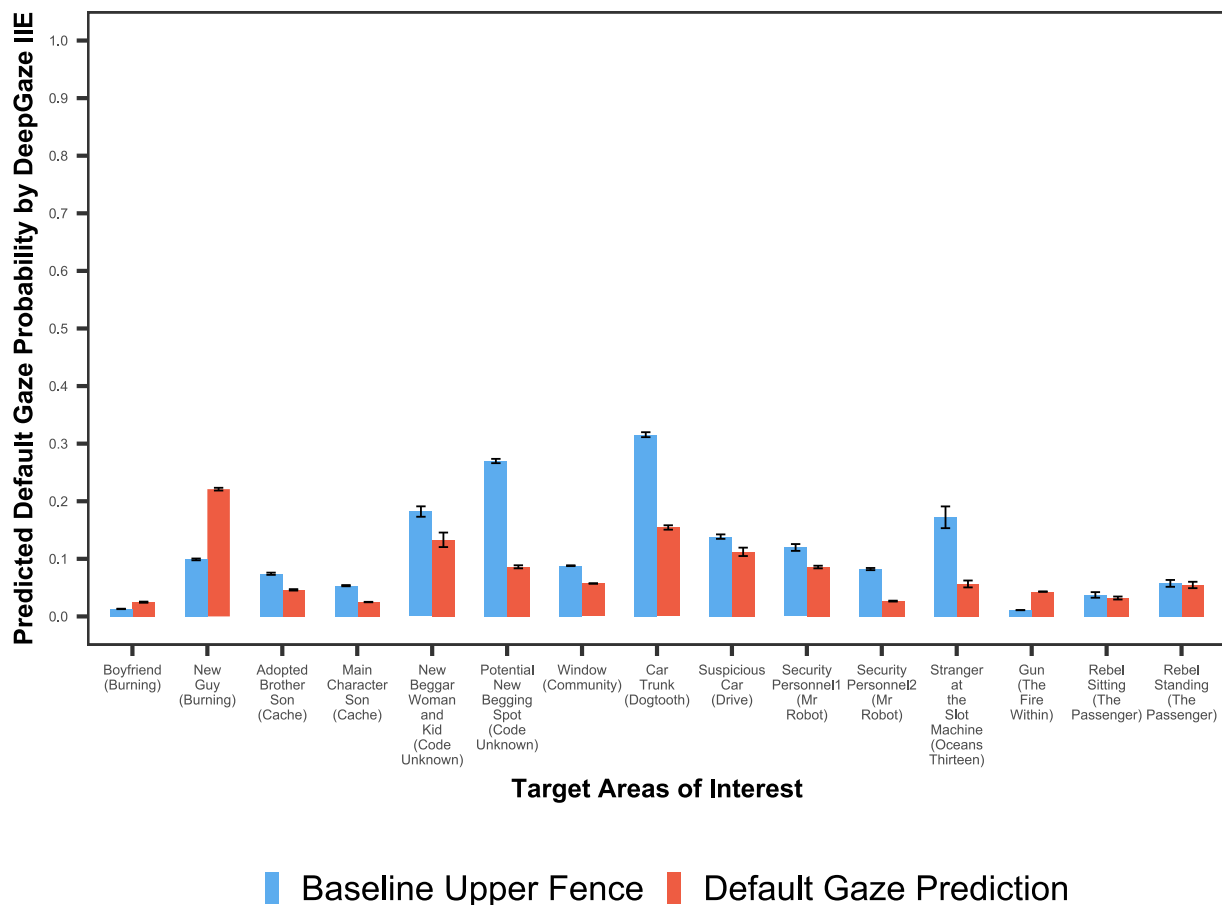
Figure 15 collapses the data in Figure 14 across all frames for each AOI, and shows the average predicted gaze probability in red and the upper fence in blue for each AOI across all 10 film clips. Here I can see that, taken as a whole, the *Stranger at the Slot Machine* AOI is indeed well below the upper fence, and would thus be predicted, a priori, to be inconspicuous, thereby meeting Criterion 2 for greatly diminishing the Tyranny of Film. The visual inspection of Figure 15 is supported by Table 4, which shows paired t-tests for each AOI and its corresponding upper fence across all 10 film clips. I set the threshold for the target AOI satisfying the *Low Default*

¹⁰ Normally, the probability is compared to the random baseline mean to determine whether the likelihood of the predicted event happening is better than chance. However, most of my Target AOIs are people or objects which the DeepGaze IIE is already designed to assign higher likelihood than the rest of the screen including an object recognition component (Linardos et al., 2021; Kümmerer & Bethge, 2023). Moreover, studies have shown that the probability of fixating on people in free-viewing of dynamic scenes is always greater than chance (e.g., Smith & Mital, 2013).

Gaze Prediction criterion to be $t < 0$ (i.e., Baseline Upper Fence > Default Gaze Prediction) and $p < .05$ (see Figure 15 & Table 4). Figure 15 and Table 4 show that most AOIs, a priori, meet Criterion 2. Nevertheless, this set of a priori default gaze prediction probabilities shows that two of the 10 film clips fail to meet Criterion 2: *Burning* (both AOIs failed) and *The Fire Within* (the single *gun* AOI failed). Additionally, one of the two AOIs in *The Passenger* failed (*the standing rebel*), while the other (*the sitting rebel*) managed to slip through. Based on these analyses, I must reduce my set of 10 film clips (13 Target AOIs) that previously passed Criterion 1 to 8 film clips (10 Target AOIs), which also passed Criterion 2.

Figure 15

Graph showing the Average predicted gaze probability of each AOI based on DeepGaze IIE estimation and baseline upper fence averaged across all the frame orders in that particular video.



Note. The error bars are equal to +/- 95% Confidence Intervals. The Target AOIs where the Default Gaze Prediction (*Red* bars) is lower than the Baseline Upper Fence (*Blue* bars) are deemed to have satisfied the a priori low default gaze prediction criterion.

Table 4

Paired t-test comparisons between the Default Gaze Prediction and the Baseline Upper Fence for each Target AOIs for all the Frames with AOIs present in the video to test the low default gaze probability criterion.

Video	Target Area of Interest	Default Gaze Prediction		Baseline Upper Fence		df	t	p	Default Gaze Criterion Satisfied (t<0 & p<.05)
		M	SD	M	SD				
Burning	Boyfriend	0.02	0.01	0.01	0.004	441	31.78	<.001	No
Burning	New Guy	0.22	0.03	0.10	0.02	458	99.23	<.001	No
Cache	Adopted Brother Son	0.05	0.03	0.07	0.04	1800	-60.82	<.001	Yes
Cache	Main Character Son	0.02	0.01	0.05	0.02	1818	-67.79	<.001	Yes
Code Unknown	New Beggar Woman and Kid	0.13	0.11	0.18	0.08	268	-6.01	<.001	Yes
Code Unknown	Potential New Begging Spot	0.09	0.04	0.27	0.05	807	-115.15	<.001	Yes
Community	Window	0.06	0.004	0.09	0.01	360	-89.42	<.001	Yes
Dogtooth	Car Trunk	0.15	0.03	0.32	0.03	206	-147.31	<.001	Yes
Drive	Suspicious Car	0.11	0.04	0.14	0.02	134	-7.84	<.001	Yes
Mr Robot	Security Personnel1	0.09	0.03	0.12	0.09	820	-15.87	<.001	Yes
Mr Robot	Security Personnel2	0.03	0.01	0.08	0.03	706	-44.11	<.001	Yes
Oceans Thirteen	Stranger at the Slot Machine	0.06	0.04	0.17	0.13	190	-16.73	<.001	Yes
The Fire Within	Gun	0.04	0.003	0.01	0.001	338	266.58	<.001	No
The Passenger	Rebel Sitting	0.03	0.01	0.04	0.03	125	-3.67	<.001	Yes
The Passenger	Rebel Standing	0.05	0.05	0.06	0.05	277	-1.07	.286	No

Note. The p-values in **bold** are statistically significant with $p < .05$. The threshold for the AOIs satisfying the Default Gaze Criterion is set to be $t < 0$ (i.e., Baseline Upper Fence > Default Gaze Prediction) and $p < .05$. Cells in bold are the ones that failed the criterion.

Thus, based on the three preliminary studies, two for Criterion 1 (the event model), and one for Criterion 2 (default gaze prediction), I was able to select eight videos that I predict will lead to a dissociation between the stimulus-based default gaze prediction and event-based importance map. This brings us to the key unanswered question I will investigate in this proposed dissertation study. Will the dissociation between the Default Gaze Prediction Maps

and the viewer's event model, as measured by both comparisons of differential elements (event indices) and Importance Maps, lead to differences in overt attention, namely eye movements?

Chapter 4 - Eye-Tracking Study

My results from Preliminary Study 1 show that 10 of my selected videos yield different predictive inferences in the Context and No-context conditions. Likewise, my results from Preliminary Study 2 show that, in the Context and No-context conditions, those 10 videos lead viewers to prioritize different screen information from moment to moment. The results of Preliminary Study 1 & 2 show that the 10 videos meet Criterion 1, creating differential understanding and prioritizing the target of attention during the common viewing period in the Context and No-context conditions. Finally, my results from Preliminary Study 3 using the state-of-the-art gaze prediction algorithm, DeepGaze IIE, show that 8 of the 10 videos meet Criterion 2, with a low default probability that the targets of attention are fixated. Hence, in the final Eye-Tracking Study, I aim to test whether these differences in event comprehension - specifically how the current event model in the working memory prioritizes differential information - influence the overt attentional selection of targets. I can measure overt attentional selection using eye-tracking.

Attentional Selection Hypotheses

Event Model Hypothesis

The viewer's eye movements (i.e., attentional selection) will change as their understanding of the narrative (i.e., comprehension) develops. Based on this, I expect two sub-hypotheses: A) Participants will attend to the target of attention more whenever they watch the common viewing period after watching the Context manipulation (Context condition) than when they watch the common viewing period without any Context (No-context condition). B) I also expect an interaction between the Condition and Frame Order, where the probability of fixating

the target AOI for all Context conditions will increase with increasing frame order, but decrease, or reach a plateau, during No-context viewing¹¹.

Tyranny of Film Hypothesis

Despite strengthening the event model using the Differential Event Model Criterion and minimizing stimulus-based features using the Low Default Gaze Probability Criterion, the presence of motion may still drive attentional selection; namely, the viewer's eye movements will always be driven by bottom-up features, and, despite differences in comprehension, they will remain the same. More specifically, since these stimuli are designed not to draw the viewer's attention to the target, in both Context and No-context, participants will not look at the target much. Hence, contrary to what I predicted in the Event Model Hypothesis, the Tyranny of Film Hypothesis will predict that there will be no difference in eye movements during the common viewing period between Context and No-context viewing conditions.

Methods

Participants

I recruited 53 undergraduate students from General Psychology courses who participated in exchange for course credit. Data from nine participants were excluded from the final analysis for the following reasons: poor calibration or loss of eye-tracking (5), falling asleep during the

¹¹ This hypothesis was initially generated based on preliminary analyses from Study 2; however, those initial models did not control for Type I error because complex random effect structures—specifically those including random slopes for Frame Order—failed to converge. These preliminary results suggested a significant interaction between time and condition, indicating that target importance increased rapidly over time in the [Context] condition, whereas it reached a plateau in the No-context condition. More recently, by employing a Complex Random Intercept approach to resolve these convergence issues, I was able to implement more robust model structures. Contrary to the preliminary findings, the updated results from the Importance Judgment Task do not support the initial prediction of a Frame order-by-condition interaction.

task (3), and failing to maintain position on the chin-rest (1). This resulted in a final sample of 44 participants. To ensure statistical power, I targeted a minimum of 40 valid participants per video. I employed a stopping rule to continue recruitment until this threshold was met, accounting for excluded or missing data. All participants were provided with informed consent under a protocol approved by the Kansas State University Institutional Review Board. Before the experiment, visual acuity was assessed using the Freiburg Visual Acuity and Contrast Test (FrACT). Inclusion required a minimum acuity of **20/30** (Bach, 2006). Although all recruited participants met this requirement, these screenings ensured they had sufficient visual acuity to maintain the validity and accuracy of their data.

Sample Size rationale. I had to base my power analysis on other related data because I did not run a pilot study for the proposed experiment. Moreover, I could not find any other study than Hutson et al. (2017) that used the *jumped-in-the-middle* paradigm and showed an effect of the viewer's event model on eye movements in film; thus, it is not directly comparable for power analyses. Apart from the 8s portion of the Houston et al. (2017) *Touch of Evil* study, any other research that has shown an effect of viewers' event models on their attention selection using a jumped-in-the-middle paradigm or a similar paradigm under special viewing conditions different from naturalistic video clip viewing was excluded. One such special viewing condition was a fixed-duration slideshow¹² version of the *Touch of Evil* video clip (Hutson et al., 2022). Another special condition was a self-paced pseudo-film/slideshow¹³ viewing condition using short silent film clips (Cabañas, 2024).

¹² 1 frame every 3 seconds is picked from the videos and each frame was presented for 3 seconds in the original order in film. Hence, the overall viewing time of the video remained the same.

¹³ 1 out of every six frames were selected from the videos and participants used their key presses to move through the frames at their own pace. This often produces a relatively smooth animation effect, but it is completely dependent on the speed with which the participant continually presses their button (which provides a novel dependent variable—the inter-button-press-interval).

For that reason, I chose the Importance Judgment Task mouse data from Preliminary Study 2 to estimate the predicted effect and run power analysis simulations. The rationale behind using the Preliminary Study 2 data for power analyses is two-fold. First, I wanted to estimate the number of participants needed to replicate similar differences in event-based importance for the selected 8 videos and 10 target AOIs. Second, I expect the effect of the condition and its interaction with frame order on mouse hit probability to resemble the probability of eye movement fixations within the AOIs¹⁴. I focused on 1) the main effect of Condition and 2) the interaction between Condition and Frame Order for the current power analyses. Nevertheless, mouse movements in the Importance Judgment Task are based on a meta-cognitive task, so I expect them to be both more deliberate and slower than eye movements. Hence, the effect I found on the Importance Judgment Task study is likely more pronounced than I would expect with eye movements. Hence, the Smallest Effect Size of Interest (SESOI) (Albers & Laken, 2018; Kumle et al., 2021) would be much lower than the current effect sizes found in the mouse-based Importance Judgment Study (Preliminary Study 2).

To determine the percentage of the Importance Judgment Task's effect size I can expect to find in the Eye-Tracking Study, I turn to the only previous study showing an effect of the event model on attentional selection while watching a film clip (Hutson et al., 2017). This is based on post-hoc analyses of the first 8 seconds of seeing the target of interest (a car, which only those in the Context condition had previously seen and knew had a bomb in its trunk)

¹⁴ Note: Initial power analyses were based on a simpler random effect structure due to convergence issues. While those earlier models may have inflated Type I error—yielding a significant interaction between Frame Order and Condition not found in the final model—the primary sample size was determined by the main effect of Condition. Since this main effect remains similar and significant under the more robust Complex Random Intercept (CRI) approach, the original power analyses remain valid.

during the common viewing period of the *Touch of Evil* opening scene. Hutson et al. (2017) found that the participants in the Context condition who showed the expected differences in their understanding had a higher proportion of fixations on the target of attention ($N = 65$, $M = 0.098$, $SD = 0.086$)(i.e., the car with the bomb in the trunk) than the No-context participants ($N = 62$, $M = 0.054$, $SD = 0.078$). I then compared the effect size, measured as Cohen's d_s for between-subject designs, reported in Hutson et al. (2017) ($d_s = 0.535$), with the effect sizes for each of the Target AOIs¹⁵ in the Preliminary Study 2 (Importance Judgment Task study). I found that this effect size from Hutson et. al. (2017) was roughly 64% and 66% smaller than the median effect sizes, measured as Cohen's d_{rm} ¹⁶ for repeated measures, of the 10 Target AOIs ($d_{rm} = 1.46$ and $d_{rm} = 1.57$) in Preliminary Study 2 (see Appendix I for effect sizes for all AOIs). Hence, a conservative rule of thumb could be to calculate the sample size needed to detect an effect that is only 65% of the effect found in the Importance Judgment Task mouse study.

Hence, I considered 65% of the fixed-effect coefficients (B) for all predictors of mouse hit probability from Preliminary Study 2 as my Smallest Effect Size of Interest (SESOI) to determine the number of participants required for the current study. For the power analysis, I employed the *mixed power* Package in RStudio (Team, 2022), as suggested by Kumle et al. (2021). Based on a multilevel binomial logistic regression from Preliminary Study 2, I conducted power simulations using 100 iterations with both the same parameter estimates from Preliminary

¹⁵ Hutson et. al. (2017) effects were based on just one video with one Area of Interest (i.e., the car with the bomb in the trunk) and post-hoc ANOVA analysis with Tukey's HSD comparisons. Hence, comparing this effect to individual effect sizes of each target of attention in preliminary experiment 2 would be the simpler and straightforward comparison.

¹⁶ Cohen's d for repeated measures (Cohen, 1988), is the standardized effect size calculation used for within-subjects data which enables comparing effects across within and between-subjects (see Appendix 8 for more details).

Study 2 (data-based) and parameter estimates reduced by 65% for all predictors (SESOI). Based on the results of this power analysis, a sample size of $n = 40$ and a 65% smaller effect size will still provide statistical power $> 80\%$ for both the Condition and the Condition $\times$ Frame Order interaction. Hence, 40 participants per video should provide sufficient statistical power to detect significant differences in the proportion of fixations on target AOIs ($\alpha = .05$) in a within-subjects design.

Materials

I used the eight pre-selected videos with 10 target AOIs that satisfied criteria 1 and 2 in the Eye-Tracking study. For these 8 video pairs, the Common Viewing Period clips have an average duration of 49.5 s, ranging from 10 s to 136 s (2 min, 16 s). Context clips have an average duration of 289.13 s (4 min, 49.13 s), a maximum of 620 s (10 min, 20 s), and a minimum of 46 s. These commercial films were shot in various aspect ratios ranging from 1.66:1 to 2.39:1. As the videos were run on a screen with a 16:9 aspect ratio, all the videos were converted to a “letterbox” format by adding black borders below and above, or on the sides, depending upon the original aspect ratio. This is done to avoid altering the original aspect ratio, ensuring that the pixels are not stretched and that the physical width and height remain the same. Hence, all the video files were presented in the MP4 format and have a resolution of 1920 x 1080. I played all the videos at 24 fps and displayed them on the same screen used in Preliminary Study 2, namely, the Iiyama G-master 27-inch screen with a 1920 x 1080-pixel resolution. Hence, the videos filled the entire screen natively.

Data Recording

I recorded all eye movements using an SR Research Eye-Link 1000 eye tracker, which tracks participants' eyes using the dark pupil and corneal reflection to estimate gaze at 1,000 Hz.

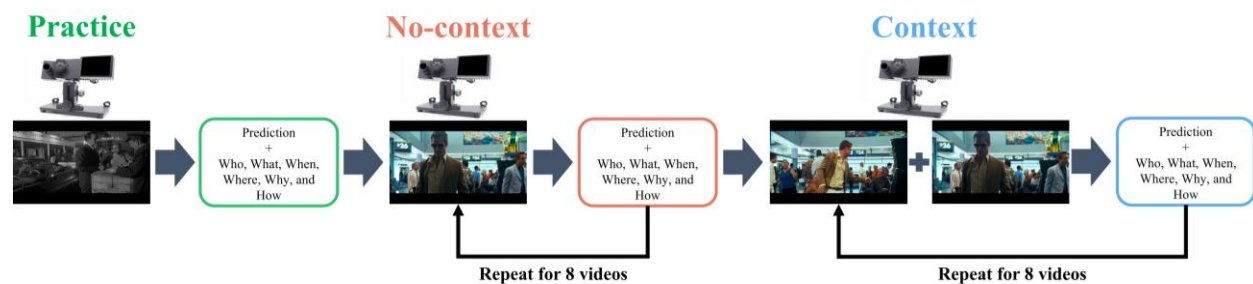
Subjects viewed videos with a forehead and chin rest to minimize sudden head movements by the participant. I performed calibration and subsequent validation at the beginning of every experimental session. I used a nine-point calibration procedure, which was further validated by ensuring an average spatial error of less than 0.5 degrees of visual angle (dva) and a maximum error of less than 1 dva, as recommended by Holmqvist and Andersson (2017).

Study Design

The eye-tracking study employed a within-subjects *jumped-in-the-middle* design (Figure 16) similar to the Preliminary Study 1 & Preliminary Study 2. The within-subject design enables one to control for any individual differences in attentional selection, such as differences due to cultural background (Chua et al., 2005; Masuda & Nisbett, 2001), working memory capacity (Unsworth & Engle, 2007; Unsworth et al., 2004), and so on.

Figure 16

Within-subject jumped-in-the-middle paradigm for the Eye-Tracking study. Their eyes will be tracked while they watch the videos.



In this within-subject study, participants first watched the No-context version of all 8 videos while their eye movements were tracked. After each video, the participant would give their predictions for what would happen next. After the No-context block of eight videos, the viewers watched each of the common viewing period videos again, this time only after watching each video's context video. The within-subject design also tests a causal effect for the Context

manipulation by employing a longitudinal comparison (time 1 vs. time 2) rather than a correlational effect between subjects.

To counterbalance any effect of viewing order (i.e., videos they see last in the sequence of 8 videos could be affected by fatigue or learning), in each of the conditions, I also have eight different orderings of the videos, also known as video sets (see Appendix J to see the order of videos in each set). I designed this counterbalancing scheme using the Latin square based on Williams design (Williams, 1949). Hence, Participants were randomly assigned to one of the eight video sets, ensuring that each video set had 5 participants (i.e., 5×8 video sets = 40 total).

Procedure

As mentioned before, I pseudo-randomly assigned participants to one of the eight video sets. All participants watched a practice video (the common viewing period of the *Touch of Evil* opening scene used in Hutson et al. (2017)) to familiarize them with the prediction task and to provide feedback on how rich I want their predictions to be. I told all participants they would watch video clips to understand them better while I tracked their eye movements. Each trial in each group began with a drift check, during which the participant was presented with a central fixation point (similar to the one used in the calibration and validation processes). They needed to press the spacebar while looking at the fixation point to start the video. Whenever participants' fixation locations during these drift checks are not within 2 degrees visual angle (dva) of the central fixation, I recalibrated them before proceeding to reduce calibration drift (i.e., a calibration spatial offset that increases over time).

Participants were presented with the common viewing period for one of the videos (No-context block) after the practice. Before each common viewing period, a central fixation cross flashes on and off at a rate of 60 ms per half-cycle, 4 times in a row, for a total of 480 ms (i.e., 4

x 120 ms = 480 ms). This should produce oculomotor capture of their gaze, ensuring that all participants start each common viewing period from the same central fixation location. After they finished watching each common-viewing-period video, they were asked, “What will happen next?” This was followed by prompts to disambiguate any vague responses by providing probing questions about who, where, when, why, and how regarding their prediction. After that, they repeated the same process for the following 7 videos. The order in which these eight videos are presented to them was based on the video-order counterbalancing set they were assigned to at the beginning of the study (see Appendix J for the video order in each set). Once they finish this No-context block, I give them a short break to avoid fatigue.

After the break, they were presented with the Context manipulation videos for the same set they had watched in the No-context block, in the same order. After they watched the first Context manipulation video, they saw the same flashing central fixation cross as in the No-context block before the common viewing period for that video began. After watching the common viewing period, I ask them the same “What will happen next?” prediction question, together with the same follow-up probing questions as in the No-context block. I asked them an extra question in this block: “Have you seen this film clip anytime before this experiment?” to which they can answer “Yes” or “No”. After these questions, they repeated the same process for the remaining seven videos. The videos presented in the Context block were in the same order as the No-context block, to maintain the same time gap between the first and second viewings of the common viewing periods across videos. After this block, I debriefed the participant and credited them for their participation.

Data Cleaning measures

To test the event model's effect in the *jumped-in-the-middle* paradigm, I need to ensure that the participant's No-context viewing is a true control and that they were not familiar with the video before this experiment. Hence, for participants who reported having seen a video clip before the experiment, I excluded their eye movements for that clip only from my analyses. Based on that, I removed a total of 9 responses: 4 from the *Ocean's Thirteen* video, 3 from the *Drive* video, and 2 from *Community* and *Mr. Robot*. 8 more responses were also removed for having a data loss rate $> 20\%$ for a video due to blinks or loss of track (Komogortsev et al., 2010). Finally, I also used a conservative cutoff of removing fixations below 40ms and above 3,000ms to clear eye movements that could be due to tracker error (similar to the cleaning procedure in Hutson et al., 2022), still retaining 98% of fixations for analyses.

Data Variables

Proportion of fixations within AOIs

To compare the time spent within the target AOIs, I compared the proportion of fixations within AOIs, a standard measure used in previous studies (e.g., Hutson et al. 2017, Hutson et al. 2022). Moreover, as mentioned before, minimal visual information is gathered during saccades due to saccadic suppression (Burr, Marrone & Ross, 1994; Dorr & Bex, 2013), so saccades within AOIs do not indicate visual processing of that region. Hence, only fixations were considered in AOI analyses. I analyzed the proportion of time spent within AOIs similarly to the mouse hit probability in Preliminary Study 2, which is by converting each participant's fixation location data on each frame, where the Target AOI is present, to a 1 (hit) if the fixation in that frame is located within the Target AOI, or 0 (miss). Based on this data, I can run multilevel binomial logistic regression models where the dependent variable is the probability of fixating

the Target AOI (i.e., AOI fixation hit probability). Details for various regression models to be run using this variable are listed in the analysis plan section.

To account for the probability of fixating on target Areas of Interest (AOIs) by chance - given both their screen area and inherent spatial viewing biases - I employed a spatial permutation procedure (adapted from Smith & Mital, 2013). For every frame containing a target, I generated a random baseline distribution of fixation hits using a Monte Carlo simulation (1,000 iterations). In each iteration, a surrogate AOI was created by sampling a random spatial location while maintaining the exact dimensions (width and height) of the observed target AOI for that specific frame. To establish a population-level estimate of viewing behavior, I pooled all fixations from both the Context and No-context conditions. I then calculated the proportion of fixations within the surrogate AOI coordinates across all 1,000 iterations. This generated a spatial baseline that accounts for the physical size of the AOIs and systematic gaze tendencies, such as the center bias (Dorr et al., 2010; Kümmerer et al., 2017; Tatler et al., 2011). While AOI size and frame duration varied across videos, the grand mean of this random fixation hit probability ($M = .031$, $SD = .069$) is utilized as the control baseline. Finally, the observed fixation proportions for the actual target AOIs were compared against this permuted distribution to determine statistical significance.

A priori Predictions from Preliminary Studies

Stimulus-based Default Gaze prediction ($P(D)$). The a priori estimate for the stimulus-based Default Gaze prediction ($P(D)$) is simply the default gaze probability for a given AOI and frame order in the video, calculated as the sum of the individual default gaze probabilities of all pixels within that AOI. The probability of each pixel in a frame is already estimated using the

DeepGaze IIE algorithm in Preliminary Study 3. For multilevel regression analyses, these values will be scaled and centered to standardize the unit change in both a priori predictions.

Event-based Importance prediction ($P(I)$). The a priori estimate for the event-based Importance is the proportion of mouse hits ($P(I)$) on a given AOI for each frame order, averaged across the 30 participants from Preliminary Study 2. This has already been estimated in Preliminary Study 2. For multilevel regression analyses, these values will be scaled and centered to standardize the unit change in both a priori predictions.

Analysis Plan

Proportion of fixations within AOIs

I analyzed the proportion of time spent within AOIs when the Target AOI is present by converting each participant's fixation location data on each frame to a 1 (hit) if the fixation in that frame is located within the Target AOI, or 0 (miss). Based on this data, I ran a multilevel binomial logistic regression where the dependent variable is the probability of fixating the Target AOI (i.e., AOI fixation hit probability). The random effect structure for the full model, which includes condition, frame order, and their interaction as predictors, was selected using a Complex Random Intercept (CRI) approach (Scandola & Tidoni, 2024) to prevent pseudoreplication (i.e., inflation of degrees of freedom) and reduce the risk of Type I errors. This method, according to Scandola & Tidoni (2024), also reduces singularity and non-convergence issues caused by overparameterization. The full-CRI model yielded a singular fit; I removed the CRI with the lowest variance, as recommended by Scandola & Tidoni (2024) (see Appendix K for details). The resultant final full random-effects structure included basic random intercepts for my grouping factors (AOI and Participant), as well as complex grouping terms representing the residual variance of AOI and Participant interacting with Condition. Finally, the continuous

variable of Frame Order was included as a random slope attached to the AOI, AOI:Condition, and Participant:Condition grouping terms.

Now, keeping this random effect structure constant, I compared nested models to test various hypotheses, starting with the Tyranny of Film (null) hypothesis, represented by a null model with only an intercept and the pre-selected random effect structure. To test the event model hypothesis, I added Condition (Context vs No-context) as a predictor and compared the model performance to the null model. The nested models are then tested using a likelihood-ratio test to identify the best-fitting model for the fixation-hit probability data and to reject the null or alternative hypotheses. The effects of frame order (time) on viewers' likelihood of fixating the target were further tested by first adding Frame Order and then adding the interaction between the Condition and Frame Order as predictors to the model. A likelihood ratio test was run at each step.

Relative Weights of stimulus-based default gaze probability and event-based importance

In this analysis, I again used the probability of fixating target AOIs as the dependent variable to estimate how well my a priori predictions fared against participants' ground-truth attention to the target. More importantly, this analysis allows us to assess the relative contributions/weights of my a priori probability of fixation estimates from Stimulus-based Default Gaze predictions and Event model-based AOI Importance (mouse hit) probability. To estimate the relative weights participants' gaze has given to these two a priori estimates, I ran multiple hierarchical multilevel logistic regressions predicting the probability of fixating the Target AOIs to evaluate how adding each a priori estimate improves the model's estimates of the dependent variable.

However, by design, the condition variable is highly correlated with the mouse hit probability estimates. In preliminary study 2, I already found that the mouse hit probability is highly dependent on condition: participants had significantly more mouse hits on the target during the Context condition than during the No-context condition. Hence, these will lead to strong multicollinearity issues in the model if I were to consider both Condition and mouse hit probability as predictors in the same model. Moreover, I am primarily interested in comparing the relative weights of the default gaze probability and mouse hit probability, rather than how the a priori predictions interact with the Condition. Hence, I ran two separate model comparisons and analyses for the Context and No-context data. Given the systematic selection of the clips, I already expected a low rate of fixations to the target during the No-context condition. Moreover, the importance of the target is expected to be uninformative for predicting fixations to the target, given the poor event model they would have in the absence of context. I still primarily looked at the No-context condition gaze data to see whether saliency could still predict the small number of fixations to the target. In other words, I can examine whether the target's relatively low saliency influences participants who end up looking at it during the No-context condition.

I kept the same random-effect structure, selected as the best for the proportion of fixation analyses; however, I now omitted the condition slopes for both AOIs and participants, since models for each condition are run separately. Hence, the random-effect structure I kept constant across nested model comparisons includes random intercepts for each participant and each Video and AOI combination (as some videos have multiple AOIs), along with random slopes for frame order varying by both participant and Video and AOI combinations. I started the hierarchical regression comparison with the *Base model*, in which the predictors of Fixation Hit Probability are only the intercept and the aforementioned random effects. I then added Default Gaze

Prediction ($P(D)$) as a new predictor to the Base model to create the second model, namely *the Base + Stimulus Features model*. I evaluated whether it improved the Base model. The third and final model, namely *Base + Stimulus Features + Event Importance model*, was run by adding the Event-based Importance prediction ($P(I)$) as the new predictor. I then again compared the new model with the previous model using a likelihood ratio test and checked whether the new model improved the model fit. These steps remain the same for both Context and No-context datasets. I expected the best model in both cases to be the Base + Stimulus Features + Event Importance model, as both a priori predictions are expected to predict viewers' attention to the target. Importantly, the fixed-effect coefficients of the best model provide estimates of the relative contributions of stimulus-based and event-based predictions to attention in both the Context and No-context conditions.

Results

Proportion of fixations within AOIs

Based on the event model hypotheses, I predicted that participants would make a greater proportion of fixations to the target AOIs in the Context condition than in the No-context condition. In contrast, the Tyranny of Film (null) hypothesis would predict no effect of the condition. A likelihood ratio test reveals that the model¹⁷ (see Table 5) with condition as fixed effects provided a better fit to the AOI fixation hit probability data compared to the Tyranny of film model ($\chi^2(1, N = 44) = 11.12, p < .001$), indicating strong evidence against the Tyranny of film hypothesis in favor of the event model hypothesis. Planned pairwise comparisons indicated a significant effect of condition; participants were 5.71 times more likely to fixate on the target

¹⁷ No significant overdispersion (*dispersion ratio* = 0.415, $p = .128$) detected using *performance* package (Lüdtke et al., 2019) in R.

when in the Context condition than in the No-context condition ($OR = 5.71, z = 4.21, p < .001$). Hence, when both criteria are satisfied, the participants will look less at the target when they are not aware of its importance (No-context: $P(\text{target fixation})^{18} = .014, SE = .005$) and will shift their attention to the target after seeing the context and know its importance (Context: $P(\text{target fixation}) = .077, SE = .023$) (see Figure 17b).

Table 5

Parameter Estimates for the Multilevel Binomial Logistic Regression Predicting AOI Fixation Hit Probability.

Variable	B	SE(B)	Z	Significance(p)	OR	95% CI for OR	
						Lower	Upper
Intercept	-3.35	0.25	-13.50	<.001	0.03		
Condition (Context)	0.87	0.21	4.21	<.001	2.39	1.59	3.59

Note: Condition was effect coded (No-context = -1; Context = 1).

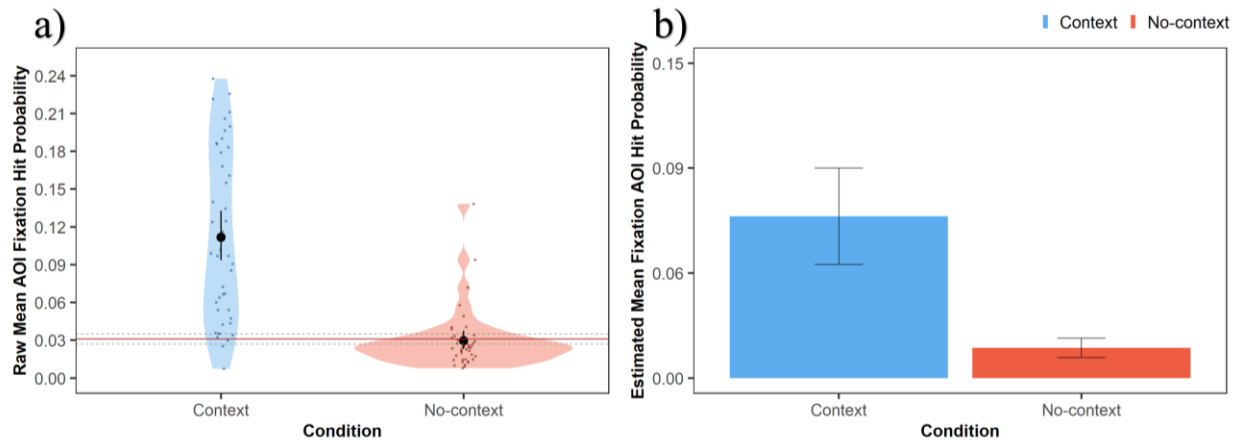
Analysis of raw fixation distribution revealed that the target AOIs attracted significantly more attention in the Context condition than expected by chance. Across all frames and videos, the raw proportion of fixation hits in the Context condition ($M = .111, SD = .097$) was significantly higher than the permuted random baseline (chance) for AOIs of same size ($M = .031, SD = .069$), as indicated by the mean Z-score ($z = 2.26, p = .01$). While in No-context condition ($M = .030, SD = .053$), the proportion of fixations within AOIs remained at chance ($z = 0.16, p = .43$) suggesting that the target AOIs in No-context condition did not attract attention beyond what would be expected by chance (see Figure 17a). Hence, these comparisons indicate that the visual features of the target AOI are insufficient to drive the viewer's attention beyond what is expected given its size and position in the No-context condition. In the Context

¹⁸ Estimated marginal mean probability of fixating the target AOI based on the multilevel logistic regression model

condition, participants looked at the target because it was important for comprehension, regardless of AOI size, location, or visual features.

Figure 17

Raw average and multilevel binomial logistic model estimated fixation hit probability on target AOIs of the condition (Context vs No-context).



Note. The error bars are equal to +/- 95% confidence intervals. The horizontal brown line and the dotted lines represent the random probability of fixating the target and the +/- 95% confidence interval of that permuted random distribution. The observed differences between raw and model-estimated means result from the generalized multilevel model's use of partial pooling, which applies shrinkage to item-level estimates. Because my AOIs exhibited large variance in fixation hit probability, raw means were susceptible to skew from atypical items; however, model-Estimated Marginal Means (EMMs) provide a more robust, generalizable estimate of the condition effect by downweighting the influence of these outliers (Gelman & Hill, 2006). While these EMMs serve as the basis for inferential testing, the raw means are retained in the above visualizations primarily to compare specific AOI performance against random-chance levels.

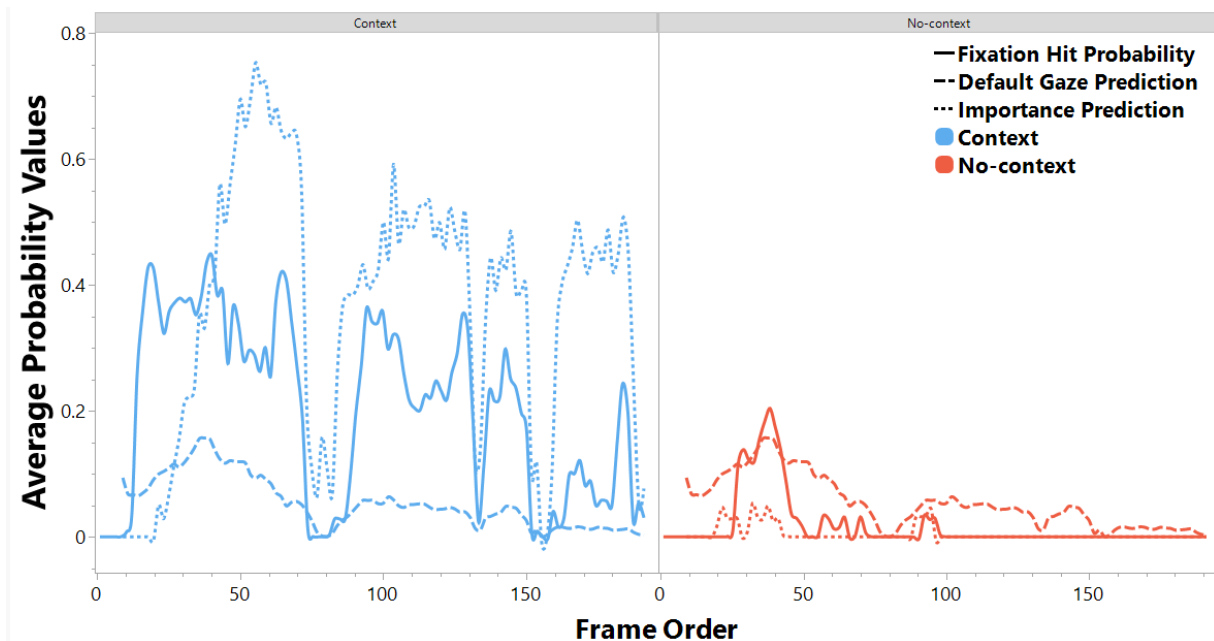
I also hypothesized that participants' likelihood of fixating on the target could increase over time on the video (frame order) as their predictions became more imminent, and that this rate of increase could vary across conditions. However, adding frame order as a predictor did not

improve the model fit compared to the condition-only model ($\chi^2(1, N = 44) = 1.81, p = .18$). Furthermore, adding an interaction between frame order and condition also did not improve the model ($\chi^2(1, N = 44) = 3.37, p = .07$) (see Appendix L for complete model comparisons). Hence, only the event-model condition appears to have a substantial effect on attentional selection. At the same time, the probability of fixating the target did not increase or decrease over time during the AOI's presence in the common viewing period video. This is probably because there is a lot of variance in the random slopes for frame order ($T_{II} = 12.90, SD = 3.79$), indicating that the rate of change in proportion of fixation with frame order varied a lot across target AOIs within the different conditions. Moreover, as you can see in the raw proportion of fixation hits on target over time for the *Ocean's Thirteen* clip in Figure 18, the proportions of fixation hits fluctuate over time, and there is not a unidirectional change with frame order. If time in the video is not causing the increase or decrease in the likelihood of fixating the target, then what is modulating these fluctuations in fixations-to-target probability over time? There will be some systematic fluctuations in probability values due to changes in the size of the dynamic target AOI across frame order; however, those large changes will be the same across the a priori predictions, as their respective calculations were made using the same AOIs. Hence, when controlled for these variations with frame order, I should be able to measure the relative impact of each a priori prediction. In fact, in Figure 18, you can see that a priori prediction seems to modulate together with Fixation hit probability, especially in the Context condition. In the Context condition, you can see that changes in Importance prediction also align mostly with changes in the Fixation hit probability. In No-context, the overall proportion of fixations to the target is very low. Still, I can at least see a major common peak in both the Default Gaze Probability and Fixation hit probability around frames 25 to 45 in the common viewing period

video. I can statistically measure these influences using the relative weights of the a priori predictors in a multilevel logistic regression aimed at predicting fixation hits to the target.

Figure 18

Frame by Frame changes in the raw average probability values of fixation hit probability and the two a priori predictions (Default Gaze Prediction and Importance Prediction) in the Ocean's Thirteen clip for both Context and No-context viewing.



Note. The fixation hit probability goes up and down over time; the large dips found in all three curves indicate large changes in AOI size or entry/exit of the AOI. Beyond that, you can see that the changes in Importance prediction (dotted lines) roughly accompany changes in Fixation hit probability in the Context condition (*Blue* lines). For default gaze probability (dashed lines), there is one major peak in the No-context (*Red* lines) that matches with the only major peak in Fixation hit probability around the Frame Order of 25 to 45. See Appendix M to see the same comparison for all other AOIs.

Relative Weights of stimulus-based default gaze probability and event-based importance

For Context condition, I found that AOI fixation hit probability predictions compared to the Base model (only intercept and random effects) improves with the addition of the default gaze probability ($\chi^2(1, N = 44) = 1135.35, p < .001$), which is further improved by the addition of the AOI importance probability ($\chi^2(1, N = 44) = 824.15, p < .001$) (see Appendix N for detailed model comparisons). This indicates that the best model is the Base + Stimulus Features + Event Importance model¹⁹, meaning that both a priori predictions provide meaningful contributions to predicting fixations to the target AOIs. Based on this model, both the Importance prediction ($B = 0.40, p < .001$) and the Default gaze prediction ($B = 0.25, p < .001$) were significant unique predictors (Table 6, for details). For a one-standard-deviation change in Importance prediction ($M = .28, SD = .24$), the odds of looking at the target increase significantly by 49% ($OR = 1.49, SE = 0.02$), which, on average, corresponds to a 3.12% increase in the fixation hit probability. Similarly, for every 1 standard-unit increase in Default gaze prediction ($M = .05, SD = .05$), the odds of a fixation hit increase significantly by 29% ($OR = 1.29, SE = 0.01$), which, on average, is an increase of 1.98% in the fixation hit probability.

To determine which a priori prediction had a greater relative impact on fixation hits on target, I compared the standardized regression coefficients using the *emmeans* package (Lenth & Piaskowski, 2017). A direct comparison of the slopes revealed that the effect of Importance Prediction was significantly stronger than that of default gaze prediction ($\Delta B = 0.14, z = 7.28, p < .001$) (see Figure 19a). Hence, event-based importance is a strong predictor of participants'

¹⁹ No significant overdispersion (*dispersion ratio* = 0.685, $p = .496$) detected using *performance* package (Lüdtke et al., 2019) in R.

gaze to the target during the Context condition, and has a stronger influence on attentional selection than stimulus-based default gaze probability. This indicates top-down biasing of gaze competition during film viewing when both criteria are satisfied.

Table 6

Parameter Estimates for the two separate Multilevel Binomial Logistic Regressions for each condition, where a priori predictions are predicting the AOI Fixation Hit Probability.

Condition	Variable	<i>B</i>	<i>SE(B)</i>	<i>Z</i>	<i>Significance(p)</i>	<i>OR</i>	95% CI for OR	
							Lower	Upper
Context								
	Intercept	-2.64	0.26	-10.09	<.001			
	Importance prediction (<i>P(I)</i>)	0.40	0.01	28.25	<.001	1.49	1.45	1.53
	Default gaze prediction (<i>P(D)</i>)	0.25	0.01	23.35	<.001	1.29	1.26	1.32
No-context								
	Intercept	-4.41	0.33	-13.24	<.001			
	Importance prediction (<i>P(I)</i>)	0.04	0.01	3.88	<.001	1.04	1.02	1.07
	Default gaze prediction (<i>P(D)</i>)	0.31	0.01	26.30	<.001	1.37	1.34	1.40

Note: Both a priori predictions, such as stimulus-based default gaze prediction and event-based importance prediction, are scaled and centered.

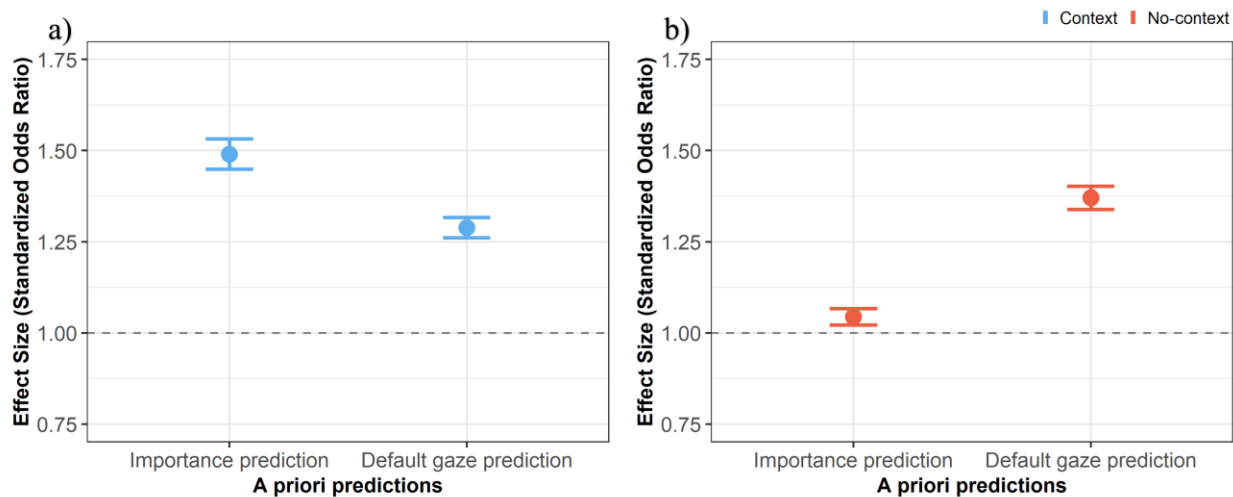
For the No-context condition, I also found that the complete Base + Stimulus Features + Event Importance model still provided a better fit to the data than the model with only default gaze prediction as a predictor ($\chi^2(1, N = 44) = 14.22, p < .001$). Furthermore, the default gaze prediction only model was better than the Base model ($\chi^2(1, N = 44) = 757.13, p < .001$) (see Appendix N. for detailed model comparisons). Based on the Base + Stimulus Features + Event Importance model²⁰, the default gaze probability has a strong effect on the likelihood of looking at the target ($OR = 1.37, SE = 0.02, p < .001$), where one standard deviation change in default gaze prediction ($M = .05, SD = .05$) increases the odds of fixating the target by 37%, which is on

²⁰ No significant overdispersion (*dispersion ratio* = 0.194, $p = .2$) detected using *performance* package (Lüdtke et al., 2019) in R.

average is an increase of 0.7%²¹ in the fixation hit probability. It is important to note that the Odds Ratio of the default gaze probability on attention in the No-context condition is slightly higher than the Context condition ($OR = 1.29$, $SE = 0.01$), indicating a stronger influence of stimulus-based features on viewers' attention to the target during the No-context viewing (see Figure 19b).

Figure 19

Effect size (Odds Ratio) of both a priori predictions on predicting AOI fixation hit probability for each condition, Context and No-context separately.



Note. The error bars are equal to $\pm$ 95% confidence intervals. The effect size, measured as an odds ratio, indicates the rate of change in the odds of fixating the target for a 1-standard-unit (SD) increase in the a priori predictions. Finally, the dotted line indicates no change in fixation hit probability.

²¹ The seemingly small average change in predicted probability (0.7%)—despite a higher Odds Ratio for the No-context condition—is a known artifact of the logistic regression's sigmoid geometry. Because the sigmoid curve is flattest at its extremes, the marginal change in probability depends heavily on the baseline position. In my data, the No-context condition lies in the lower 'tail' of the curve (95% CI: 2.6–2.8%), where the slope is nearly horizontal. Conversely, the Context condition (95% CI: 12.4–12.8%) sits further up the curve's steeper section. Consequently, a high Odds Ratio in the No-context condition results in a smaller absolute probability shift than a lower Odds Ratio would in the Context condition.

Finally, while the model indicated a statistically significant positive association between importance prediction and target gaze, the effect size was minimal. A one-standard-unit increase in Importance prediction ($M = .01$, $SD = .03$) resulted in only a 4% increase in the odds of a target hit ($OR = 1.04$, $SE = 0.01$, $p < .001$). When transformed to the probability scale, this effect yielded a significant average marginal slope close to zero ($Slope = 0.001$, $z = 2.78$, $p = .005$), corresponding to an average of 0.1% increase in fixation hit probability. A direct comparison of the slopes revealed that the effect of default gaze prediction was significantly stronger than that of Importance prediction in the No-context condition ($\Delta B = 0.27$, $z = 16.73$, $p < .001$).

In conclusion, I have a contrasting pattern between the Context and No-context conditions. In the Context condition, importance prediction has a relatively greater influence on fixations to the target than the default gaze probability. In No-context, the default gaze probability has a stronger influence on attention to the target than the minimal influence of importance prediction. These trends indicate the successful dissociation of bottom-up stimulus-based default gaze prediction from event-based importance predictions for attentional selection of the target.

Chapter 5 - Discussion

Many people might assume that viewers' eye movements are influenced by their moment-to-moment comprehension as they watch films or TV. Surprisingly, several recent studies have shown evidence of the Tyranny of film, namely that big differences in comprehension nonetheless produce nearly identical gaze patterns (Davis et al., 2021; Huff et al., 2014; Hutson et al., 2017; Loschky et al., 2015). This study has asked whether this is true for all directed videos, or whether there are specific conditions under which differences in event comprehension lead to differences in attentional selection. The answer to this question has important implications for theories of the relationship between comprehension and attention (e.g., Loschky et al., 2018, 2020, 2025) and has real-world applications in mass media and education.

My results have shown that when there is a target of attention that satisfies both of the two proposed criteria, namely, 1) differential prioritization of the target based on the context (Differential Event Model Criterion), and 2) low default gaze probability of the target, then the viewer's event model can influence attentional selection, breaking the Tyranny of film.

My results from the Eye-Tracking Study showed that, once participants in the Context condition became aware of the importance of the target of attention, they were more likely to look at the target than when in the No-context condition. Preliminary Study 2, which used the mouse-based Importance Judgment Task, showed that participants in the Context condition perceived the target as more important than in the No-context condition, aligning with the gaze results in the Eye-Tracking Study. In Preliminary Study 3, I showed that the targets of attention have relatively low default gaze probability and do not stand out from the rest of the scene, thereby verifying the Low Default Gaze Probability criterion. Importantly, the Eye-Tracking Study found that this was also reflected in the gaze data: participants in the No-context condition,

who lacked strong event-model guidance, only looked at the target at a frequency expected by chance. Hence, as predicted previously in my Figure 7, in the No-context condition, satisfying the Low Default Gaze Probability Criterion means that the stimulus features should not reflexively draw viewers' attention to the target, and that their impoverished event-model should likewise lack the prior knowledge needed to direct their attention to it. Conversely, in the Context condition, once the target of attention's moment-to-moment importance increases, satisfying the differential event model criteria, a dissociation between stimulus-based features and event-model-based importance is created. This opens the opportunity for the event model to influence attention to the target, and my Eye-Tracking Study showed that it did.

In both Preliminary Study 2 and the Eye-Tracking Study, time did not emerge as a significant fixed effect. This suggests that neither the perceived importance of the target nor the likelihood of fixating upon it increases or decreases monotonically as a function of its duration on screen. Instead, temporal trajectories appear highly AOI-specific: while some targets showed a strong increase in attention and importance over time, others declined or remained stable. These results indicate that temporal changes are unique to individual AOIs, necessitating follow-up content analyses to identify specific patterns in event model and gaze behavior. Although I initially hypothesized a monotonic relationship, these findings suggest that the influence of time may be non-monotonic—potentially following a quadratic, sinusoidal, or inverted-U curve. Furthermore, clip-based differences in the timing of shifts in event-model relevance and default gaze probability may be obscuring global temporal effects. Consequently, my findings demonstrate that fluctuations in the likelihood of looking at a target are modulated more effectively by moment-to-moment changes in default gaze probability and situational importance than by the mere duration of the target AOI's on-screen presence.

In the Eye-Tracking Study, in the No-context condition, the default gaze probability had a moderate effect on the likelihood of fixating the target. A naïve assumption would hypothesize that stimulus-based default gaze features have the same effect on attention across conditions. However, in the Context condition, the default gaze probability was again a moderate predictor of fixations to the target, but this time that influence was reduced compared to the No-context condition; specifically, the event-based importance was now a stronger predictor and was calculated to have a higher weight compared to the default gaze probability. The above results are consistent with the Biased Competition Theory (Adeli & Zelinsky, 2018; Desimone & Duncan, 1995; Tsotsos et al., 1995), which employs a winner-take-all approach where an increase in the influence of the top-down event model factor will lead to a decrease in the relative influence of stimulus-based bottom-up features. This is especially true when bottom-up stimulus-based features are dissociated from top-down influences, as accomplished in this study by verifying, in Preliminary Studies 1-3, that both hypothesized criteria for reducing the Tyranny of film are met. Thus, when the event model is richer and has a top-down prioritization of the target, the importance map modulates viewers' attention to the target. However, in the No-context condition, without any strong event-model biasing toward the target, their bottom-up default gaze probability becomes the primary factor driving attention to the target. In this condition, the viewer is still likely constructing a default event model of the primary character(s) and events, which influences gaze toward those key components. Furthermore, the creation of this default event model is initially heavily influenced by bottom-up factors, leading to high congruence between top-down and bottom-up influences. However, this default event model often does not include the target of attention and results in showing only a minimal effect of Importance on attention to the target during the No-context condition.

Broader Application of the Current Method of Breaking the Tyranny of Film

The current set of studies provides a method for breaking the Tyranny of film. This could therefore be used with a new set of videos by following a similar procedure. Methods used to achieve Criteria 1 and 2 can be used to identify and create new videos, and new additions to the criteria can also be employed to test new hypotheses. The experiment results should be interpretable as long as Criterion 1 (Differential Event Model Criterion) helps to create a “strong” and measurable event model that points to the target of attention, and Criterion 2 (Low Default Gaze Prediction Criterion) minimizes the visual saliency of the target of attention one way or another; any other controlled changes should be workable. One can run Deep Neural Network-based default gaze probability maps of the selected videos to confirm whether the dissociation of the theoretically-driven target of attention (mostly linked to the differential elements) is inconspicuous. One can also use the threshold I used in this study, where I compared the default gaze probability values to the upper fence of the random baseline distribution to determine its conspicuity. However, this threshold could be made either conservative or liberal. It may need further refinement based on follow-up studies involving the clips/AOIs that failed the Default Gaze Criterion selection, or other studies involving other videos. I have developed the two criteria above for situations in which I believe pre-existing film clips can be used to demonstrate the impact of the viewer's event model on their attentional selection. I propose that future studies use existing film clips or, even better yet, create experimenter-generated films that adhere to these criteria.

Moreover, Importance maps, with further validation, could be used to empirically identify what viewers' event models consider essential in a visual narrative, especially in dynamic images. Point-click maps have already been used to generate static narratives and

stimuli to produce “importance” maps (Hutson et al., 2018; Kim et al., 2017). However, to my knowledge, no previous empirical method can help identify what is important in a dynamic scene based on the viewer's event model (i.e., their moment-to-moment understanding). Suppose that the importance map can effectively filter default gaze prediction maps, thereby improving their gaze movement predictions in the Context condition. In that case, it could also imply that importance maps can be a practical way to identify priority maps for event model processes.

Event model and default gaze congruence spectrum of videos

Imagine a spectrum in which the right end consists of film clips in which event-model-driven biasing is fully congruent with the targets or locations of highest default gaze probability. Films at this end of the spectrum feature event-relevant targets that are also the most visually salient, typical of most conventional cinema, where filmmakers successfully direct attention to emphasize narrative importance, leading to the Tyranny of Film phenomenon. I could claim that this end of the spectrum represents high guidance, characterized by explicit attentional direction by the creator.

By contrast, the left end of the spectrum consists of clips in which event-model influences are fully incongruent with default gaze patterns. Films at this end feature event-relevant targets that are among the least salient objects or locations in the scene. This end also includes clips where the filmmaker intentionally allows the viewer to explore the scene to find relevant targets; my selected videos fall into this category. These stimuli can be said to have low guidance. While most videos lie between these extremes, the level of congruence between event-model priority and default gaze prediction decreases from right to left.

My Low Default Gaze Probability criterion identifies the boundary on this spectrum beyond which an event-model-relevant target, when prioritized, can successfully influence

attentional selection. While the examples considered here are film clips, I hypothesize that this spectrum applies to all dynamic stimuli. For instance, CCTV or amateur footage likely lies at the far-left end, where event targets are not necessarily the most salient items. This is why news editors or documentary filmmakers often add explicit visual cues, such as a red circle or arrow. By adding these cues, they are essentially shifting the clip from the incongruent end of the spectrum toward the congruent end.

This spectrum may also explain why naturalistic stimuli have been particularly successful in exhibiting event-model effects on attention. For example, Wahlheim et al. (2022) used videos of everyday activities with presumably equal default gaze probabilities. The effect of the event model was demonstrated through viewers' attention to a target versus an alternative item (e.g., an actor placing a key in a doorknob vs. a deadbolt) based on prior knowledge. In this case, the elements were designed not to differ in salience; therefore, neither would reflexively draw attention without prior knowledge. Consequently, such stimuli would likely meet my second criterion.

While I have not yet tested this hypothesis using *DeepGaze III*, these examples satisfy my first criterion, i.e., viewers differentially prioritize targets based on their stored event model. By extrapolation, it is possible that these two criteria—Differential Event Model and Low Default Gaze Probability—could be generalized to all previous research showing event-model effects on attention in moving images. A subsequent meta-analysis of such studies could confirm whether these criteria serve as a universal framework for understanding top-down attentional selection across all filmic and non-filmic media.

Role of familiarity

One of the advantages of this study is its within-subject design, which allows for causal inference about the relationship between context and attentional selection and controls for individual differences (Chua et al., 2005; Masuda & Nisbett, 2001; Unsworth & Engle, 2007; Unsworth et al., 2004). However, the same design could also introduce a caveat: participants will watch the common viewing period twice, first without context and second with context. Hence, familiarity with the common viewing period during the second viewing could influence viewers. Viewing the *common viewing period* a second time influences eye movements (Dorr et al., 2010). Even if it does not explain eye movements to the target AOIs, it could still influence viewers' overall eye movements and thus reduce or increase the frequency of fixations to the targets. Participants in the repeated-viewing condition will show less attentional synchrony than during their first viewing. Namely, participants tend to look at similar locations during their first viewing. However, during repeated viewing, they tend to look at different parts of the scene and explore it more (Dorr et al., 2010). For this reason, conducting a future study using a between-subjects design could eliminate that issue and would therefore be a good idea. Our laboratory is currently doing that.

Role of individual differential event elements in attentional selection

Based on this research, I foresee clear future research directions for exploring the effects of the event model on attentional selection across both theoretical and applied sciences. The results of Eye-Tracking Study 4 showed an apparent effect of the event model on attentional selection for my selected videos. In doing so, it has provided an existence proof that understanding events in certain conditions can overcome the Tyranny of Film and influence attention selection, even in commercial film clips. Future studies can be designed to investigate

the underlying necessary and sufficient conditions for breaking the Tyranny of Film. For example, one can explore how crucial each differential element is for creating differences in attentional selection. Can the effect be just an agent effect (Hutson et al., 2017), namely, does the participant attend primarily to a character or object newly introduced in the Context condition? If so, the attentional effects should persist even after removing other event indices from the event model. Researchers can test the *agent effect* by comparing eye movements between two groups of participants, with one group receiving the same context manipulation as in the current study. The second group could have the context manipulation video frames randomized across 1s bins (for example, Dorr et al., 2010), so it would be very difficult to understand the entire context (hence, difficult to identify many other differential elements other than the entities, particularly time, goals, causality, and temporal operativity). Still, they will have enough visual exposure to the differential entities to identify them (and probably also locations, actions, and emotions) in the common viewing period.

Implications for filmmaking and viewers' attentional modulation

Famous film director/screenwriter Alexander Mackendrick, in his book *On film-making: An introduction to the craft of the director*, said: “What a film director really directs is the audience’s attention” (Mackendrick & Cronin, 2005, p. 200). Over time, many filmmakers have perfected this art of directing attention to highlight what’s important to a story and to de-emphasize what’s not. They employ various techniques to achieve this, including visual composition, ensemble staging, focus pulls, camera movements, sudden character movements, etc. (Bordwell et al., 2024). These create strong synchronization across viewers’ eye movements. As their attention is typically stimulus-controlled by the filmmakers, this produces effortless, primarily attentionally passive viewing for the audience. This close link between what’s

important to the film's narrative and the most attention-grabbing scene elements could be a major reason why studies in the fields of scene perception, and event comprehension—using conventional film clips—failed to show any strong effects of the event model on attention known as the Tyranny of Film (Hutson et al., 2017; Loschky et al., 2015). Although the Tyranny of Film seems to be the norm in many well-directed films, many filmmakers also choose not to actively guide viewers' attention, instead allowing them to explore the scene freely. Specifically, Austrian director Michael Haneke has stated that he prefers to “...give the spectator the possibility of participating. The audience completes the film by thinking about it; those who watch must not be just consumers ingesting spoon-fed images” (Conrad, 2012).

In the clips used in the current study, the filmmakers placed strong emphasis on the target by establishing context. Still, they provided little or no visual emphasis on it during the common viewing period, thereby allowing the viewer's event model to fill in the gaps. This approach was most effective for time-sensitive events, analogous to Hitchcock's hypothetical practice of setting a bomb's timer at the scene's outset to build suspense. If the event trigger is clearly emphasized early on, the filmmaker can step back, allowing the viewer's understanding to focus attention on the target in anticipation of the outcome. For example, as discussed before in the *Ocean's Thirteen* clip, the filmmaker earlier in the film shows a lot of emphasis on the special coins and their impact on slot machines by showing the tactics employed by Brad Pitt's character to slyly give the coin to a stranger and show the clear effect of its use on the slot machine by showing a stranger winning big. This helps establish the cause-and-effect relationship between the special coin and hitting the jackpot, which should now be stored in the viewer's working memory and later in their long-term episodic memory. Later, a few seconds before the start of the common viewing period, the filmmaker initially drew the audience's attention to the transfer of the special

coin to the stranger by Brad Pitt, by showing that interaction in close-up. However, for the next 10 seconds in the common viewing period, the director decided to stay back and do nothing more to draw the viewer's attention to the stranger in the background. The filmmaker, in fact, moved the camera farther, reducing the target's default gaze probability. However, since the viewer's episodic memory contains the previous event with the coin and its effects, they can anticipate what will happen next. They can shift their attention to the stranger in the background occasionally to verify their prediction. For some videos used in the current studies, maintaining a consistent visual composition or camera perspective when the target was introduced in the context provided a schema for the location, enabling viewers to locate the target later without the filmmaker explicitly guiding their visual attention. Low visual emphasis on the target during the primary (common viewing period) clip can be achieved by placing the target off-center, out of focus, with limited motion, or by surrounding it with numerous distractors. When these methods are used wisely, the filmmaker can draw the viewer into actively participating in creating the narrative—encouraging them to make inferences, recognize dramatic irony, and anticipate suspenseful outcomes. Moreover, it provides a deeper understanding of how filmmakers can control attention, or, conversely, allow viewers to direct their attention to narrative content. This has perhaps been nowhere more clearly shown than in the final static-camera long shot from Haneke's *Cache* (Haneke, 2006) used in the current study, which notoriously allowed, and required, the film viewer to guide their own attention. In fact, the scene from *Cache* was a touchstone for the entire study and had one of the largest effect sizes of any clip, despite having targets that were noticed by only 62% of participants. This is because participants have to put effort into finding targets in a crowd. Still, once they find them, they tend to give their undivided attention to those targets (see Appendix A, *Cache* section, for target AOIs and videos).

Chapter 6 - Conclusion

The present study demonstrates that the Tyranny of Film can be reduced and that a stronger role for event comprehension in directing attention can be achieved when two criteria are satisfied by the video clips. From the results, I conclude that when there are minimal bottom-up stimulus-driven or medium-specific effects on attention to a particular target on the screen, the event model's prioritization of that target can still influence viewers' gaze toward it. Additionally, it indicates that any research on attentional selection using found-footage film, or indeed any dynamic stimuli, must ensure a clear separation of stimulus-based effects from top-down influences on relevant locations or object features, so that their individual impacts can be distinguished in the findings. This study makes important theoretical contributions by providing a robust test of SPECT and demonstrating that attentional selection during dynamic stimuli should be understood as an interaction between the strength of bottom-up influences, such as motion in the stimuli, and top-down influences, such as event model prioritization, to predict where viewers will fixate. If bottom-up effects are very strong, the influence of top-down event models is negligible or cannot be separated from them. Conversely, if top-down influence is dominant, attention can be directed even toward less salient objects. Furthermore, it offers a deeper insight into how filmmakers can guide attention or enable viewers to focus on narrative content themselves. Future research can explore where the thresholds of these criteria lie, which components of the event model are most critical in attentional selection, and how to model attention by predicting potential interactions between viewers' understanding and stimulus features.

References

- Adeli, H., & Zelinsky, G. (2018). Deep-BCN: Deep networks meet biased competition to create a brain-inspired model of attention control. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 1932-1942).
- Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of experimental social psychology*, 74, 187-195.
- Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of verbal learning and verbal behavior*, 17(1), 1-12.
- Bach, M. (2006). The Freiburg Visual Acuity Test-variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 245, 965-971.
- 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.
- Blmeco (version 1.4) (2022, December 12). dispersion_glmer: Measures dispersion in a glmer-model. *RDocumentation*.
https://www.rdocumentation.org/packages/blmeco/versions/1.4/topics/dispersion_glmer
- Bordwell, D., Thompson, K., & Smith, J. (2024). *Film art: An introduction* (Thirteenth Edition). New York: McGraw-Hill.
- Burr, D. C., Morrone, M. C., & Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature*, 371(6497), 511-513.

- Cabanas Gonzalez, C. C. (2024). *Investigating the role of spontaneous theory of mind on the processing of dramatic irony in filmed narratives* (Doctoral dissertation, Birkbeck, University of London).
- Carmi, R. and L. Itti, *Visual causes versus correlates of attentional selection in dynamic scenes*. Vision Research, 2006. 46(26): p. 4333-1761 4345.
- Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of vision*, 9(12), 10-10.
- Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences*, 102(35), 12629-12633.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Routledge.
<https://doi.org/10.4324/9780203771587>
- Conrad, P (2012, November 4). *Michael Haneke: There's no easy way to say this...* The Guardian. <https://www.theguardian.com/film/2012/nov/04/michael-haneke-amour-director-interview>
- Cutting, J. E. (2021). Three filmmaking practices that guide our attention to popular cinema. *Art & Perception*, 10(1), 63-88.
- Davis, E. E., Chemnitz, E., Collins, T. K., Geerligs, L., & Campbell, K. L. (2021). Looking the same, but remembering differently: Preserved eye-movement synchrony with age during movie watching. *Psychology and Aging*, 36(5), 604.

- DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, 17(6-7), 790-811.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1), 193-222.
- Dorr, M., & Bex, P. J. (2013). Peri-saccadic natural vision. *Journal of Neuroscience*, 33(3), 1211-1217.
- Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of vision*, 10(10), 28-28.
- Enns, J. T., & MacDonald, S. C. (2013). The role of clarity and blur in guiding visual attention in photographs. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 568.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*. John Wiley & Sons.
- Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, 37(4), 571-583.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
- Gernsbacher, M. A. (2013). *Language comprehension as structure building*. Psychology Press.
- Haneke, M. (Director). (2006). *Cache* [Film]. France 3 Cinema.
- Hamker, F. H. (2005). The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebral cortex*, 15(4), 431-447.

- Holmqvist, K. and Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms and measures*, Lund, Sweden: Lund Eye-Tracking Research Institute.
- Humphrey, K., & Underwood, G. (2010). The potency of people in pictures: Evidence from sequences of eye fixations. *Journal of Vision*, 10(10), 19-19.
- Hutson, J. P., Chandran, P., Magliano, J. P., Smith, T. J., & Loschky, L. C. (2022). Narrative comprehension guides eye movements in the absence of motion. *Cognitive Science*, 46(5), e13131.
- Hutson, J. P., Magliano, J. P., & Loschky, L. C. (2018). Understanding moment-to-moment processing of visual narratives. *Cognitive science*, 42(8), 2999-3033.
- Hutson, J. P., Smith, T. J., Magliano, J. P., & Loschky, L. C. (2017). What is the role of the film viewer? The effects of narrative comprehension and viewing task on gaze control in film. *Cognitive Research: Principles and Implications*, 2(1), 1-30.
- Huff, M., Meitz, T. G., & Papenmeier, F. (2014). Changes in situation models modulate processes of event perception in audiovisual narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1377.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10-12), 1489-1506.
- Kaakinen, J. K., Hyönä, J., & Viljanen, M. (2011). Influence of a psychological perspective on scene viewing and memory for scenes. *Quarterly Journal of Experimental Psychology*, 64(7), 1372-1387.
- Kim, N. W., Bylinskii, Z., Borkin, M. A., Gajos, K. Z., Oliva, A., Durand, F., & Pfister, H. (2017). BubbleView: an interface for crowdsourcing image importance maps and

- tracking visual attention. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(5), 1-40.
- Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on biomedical engineering*, 57(11), 2635-2645.
- Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior research methods*, 53(6), 2528-2543.
- Kümmerer, M., & Bethge, M. (2023). Predicting visual fixations. *Annual Review of Vision Science*, 9(1), 269-291.
- Kümmerer, M., Bethge, M., & Wallis, T. S. (2022). DeepGaze III: Modeling free-viewing human scanpaths with deep learning. *Journal of Vision*, 22(5), 7-7.
- Kümmerer, M., Wallis, T. S., Gatys, L. A., & Bethge, M. (2017). Understanding low-and high-level contributions to fixation prediction. In *Proceedings of the IEEE international conference on computer vision* (pp. 4789-4798).
- Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., & Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, 100, 316-324.
- Langton, S. R., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, 107(1), 330-342.
- Lenth, R. V., & Piaskowski, J. (2017). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (p. 2.0.1) [Dataset]. <https://doi.org/10.32614/CRAN.package.emmeans>

- Lüdecke, D., Makowski, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., Wiernik, B. M., & Thériault, R. (2019). *performance: Assessment of Regression Models Performance* (p. 0.16.0) [Dataset]. <https://doi.org/10.32614/CRAN.package.performance>
- Linardos, A., Kümmerer, M., Press, O., & Bethge, M. (2021). DeepGaze IIE: Calibrated prediction in and out-of-domain for state-of-the-art saliency modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 12919-12928).
- Loschky, L. C., Hutson, J. P., Smith, M. E., Smith, T. J., & Magliano, J. P. (2018). Viewing static visual narratives through the lens of the scene perception and event comprehension theory (SPECT). In *Empirical Comics Research* (pp. 217-238). Routledge.
- Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would Jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PloS one*, *10*(11), e0142474.
- Loschky, L. C., Larson, A. M., Smith, T. J., & Magliano, J. P. (2020). The scene perception & event comprehension theory (SPECT) applied to visual narratives. *Topics in cognitive science*, *12*(1), 311-351.
- Loschky, L. C., Smith, M. E., Chandran, P., Hutson, J. P., Smith, T. J., & Magliano, J. P. (2026). The role of event understanding in guiding attentional selection in real-world scenes: The scene perception & event comprehension theory (spect). *Attention, Perception, & Psychophysics*, *88*(3), 92.
- Mackendrick, A., & Cronin, P. (2005). *On film-making: An introduction to the craft of the director* (Paperback edition). Faber and Faber, Inc., a division of Farrar, Straus and Giroux, LLC.

- Magliano, J., Kopp, K., McNerney, M. W., Radvansky, G. A., & Zacks, J. M. (2012). Aging and perceived event structure as a function of modality. *Aging, Neuropsychology, and Cognition*, 19(1-2), 264-282.
- MaryT_Intel. New Computer Vision Tool Accelerates Annotation of Digital Images and Video. 2019. Available online: <https://community.intel.com/t5/Blogs/Tech-Innovation/Artificial-Intelligence-AI/New-Computer-Vision-Tool-Accelerates-Annotation-of-Digital/post/1335704#gs.mMeW91jx> (accessed on 27 February 2025).
- Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans. *Journal of personality and social psychology*, 81(5), 922.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2017). *Designing experiments and analyzing data: A model comparison perspective*. Routledge.
- Milisavljevic, A., Abate, F., Le Bras, T., Gosselin, B., Mancas, M., & Doré-Mazars, K. (2021). Similarities and differences between eye and mouse dynamics during web pages exploration. *Frontiers in Psychology*, 12, 554595.
- Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive computation*, 3(1), 5-24.
- Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1243.
- Peterson, J. J. (2018). *The interaction between visual resolution and task-relevance in guiding visual selective attention*. Kansas State University.

- Ptak, R. (2012). The frontoparietal attention network of the human brain: action, saliency, and a priority map of the environment. *The Neuroscientist*, 18(5), 502-515.
- Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science*, 21(4), 269-272.
- Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. Oxford University Press.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372.
- Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., ... Truong, T. (2020). opencv/cvat: v1.1.0 (Version v1.1.0). doi:10.5281/zenodo.4009388
- Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, 6(1), 1-27.
- Smith, T. (2012). The relationship between overt attention and event perception during dynamic social scenes. *Journal of Vision*, 12(9), 407-407.
- Soderbergh, S. (Director). (1979). *Ocean's Thirteen* [Film]. Village Roadshow Pictures.
- Solomon, R. L. (1949). An extension of control group design. *Psychological bulletin*, 46(2), 137.
- Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 31(2), 248.
- Stoll J. (December 02, 2024). Average daily time spent with video content among adults in the United States from 2019 to 2028, by device (in minutes) [Graph]. In *Statista*. Retrieved March 01, 2025, from <https://www.statista.com/statistics/420791/daily-video-content-consumption-usa-device/>
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of vision*, 11(5), 5-5.

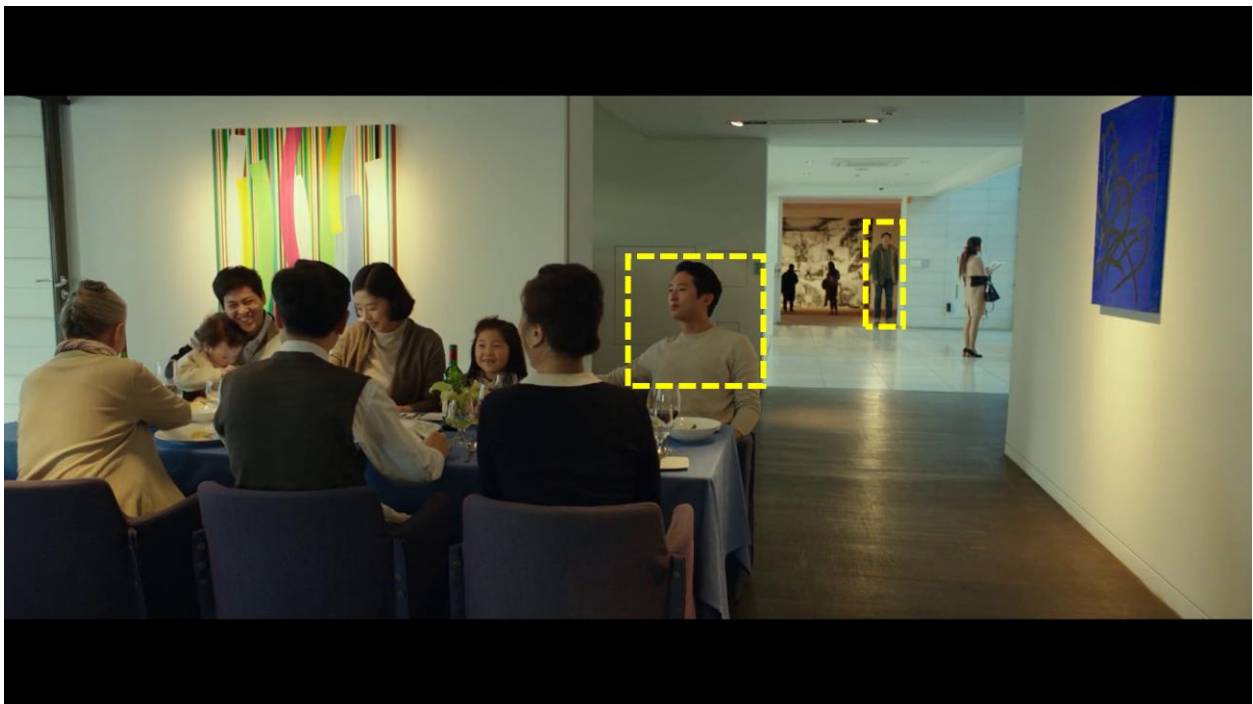
- Taya, S., Windridge, D., & Osman, M. (2012). Looking to score: The dissociation of goal influence on eye movement and meta-attentional allocation in a complex dynamic natural scene. *PLoS One*, 7(6), e39060.
- Team, P. (2022). *RStudio: Integrated Development Environment for R*. Posit Software; 2022.
- Theeuwes, J., Kramer, A. F., Hahn, S., Irwin, D. E., & Zelinsky, G. J. (1999). Influence of attentional capture on oculomotor control. *Journal of experimental psychology: human perception and performance*, 25(6), 1595.
- Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial intelligence*, 78(1-2), 507-545.
- Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131-160). Reading, MA: Addison-wesley.
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological review*, 114(1), 104.
- Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1302.
- Van den Broek, P. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. *Comprehension processes in reading*, 423, 446.
- Wahlheim, C. N., Eisenberg, M. L., Stawarczyk, D., & Zacks, J. M. (2022). Understanding everyday events: Predictive-looking errors drive memory updating. *Psychological Science*, 33(5), 765-781.
- Welles, O. and A. Zugsmith, *Touch of Evil [Film]*. 1958, Universal Pictures.

- Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry*, 2(2), 149-168.
- Williams, C. C., & Castelhana, M. S. (2019). The changing landscape: High-level influences on eye movement guidance in scenes. *vision*, 3(3), 33.
- Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision* (pp. 171-211). Springer, Boston, MA.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological science*, 6(5), 292-297.
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of experimental psychology: Learning, memory, and cognition*, 21(2), 386.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123(2), 162.
- Zelinsky, G. J., & Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Annals of the new York Academy of Sciences*, 1339(1), 154-164.

Appendix A - Videos and their targets of attention

Burning²²

This clip is from *Burning* (Chang-dong, 2018), a South Korean Psychological Thriller. In this clip, there are two Targets of Attention: the “Boyfriend” in the Background and the “New Guy” seated at the dining table. See the still from its Common Viewing Period, where the yellow bounding boxes mark the targets of attention.



You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation [here](#).

²² Not selected for Eye-tracking study but used in Preliminary studies 1,2, & 3

Cache²³

This clip is from *Cache* (Haneke, 2005), a French Thriller. In this clip, there are two Targets of Attention: the Adopted Brother Son, the older boy on the left, and the Main Character Son, the younger boy on the right. See the still from its Common Viewing Period, where the yellow bounding boxes mark the targets of attention.



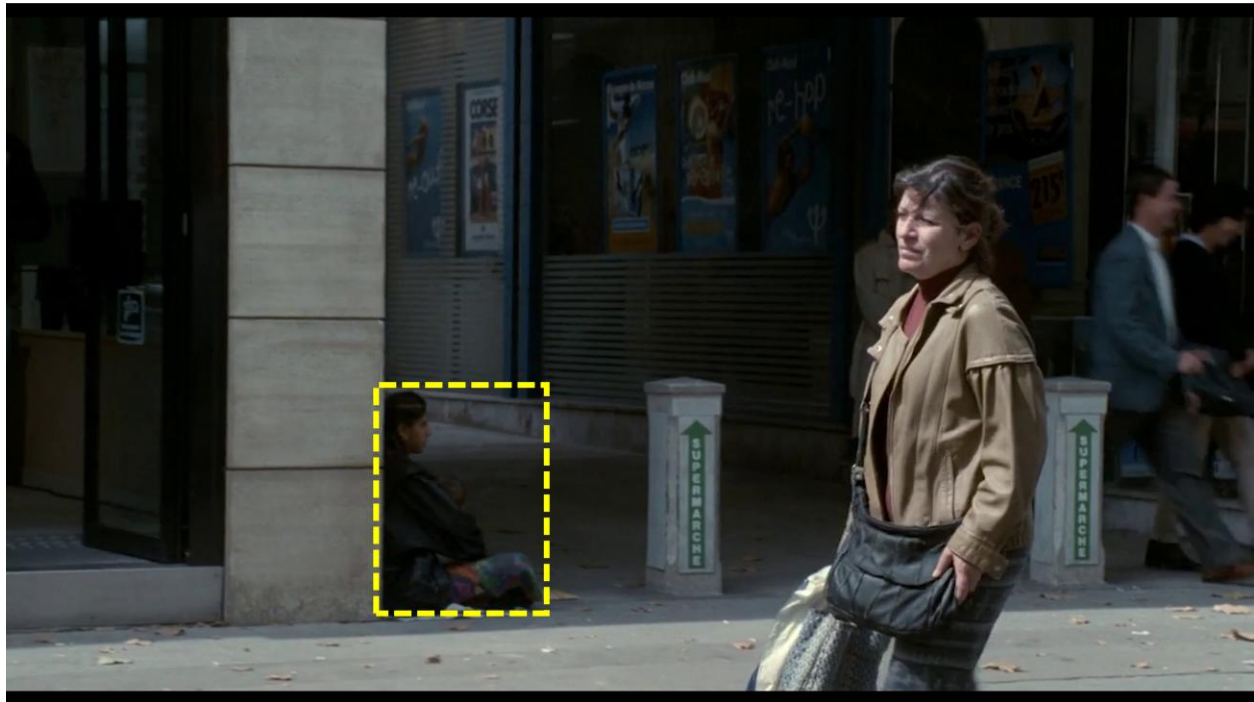
You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation [here](#).

²³ Selected for Eye-Tracking Study

Code Unknown²⁴

This clip is from *Code Unknown* (Haneke, 2000), a French drama. In this clip, there are two Targets of Attention: the first shot is the New Beggar Woman and Kid, and the second shot is the Potential New Begging Spot. The yellow bounding boxes mark the targets of attention.



²⁴ Selected for Eye-Tracking Study but only one target AOI used, *New Beggar Woman and Kid*.



You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation [here](#).

Community²⁵

This clip is from *Community*, Season 3, Episode 7 (2011), an American sitcom. In this clip, there is only one target of attention: the half-open right-side window in the background. The yellow bounding boxes mark the targets of attention.



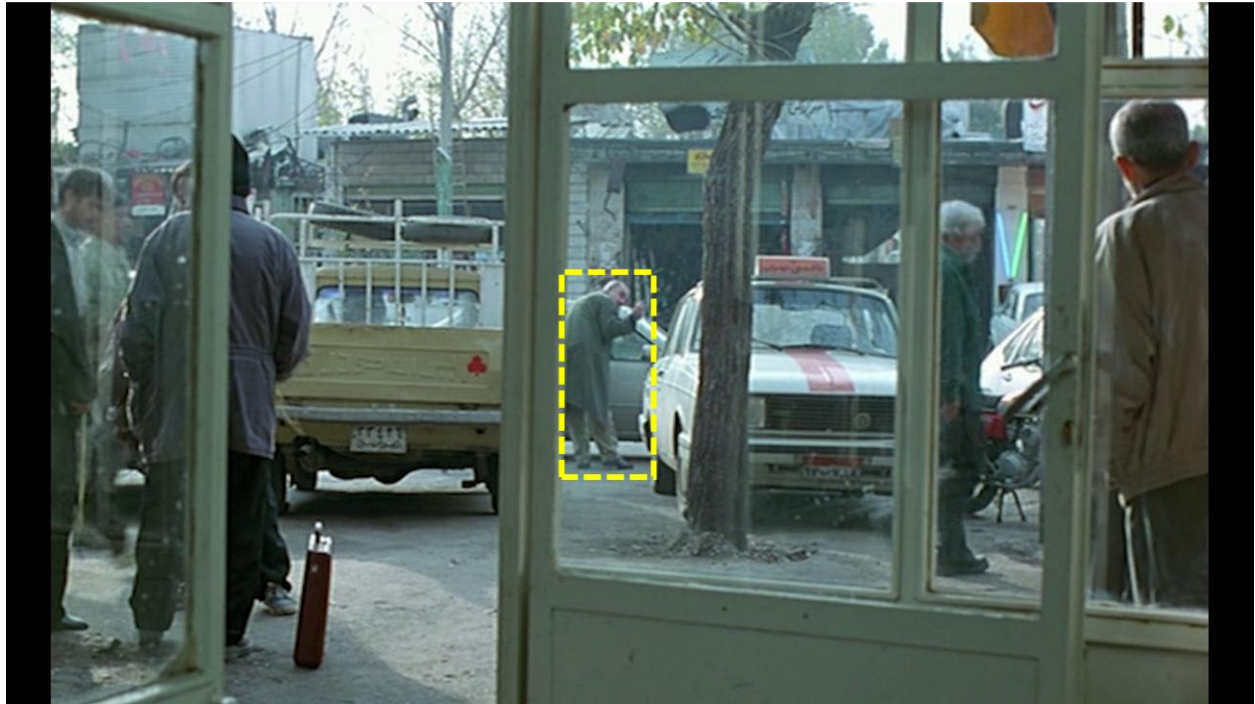
You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation [here](#).

²⁵ Selected for the Eye-Tracking Study.

Crimson Gold Version 1 & Version 2²⁶

This clip is from *Crimson Gold* (2003), an Iranian Crime drama. This video had two versions that varied in when the common viewing period started; version 1 started a little earlier than version 2. In these clips, there is only one Target of Attention: the man in the center talking to the taxi driver. The yellow bounding boxes mark the targets of attention.



You can see the No-context version 1 (i.e., common viewing period) [here](#).

You can see the Context manipulation version 1 [here](#).

You can see the No-context version 2 (i.e., common viewing period) [here](#).

You can see the Context manipulation version 2 [here](#).

²⁶ Both versions, not selected for Preliminary Study 2 & Eye Tracking Study, used in Preliminary Studies 1.

Dogtooth²⁷

This clip is from *Dogtooth* (2009), a Greek absurdist psychological drama. This video has only one target of attention: the car trunk. The yellow bounding boxes mark the targets of attention.



You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation version [here](#).

²⁷ Selected for the Eye-Tracking Study

Drive²⁸

This clip is from *Drive* (2011), an American action drama. This video has only one target of attention: the suspicious car parked in the background. The yellow bounding boxes mark the targets of attention.



You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation version [here](#).

²⁸ Selected for Eye-Tracking Study

Mr. Robot²⁹

This clip is from *Mr. Robot*, Season 3, Episode 5 (2017), an American thriller drama television series. This video has two targets of attention, both of which represent the security personnel standing and discussing in the background. The yellow bounding boxes mark both targets of attention.



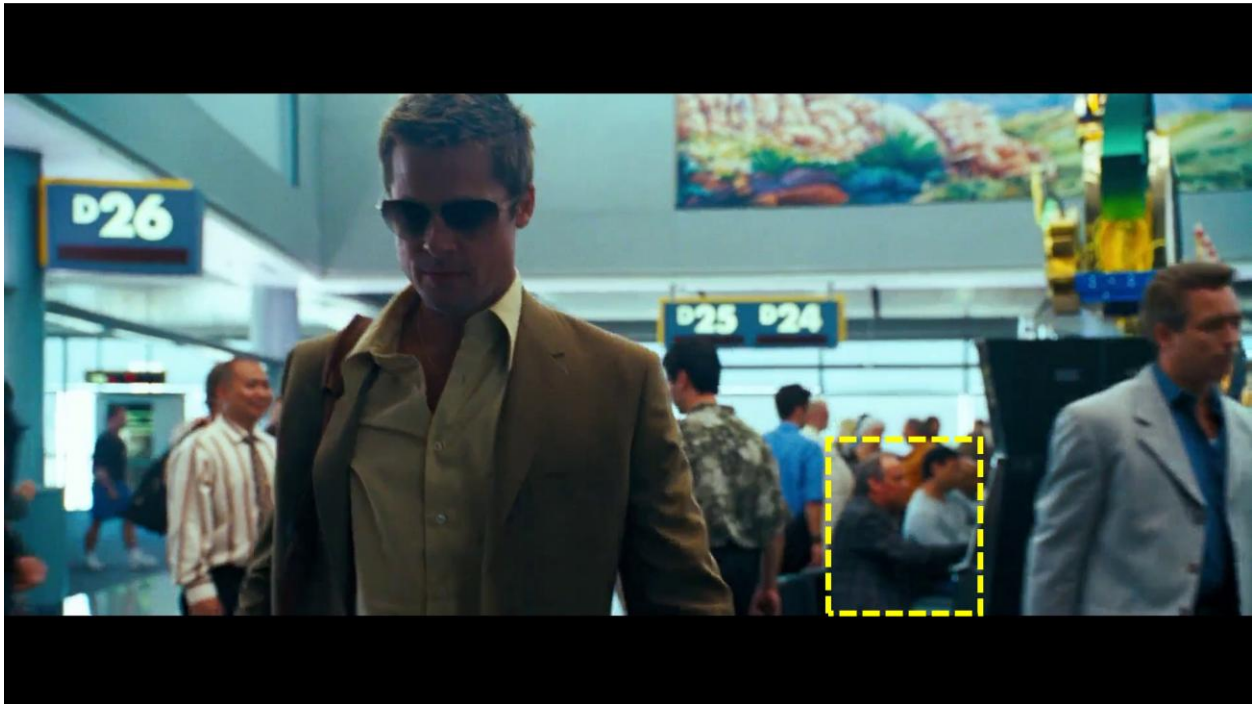
You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation version [here](#).

²⁹ Selected for Eye-Tracking Study

Ocean's Thirteen³⁰

This clip is from *Ocean's Thirteen* (2007), an American heist thriller. This video has only one target of attention: the stranger at the slot machine in the background. The yellow bounding boxes mark the targets of attention.



You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation version [here](#).

³⁰ Selected for Eye-Tracking Study

Revanche³¹

This clip is from *Revanche* (2008), an Austrian thriller. This video has only one target of attention: the photo of the woman on the table. The yellow bounding boxes mark the targets of attention.



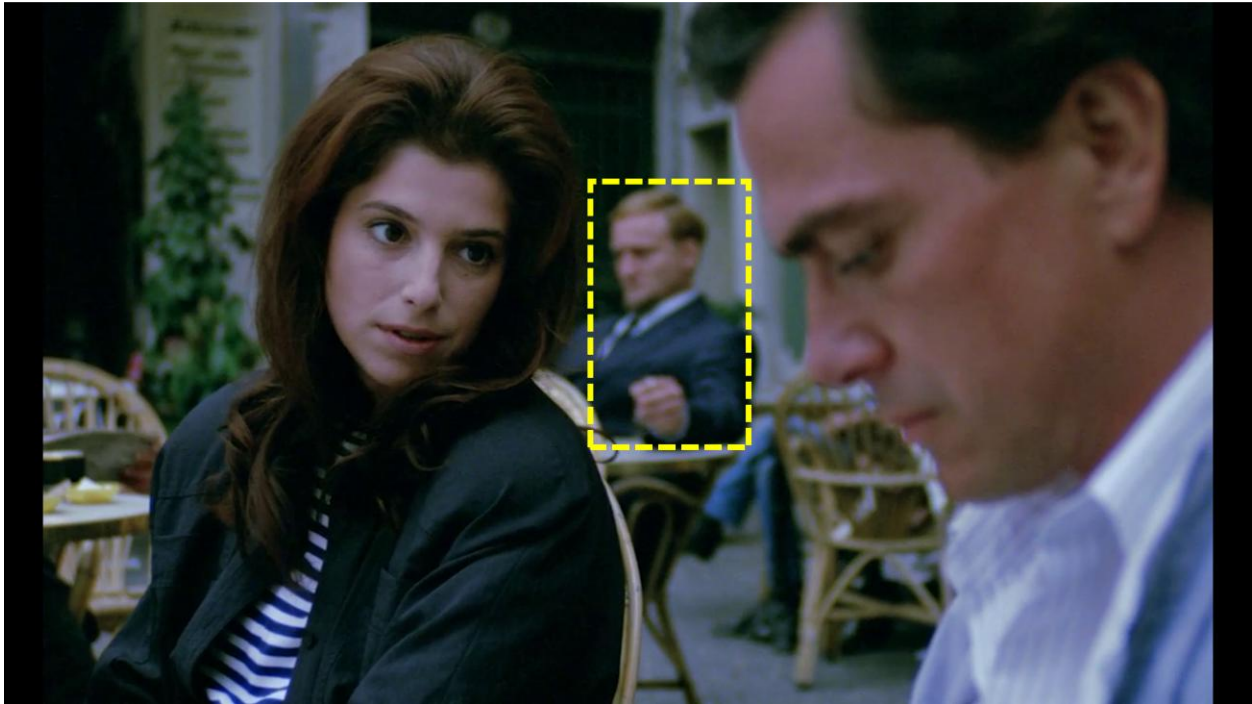
You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation version [here](#).

³¹ Not selected for both the Preliminary Study 2 and Eye-tracking Study. Only used for Preliminary Study 1.

Spoorloos Version 1 & Version 2³²

This clip is from *Spoorloos* (1988), a Danish horror/crime thriller. This video has only one target of attention: the kidnapper sitting in the background. The yellow bounding boxes mark the targets of attention. The only difference between the versions is that the common viewing period ends earlier in version 2 than in version 1.



You can see the No-context version 1 (i.e., common viewing period) [here](#).

You can see the No-context version 2 (i.e., common viewing period) [here](#).

You can see the Context manipulation version 1 & 2 [here](#).

³² Both versions not selected for both the Preliminary Study 2 and Eye-tracking Study. Only used for Preliminary Study 1.

The Fire Within Version 1³³ & Version 2³⁴

This clip is from *The Fire Within* (1963), a French drama. This video has only one target of attention: the gun on the bedside table in the background. The yellow bounding boxes mark the targets of attention. The only difference between the versions is that the common viewing period ends earlier in version 2 than in version 1.



You can see the No-context version 1 (i.e., common viewing period) [here](#).

You can see the No-context version 2 (i.e., common viewing period) [here](#).

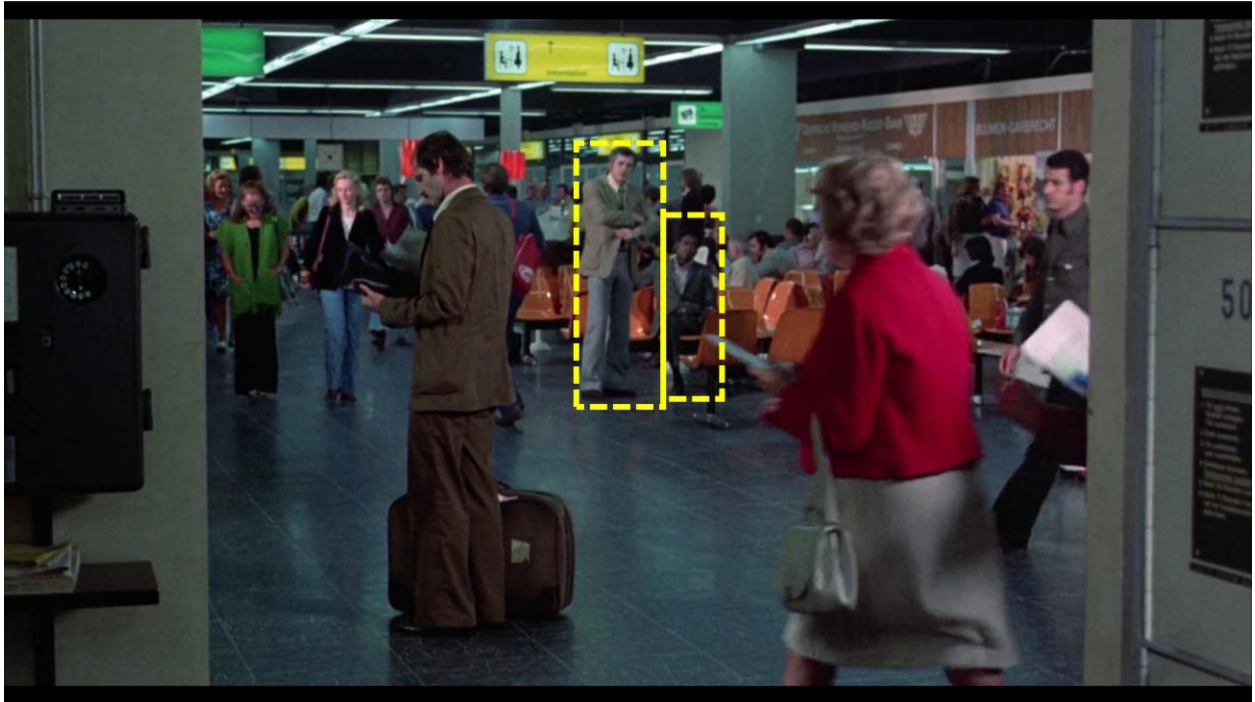
You can see the Context manipulation version 1 & 2 [here](#).

³³ Not selected for both the Preliminary Study 2 and Eye-tracking Study. Only used for Preliminary Study 1.

³⁴ Not selected for Eye-tracking study but used in Preliminary studies 1,2, & 3

The Passenger³⁵

This clip is from *The Passenger* (1975), an Italian drama thriller. This video has two targets of attention: the standing rebel on the left and the sitting rebel on the right. The yellow bounding boxes mark the targets of attention.



You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation version [here](#).

³⁵ Selected for the Eye-tracking Study but only one Target was used, *the sitting rebel*.

The Secret in Their Eyes³⁶

This clip is from The Secret in Their Eyes (2009), an Argentinian crime thriller. This video has one target of attention, the person in the right corner of the screen. The yellow bounding boxes mark the targets of attention.



You can see the No-context version (i.e., common viewing period) [here](#).

You can see the Context manipulation version [here](#).

³⁶ Not selected for both the Preliminary Study 2 and Eye-tracking Study. Only used for Preliminary Study 1.

Appendix B - Schema for Steps in Creating a Context Manipulation

- a. Find a movie clip meeting the two criteria of
 - i. having an animate or inanimate entity, or an innocuous location, but
 - ii. it suddenly becomes very important to the story after a period of time in which it had seemed completely unimportant.
- b. Analyze and list out the set of differential elements that contrast with
 - i. before, when your event model suggested that the entity or location was completely unimportant, and
 - ii. after, when your updated event model made you realize that the entity or location was very important. That is your Target of Attention.
- c. Check if there is an important action or event involving the Target of Attention that could be predictively inferred even if it were not shown. If so, that is your predictive inference.
- d. Write a summary of the minimal narrative that includes all of the differential elements, but stops just before the predictive inference should occur.
- e. Find footage from the video that conveys all, or most, of the differential elements.
 - i. In cases in which you either
 1. cannot find footage that conveys one or more of the differential elements, or
 2. cannot do so without spending too much time,
 3. convey that information as text (i.e., intertitles).
 - a. In such cases, try to be as concise and informative in conveying those missing differential elements as possible.
 - b. Present the intertitles for the amount of time predicted to be necessary to read them, based upon subtitle reading time rules.
- f. Check to make sure that you do not explicitly verbalize the importance of the target of attention in your intertitles. That importance should be determined solely by the differential elements in your updated event model.
- g. Pilot test your pilot No-context condition, and elicit a predictive inference. Then try again after showing your Context manipulation. Do you get large differences in correct versus incorrect predictive inferences? If so, then your manipulation of Event Models may be successful. Otherwise, revise your Context video and intertitles as needed.

Appendix C - Preliminary Experiment 1 Parameter Estimates

Table 7

Parameter Estimates for all six independent Multilevel Poisson Regression, predicting the average rate/expected count of each differential element in viewers' predictions.

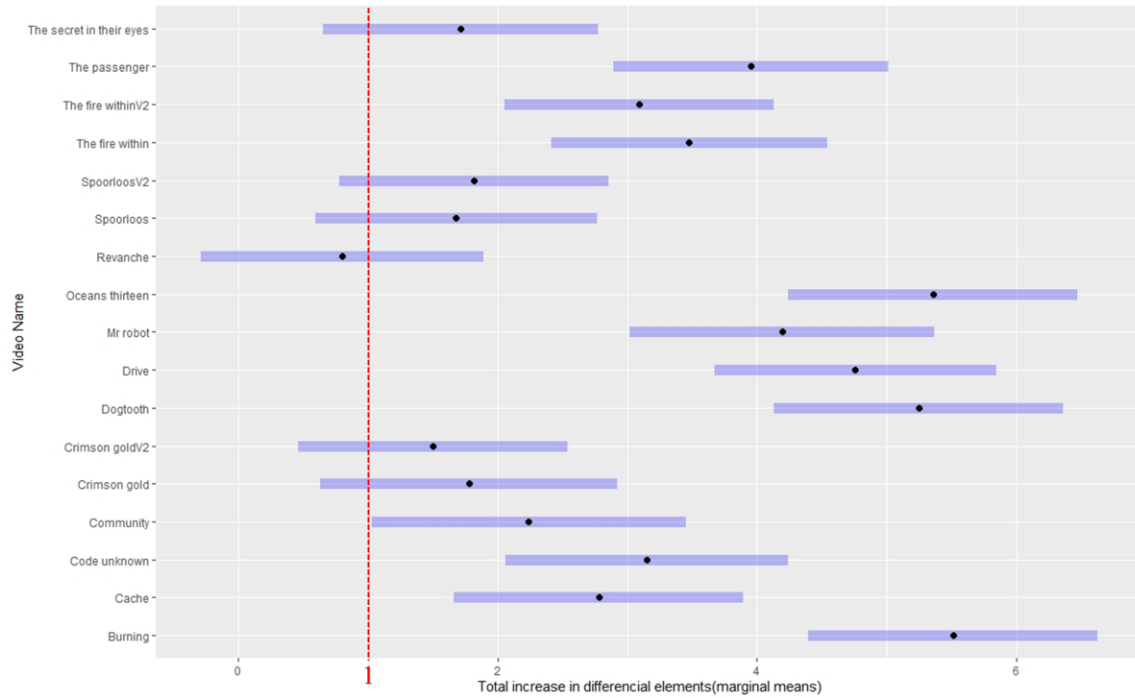
Condition	Variable	B	SE(B)	Z	Significance(p)	Dispersion Ratio	Chisq	p
Entity		-				0.71	476.52	1
	Intercept	1.17	0.17	-7.02	<.001			
	Condition (Context)	0.99	0.08	12.46	<.001			
Goals		-				0.79	502.208	1
	Intercept	1.03	0.11	-8.99	<.001			
	Condition (Context)	1.08	0.08	13.14	<.001			
Enabling Features		-				0.46	284.94	1
	Intercept	3.10	0.47	-6.65	<.001			
	Condition (Context)	1.01	0.15	6.84	<.001			
Temporal Operativity		-				0.63	417.37	1
	Intercept	1.03	0.14	-7.21	<.001			
	Condition (Context)	0.53	0.06	8.19	<.001			
Spatial Proximity		-				0.68	455.782	1
	Intercept	1.85	0.21	-8.78	<.001			
	Condition (Context)	0.79	0.10	8.01	<.001			
Emotions		-				0.54	253.68	1
	Intercept	2.35	0.38	-6.22	<.001			
	Condition (Context)	1.11	0.15	7.43	<.001			

Note: Condition was effect coded (No-context = -1; Context = 1).

Appendix D - Selecting Effective Videos

Figure 20

Model estimated marginal means for the total increase in differential elements for all 17 video clips. The dotted red vertical line represents our conservative statistical criterion, marginal mean greater than 1. We selected only those videos where the lower bound of their 95% Confidence Interval (CI) for the marginal mean was greater than this line.

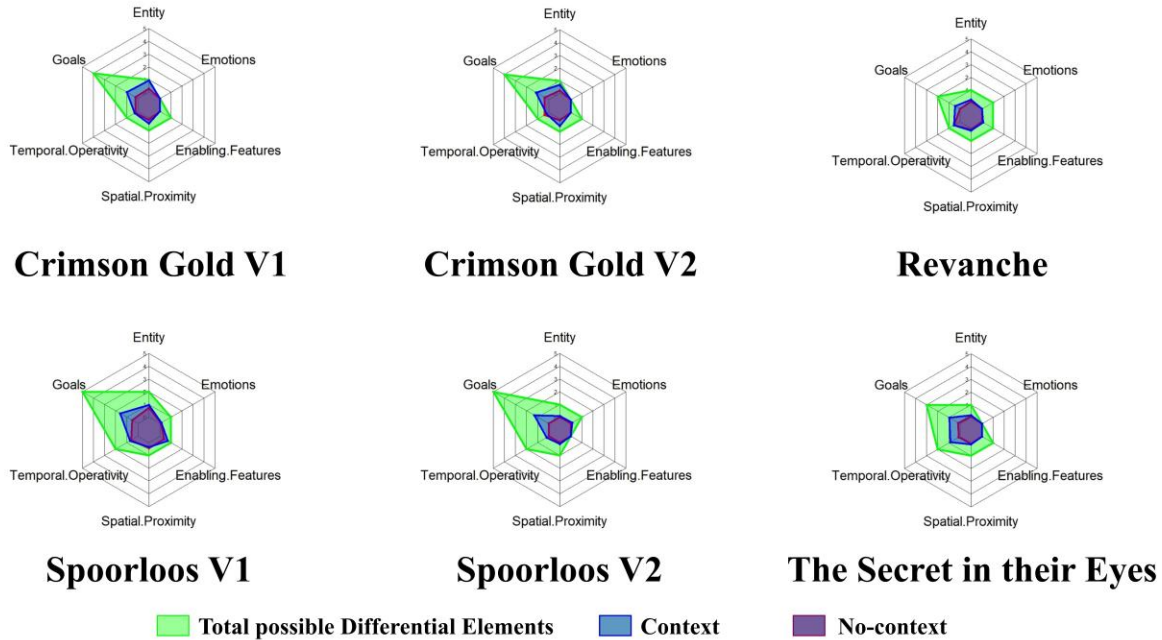


As shown in the figure above, the bars that crossed the line were removed from future studies. Here they are: *Crimson Gold V1*, *Crimson Gold V2*, *Revanche*, *Spoorloos V1*, and *Spoorloos V2*. Both versions of *The Fire Within* were above the red line, and they were very comparable in terms of marginal means; hence, we went with the version where the percentage of people mentioning the target in the No-context was the lowest.

Appendix E - Radar Charts for “ineffective” clips

Figure 21

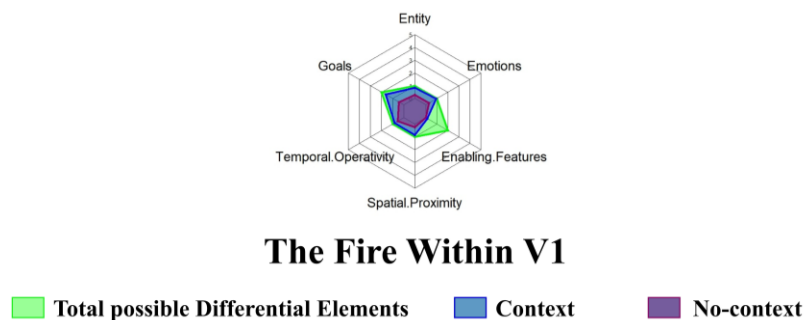
Radar charts depicting the differences in the average number of differential elements represented in viewers’ predictions in Context and No-context conditions for the videos not selected (“ineffective” clips) from Preliminary Experiment 1 for future analyses.



As shown in the radar plots above, there is significant overlap between Context and No-context, suggesting little difference in their event models across all 6 elements.

Figure 22

Radar charts depicting the differences in the average number of differential elements represented in viewers’ predictions in Context and No-context conditions for The Fire Within Version 1 clips from Preliminary Experiment 1.



The Fire Within Version 1 satisfied our selection criteria, and Version 2 also satisfied them, as the lower bound of its 95% Confidence Interval (CI) for the marginal mean was greater than 1. We ended up not selecting this version only because the number of differential elements was higher in the No-context condition; as you can see, the maroon region is larger than the smallest hexagon in the plot.

Appendix F - Video Specifications

Table 8

Video specifications for the 10 videos used for verification in Preliminary Studies 2 & 3.

Video	Native Aspect Ratio	Common Viewing Period Duration (s)	Context Manipulation Duration (s)	# of Target AOIs	Common Viewing Period # of Frames
Burning	2.35 : 1	19	498	2	466
Cache	16 : 9	86	620	2	2080
Code Unknown	1.85 : 1	136	254	2	3275
Community	16 : 9	15	46	1	361
Dogtooth	2.35 : 1	18	196	1	452
Drive	2.35 : 1	36	139	1	864
Mr Robot	16 : 9	34	477	2	822
Oceans Thirteen	2.39 : 1	10	140	1	243
The Fire Within	1.66 : 1	32	388	1	441
The Passenger	1.85 : 1	61	441	2	1465

Note: The shaded videos were used in the final Eye-tracking study.

Appendix G - Complex Random Intercept Selection and Nested Regression for Mouse Hit Probability Models.

Table 9

Model comparisons to find the best random-effect structure for a multilevel logistic regression, where Condition, Frame order, and their interaction predict Mouse hit probability to the target, based on Preliminary Study 2.

Model Label	Model random effect structure description	AIC	BIC	Convergence?	Singular?
Model 1	(1 + Frame order Participant) + (1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	307285.5	307466.9	No	Yes
Model 2	(1 Participant) + (1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	307283.5	307442.2	No	Yes
Model 3	(1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	307281.5	307428.9	Yes	No

Note: Following the approach proposed by Scandola & Tidoni (2024), the model selection process started with a full Complex Random Intercept (CRI) model (Model 1) and iteratively removed the CRI with the lowest variance until a nonsingular fit was achieved. Here, Model 3 was nonsingular and converged successfully, and it was used as the final random-effect structure for logistic regression models predicting Mouse hit probability at the target.

Table 10

Nested model comparison for hypothesis testing and finding the best model fit for Mouse hit probability using multilevel logistic regressions, based on Preliminary Study 2.

Label	Model Description	df	AIC	BIC	logLik	Chisq	Ddf	p
Null Model	1 + (1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	10	307296	307409	153638	-		
Event Model	Condition + (1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	11	307280	307405	153629	17.14	1	<.001
Event + Frame Order Model	Condition + Frame order + (1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	12	307282	307418	153629	0.56	1	.456
Full Model	Condition*Frame order + (1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	13	307281	307429	153628	2.34	1	.126

Note: The random effect structure is the same across all models: it includes a basic random

intercept for the grouping factor of AOI, as well as complex random intercepts representing the residual variance of AOI and Participant interacting with Condition (Context vs No-context).

Finally, Frame Order was included as a random slope attached to each of these grouping terms.

Appendix H - Graphs for Frame Order and Gaze Probability

Figure 23

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for the Boyfriend AOI in Burning common viewing period along with the baseline distribution measures.

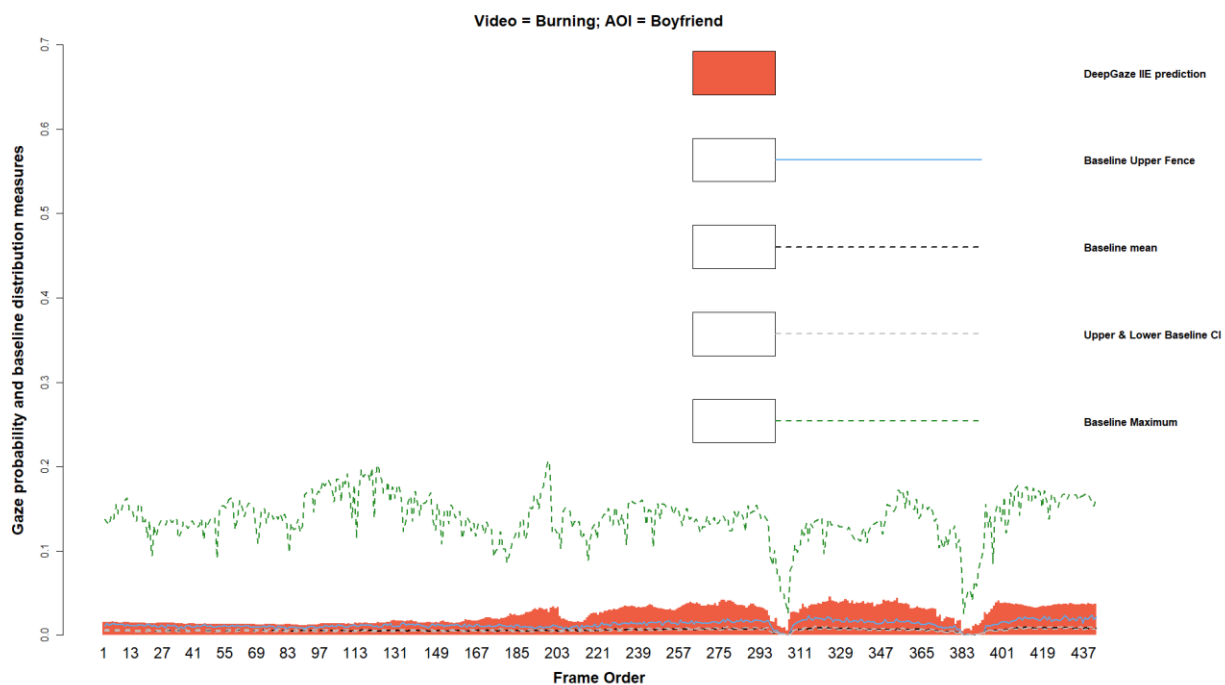


Figure 24

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for the New Guy AOI in Burning common viewing period along with the baseline distribution measures.

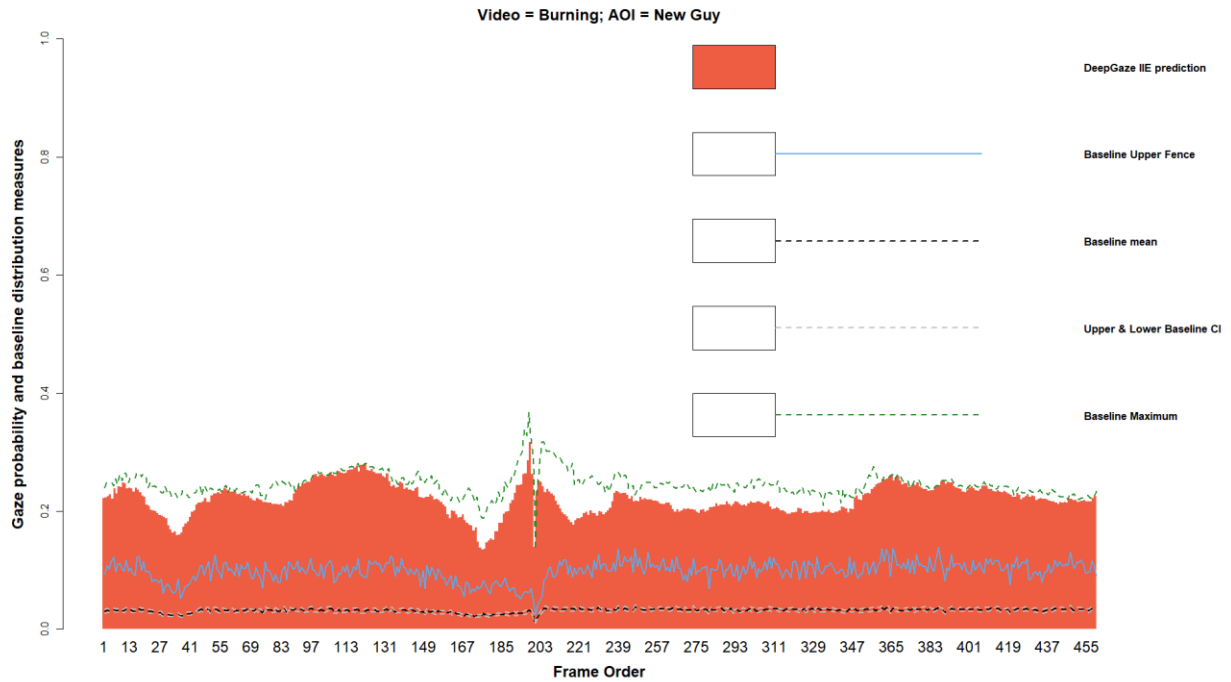


Figure 25

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for the Adopted Brother Son AOI in Cache common viewing period along with the baseline distribution measures.

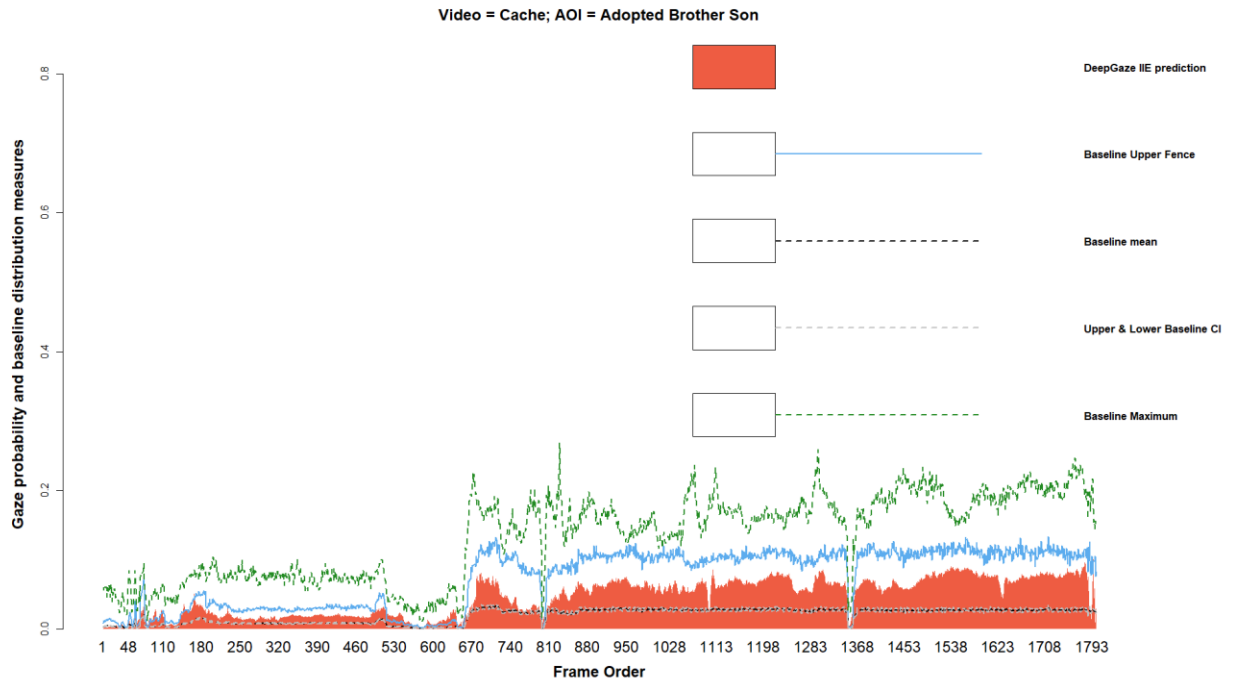


Figure 26

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for the Main Character Son AOI in Cache common viewing period along with the baseline distribution measures.

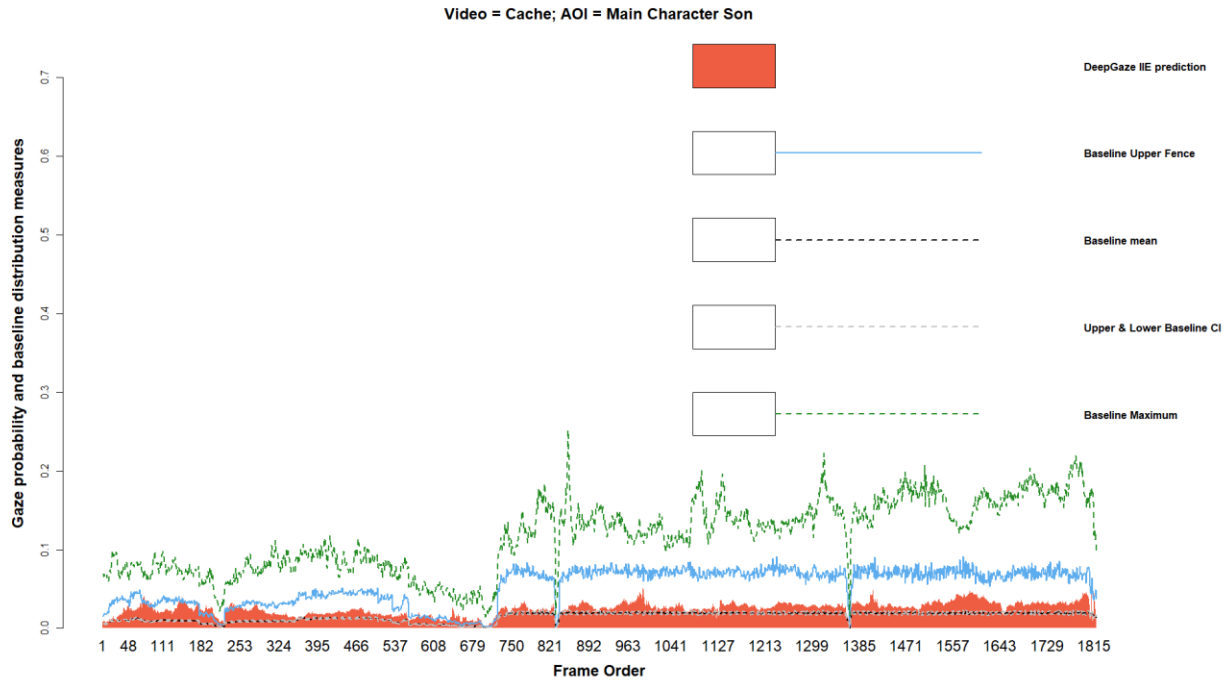


Figure 27

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for the New Beggar Woman and Kid AOI in Code Unknown common viewing period along with the baseline distribution measures.

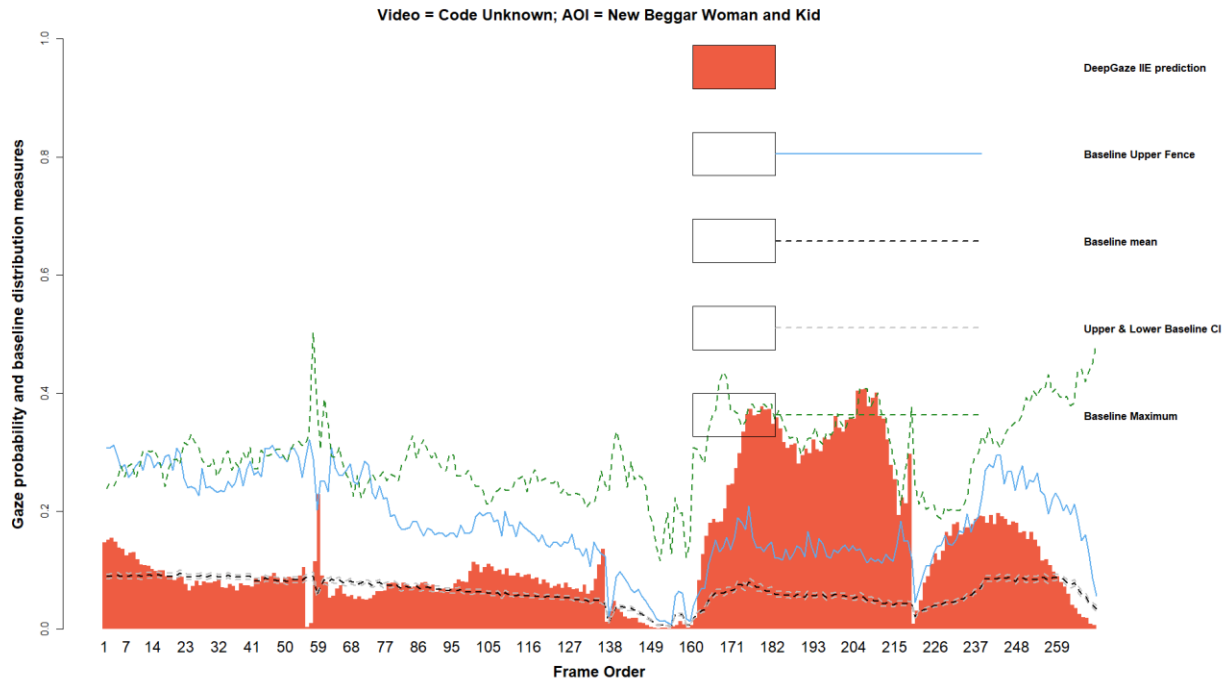


Figure 28

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for Potential New Begging Spot AOI in Code Unknown common viewing period along with the baseline distribution measures.

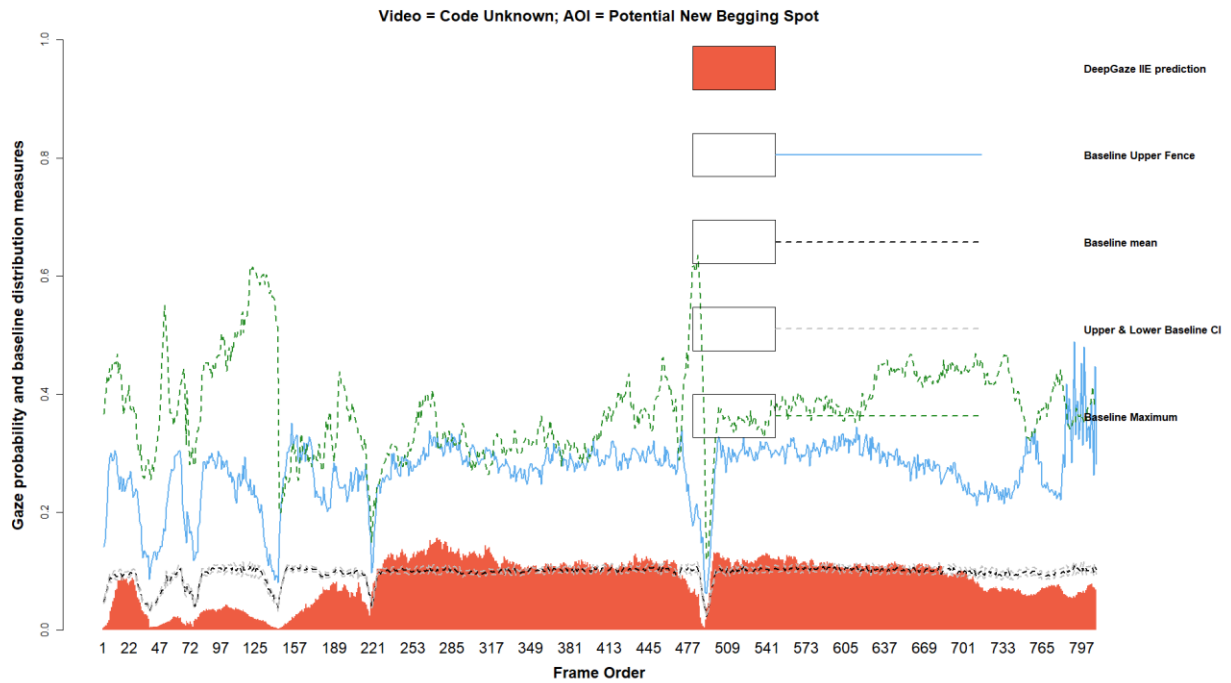


Figure 29

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for Window AOI in Community common viewing period along with the baseline distribution measures.

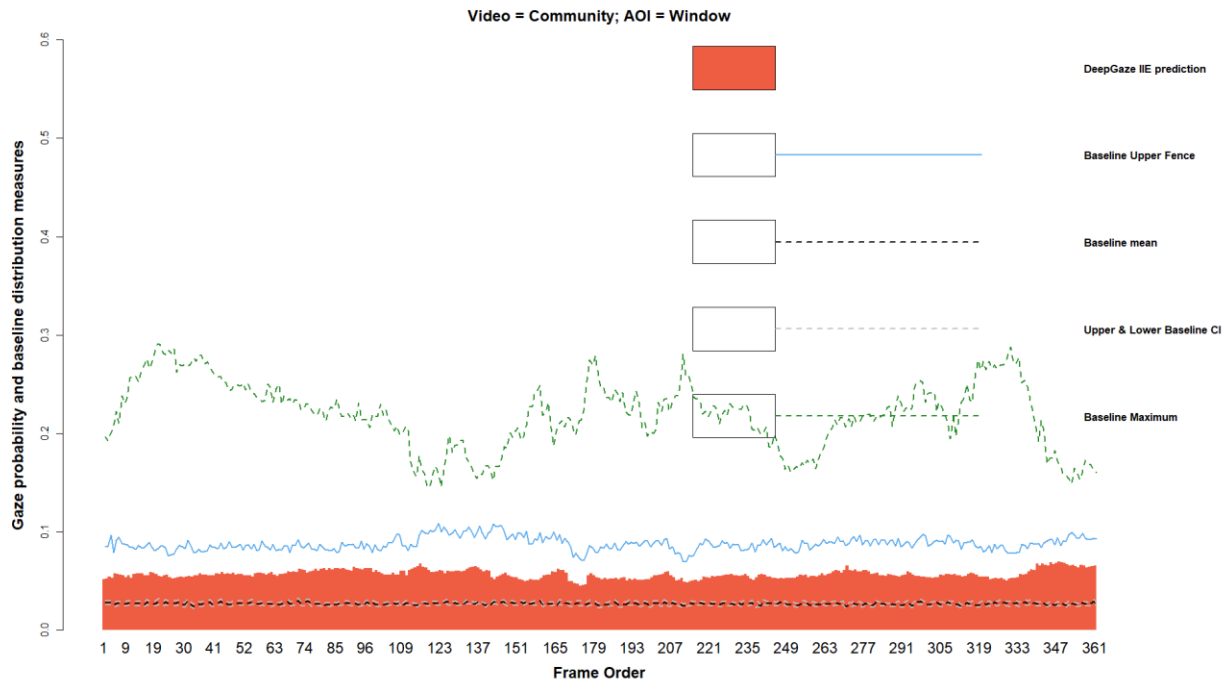


Figure 30

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for Car Trunk AOI in Dogtooth common viewing period along with the baseline distribution measures.

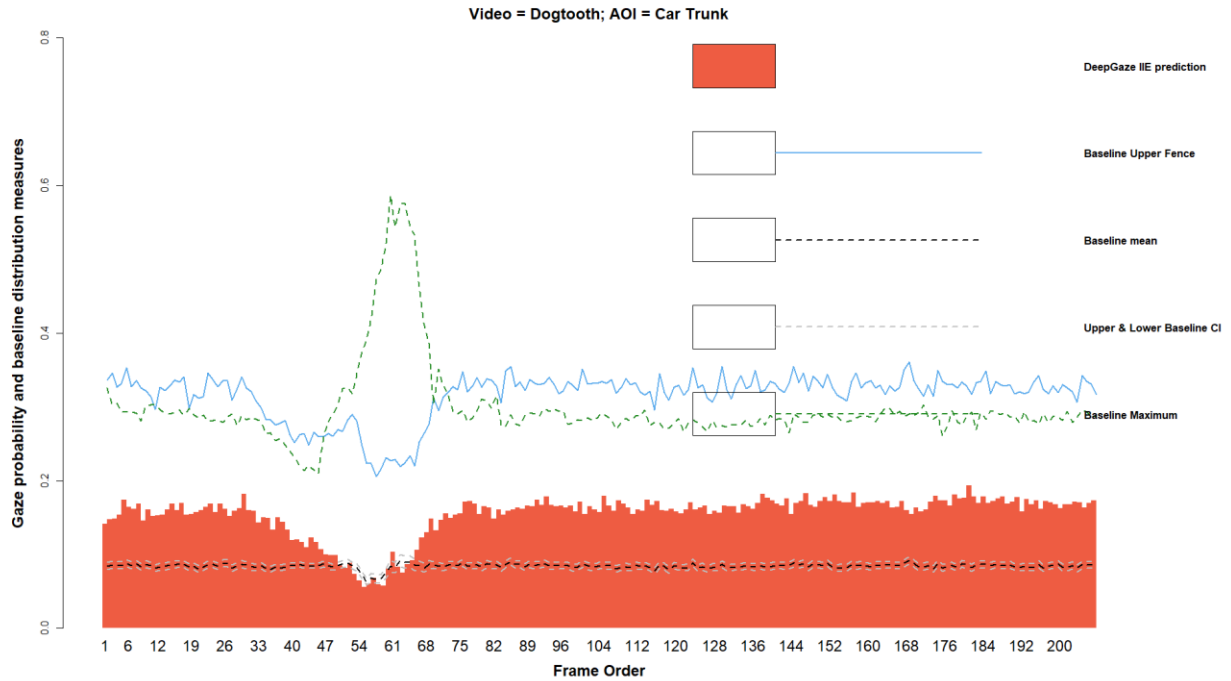


Figure 31

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for Suspicious Car AOI in Drive common viewing period along with the baseline distribution measures.

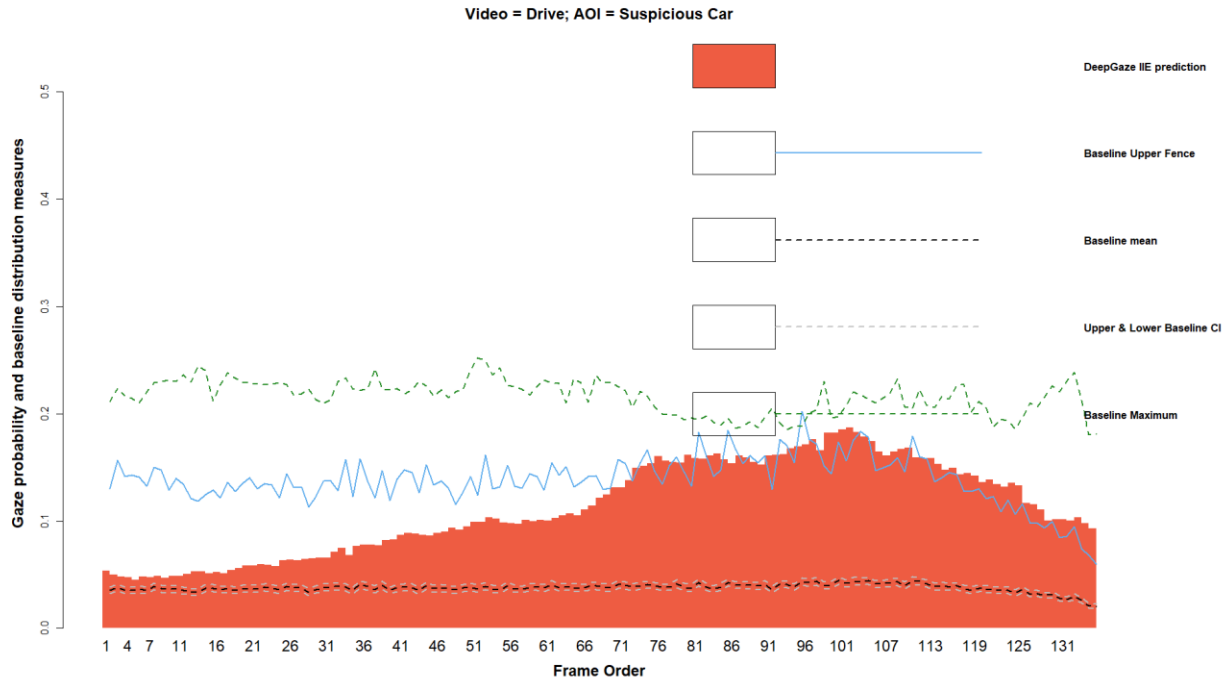


Figure 32

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for Security Personnel AOI in Mr. Robot common viewing period along with the baseline distribution measures.

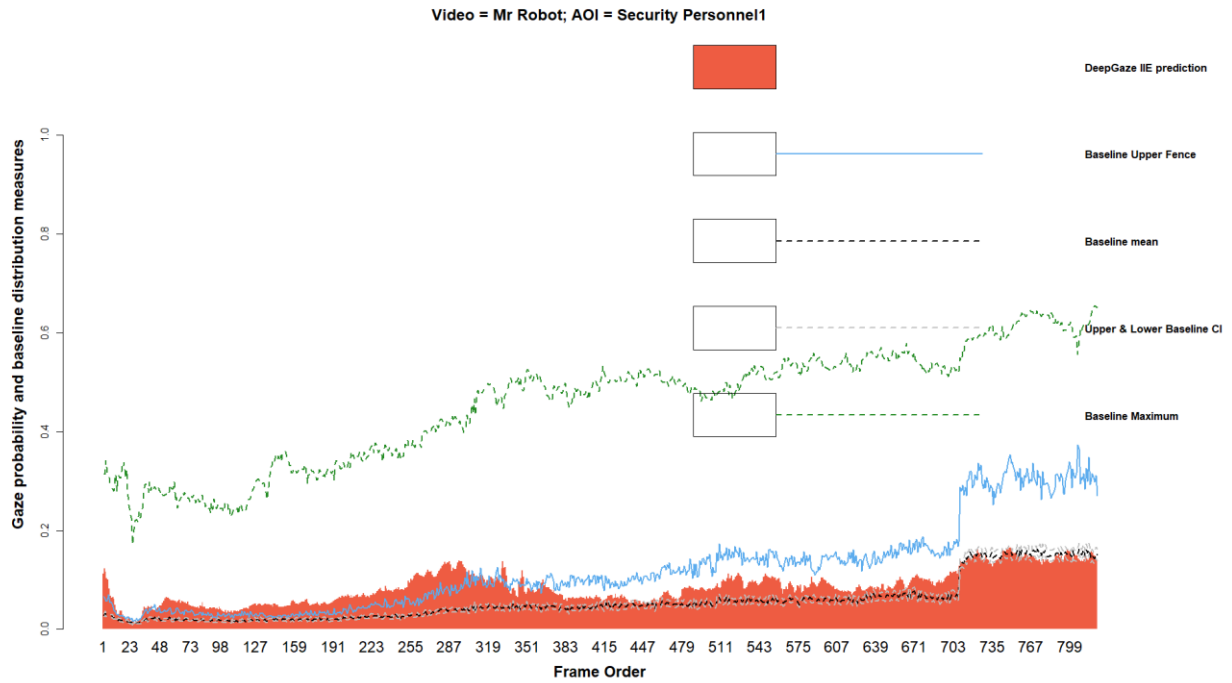


Figure 33

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for Security Personnel2 AOI in Mr. Robot common viewing period along with the baseline distribution measures.

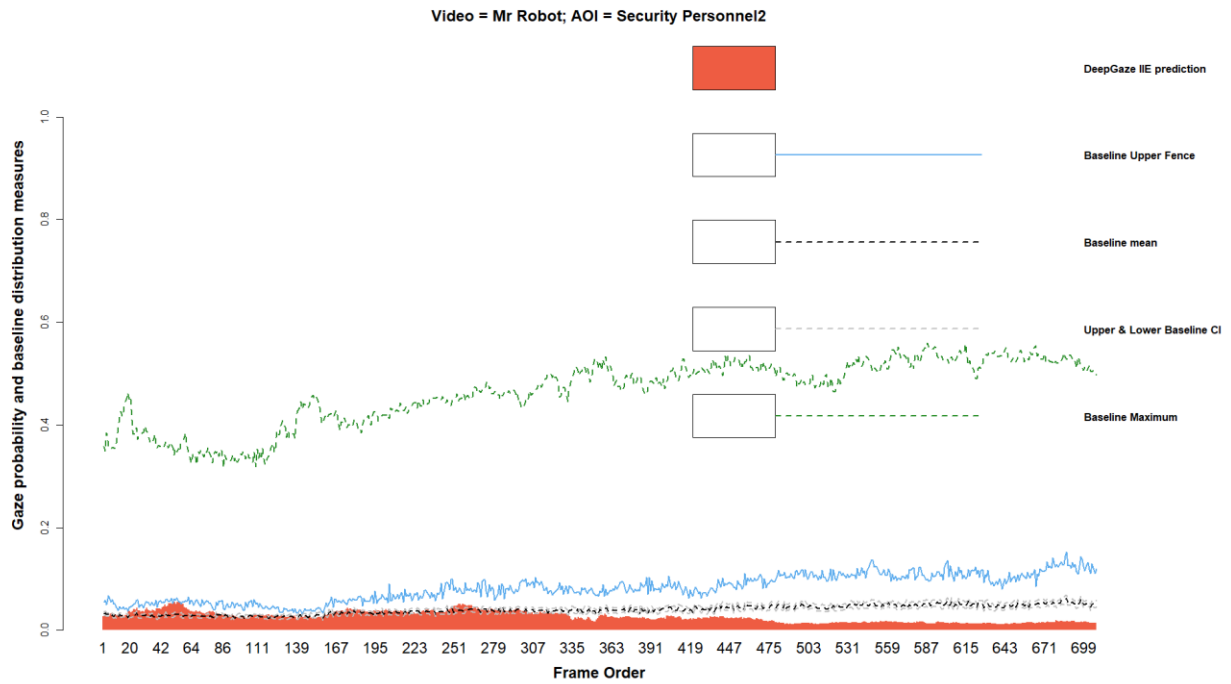


Figure 34

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for Gun AOI in The Fire Within common viewing period along with the baseline distribution measures.

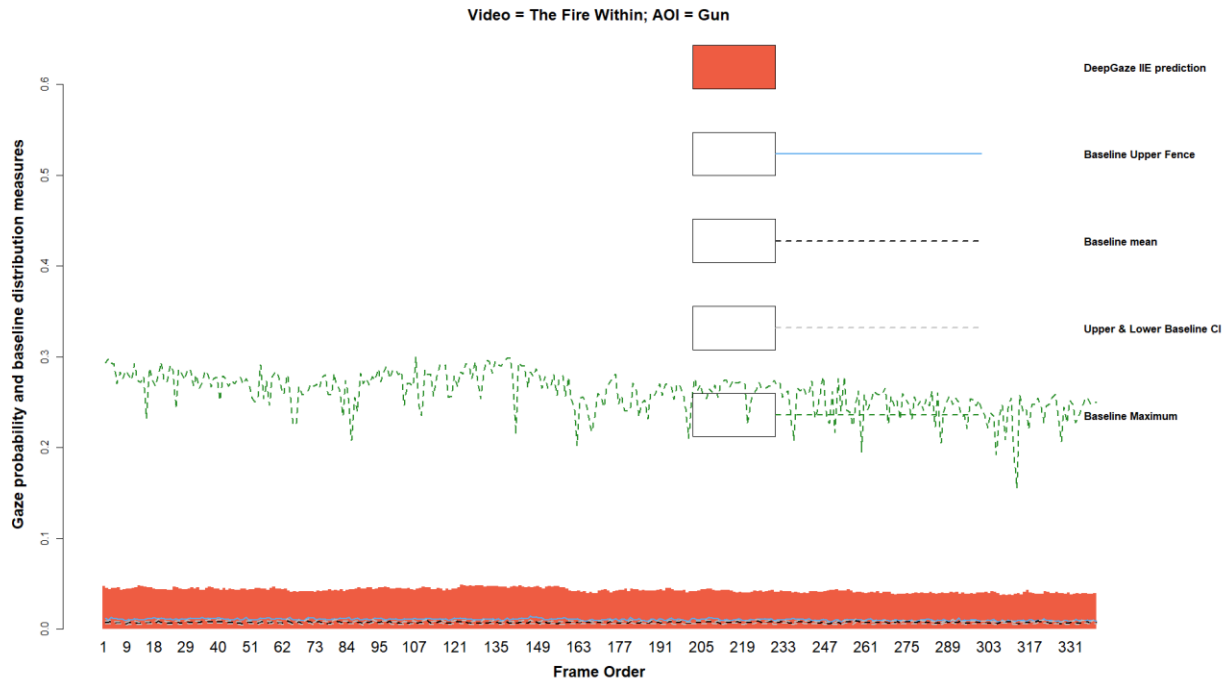


Figure 35

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for Rebel Sitting AOI in The Passenger common viewing period along with the baseline distribution measures.

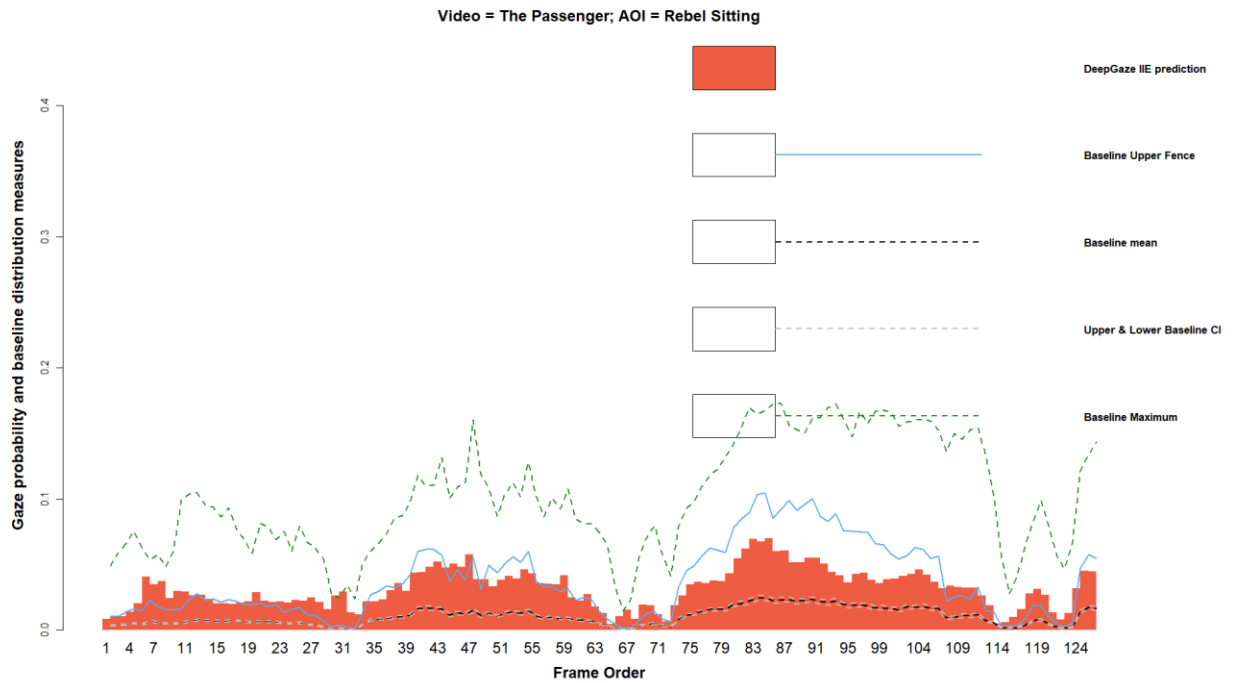
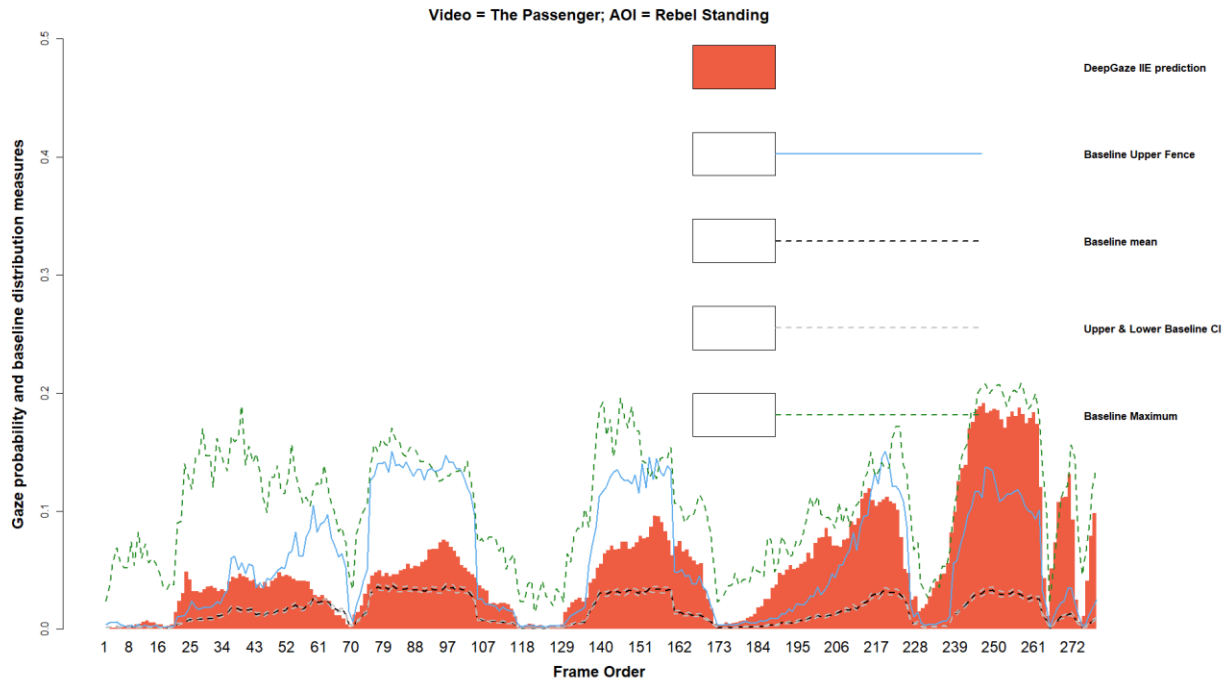


Figure 36

Frame by Frame Default Gaze Probability Prediction by DeepGaze IIE for Rebel Standing AOI in The Passenger common viewing period along with the baseline distribution measures.



Appendix I - Power Analysis Effect Size Comparisons

Table 11

Cohen's d for repeated measures was calculated separately for each Target AOI in Preliminary Study 2.

Video	Target Area of Interest	No Context		Context		<i>df</i>	<i>t</i>	<i>p</i>	<i>r</i>	<i>d_{rm}</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
Cache	Adopted Brother Son	0.004	0.01	0.56	0.24	29	-12.60	<.001	.18	2.99
Cache	Main Character Son	0.005	0.01	0.18	0.14	29	-6.70	<.001	.17	1.60
Code Unknown	New Beggar Woman and Kid	0.07	0.10	0.17	0.12	29	-4.08	<.001	.12	0.97
Community	Window	0.00	0.00	0.20	0.20	29	-5.38	<.001	.00	1.47
Dogtooth	Car Trunk	0.01	0.02	0.36	0.24	27	-7.67	<.001	.04	2.02
Drive	Suspicious Car	0.03	0.10	0.20	0.19	29	-4.33	<.001	.03	1.13
Mr Robot	Security Personnel1	0.01	0.04	0.09	0.15	28	-2.74	.011	.04	0.72
Mr Robot	Security Personnel2	0.02	0.02	0.07	0.11	28	-2.19	.037	.10	0.57
Oceans Thirteen	Stranger at the Slot Machine	0.005	0.02	0.36	0.22	28	-8.66	<.001	.01	2.27
The Passenger	Rebel Sitting	0.01	0.04	0.18	0.14	29	-6.46	<.001	.12	1.57

Appendix J - Video set order

If there is a carry-over effect from the previous video they watched (either in a context or a No-context block) on their prediction and attention on the current video, then we want to balance those carry-over effects. They are balanced using the Williams design, a generalized Latin square (Williams, 1949). Hence, we are counterbalancing the effect of the order in which the videos are presented to the participants. However, this method controls only for first-order carry-over, i.e., all videos are preceded by one of the other 7 videos or none (i.e., it is the first video in the order) at least once, as shown in the Latin square below. Moreover, this method ensures that all the videos are presented in. For example, if you notice the Latin square below, the letter A representing a video is presented in all 8 positions across the 8 video sets.

Here is the suggested 8x8 Williams Latin Square, where each row corresponds to one of the 8 video sets, and each column corresponds to the position in the order in which videos will be presented in a block. And each cell represents the video to be presented at that position.

Video Set A: 1 2 8 3 7 4 6 5

Video Set B: 2 3 1 4 8 5 7 6

Video Set C: 3 4 2 5 1 6 8 7

Video Set D: 4 5 3 6 2 7 1 8

Video Set E: 5 6 4 7 3 8 2 1

Video Set F: 6 7 5 8 4 1 3 2

Video Set G: 7 8 6 1 5 2 4 3

Video Set H: 8 1 7 2 6 3 5 4

Each alphabet represents a video clip, and here is the mapping of the letters and videos, which is currently just assigned based on their alphabetical order:

- 1 – Cache
- 2 – Code Unknown
- 3 – Community
- 4 – Dogtooth
- 5 – Drive
- 6 – Mr. Robot
- 7 – Oceans Thirteen
- 8 – The Passenger

Each of these video sets was pseudo-randomly assigned to the participants along with the 4 groups. Hence, the 40 participants in a group gave at least 5 participants each in one of the video sets listed above.

Appendix K - Selecting Random Effect Structure for Fixation Hit Probability

Table 12

Model comparisons to identify the best random-effect structure for a multilevel logistic regression, where Condition, Frame order, and their interaction predict the target fixation hit probability, based on the Eye-tracking Study.

Model Label	Model random effect structure description	AIC	BIC	Convergence?	Singular?
Model 1	(1 + Frame Order Participant) + (1 + Frame Order Participant:Condition) + (1 + Frame Order AOI) + (1 + Frame Order AOI:Condition)	168684.4	168860.7	Yes	Yes
Model 2	(1 Participant) + (1 + Frame Order Participant:Condition) + (1 + Frame Order AOI) + (1 + Frame Order AOI:Condition)	168682.4	168836.7	Yes	No

Note: Following the approach proposed by Scandola & Tidoni (2024), the model selection process started with a full Complex Random Intercept (CRI) model (Model 1) and iteratively removed the CRI with the lowest variance until a nonsingular fit was achieved. Here, Model 2 was nonsingular and converged successfully, and it was used as the final random-effect structure for logistic regression models predicting the target fixation hit probability.

Appendix L - Nested Model Comparison for Hypothesis Testing for Fixation Hit Probability

Table 13

Nested model comparison for hypothesis testing and finding the best model fit for Fixation Hit Probability using multilevel logistic regressions, based on the Eye-tracking Study.

Label	Model Description	df	AIC	BIC	logLik	χ^2	Δdf	p
Null Model	1 + (1 Participant) + (1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	11	168689	168810	-84334			
Event Model	Condition + (1 Participant) + (1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	12	168680	168812	-84328	11.12	1	<.001
Event + Frame Order Model	Condition + Frame order (1 Participant) + (1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	13	168684	168827	-84329	1.81	1	.179
Full Model	Condition*Frame order (1 Participant) + (1 + Frame order Participant:Condition) + (1 + Frame order AOI) + (1 + Frame order AOI:Condition)	14	168682	168837	-84327	3.37	1	.067

Note: The random effect structure is the same across all models: included basic random

intercepts for my grouping factors (AOI and Participant), as well as complex grouping terms representing the residual variance of AOI and Participant interacting with Condition. Finally, the continuous variable of Frame Order was included as a random slope attached to the AOI, AOI:Condition, and Participant:Condition grouping terms.

Appendix M - Frame-by-frame changes in the raw average probability values of fixation hit probability and the two a priori predictions.

Figure 37

Frame by Frame changes in the raw average probability values of fixation hit probability and the two a priori predictions (Default Gaze Prediction and Importance Prediction) of the Adopted Brother Son AOI in the Cache clip for both Context and No-context viewing.

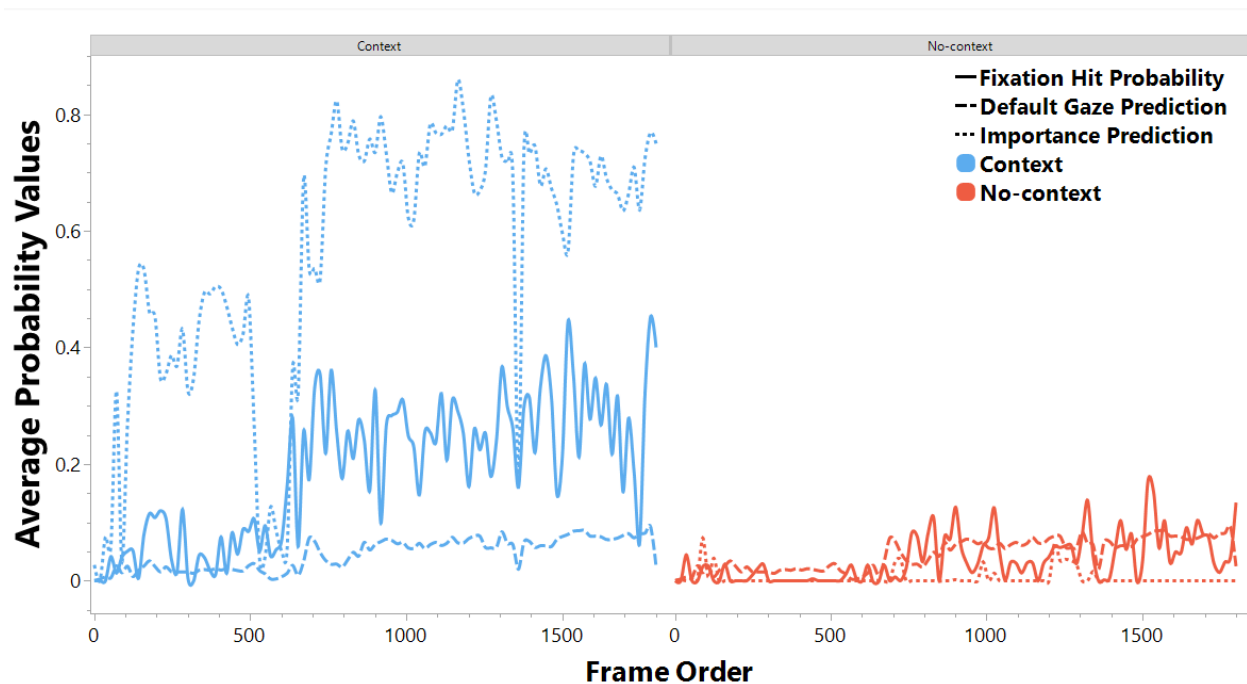


Figure 38

Frame by Frame changes in the raw average probability values of fixation hit probability and the two a priori predictions (Default Gaze Prediction and Importance Prediction) of the Main Character Son AOI in the Cache clip for both Context and No-context viewing.

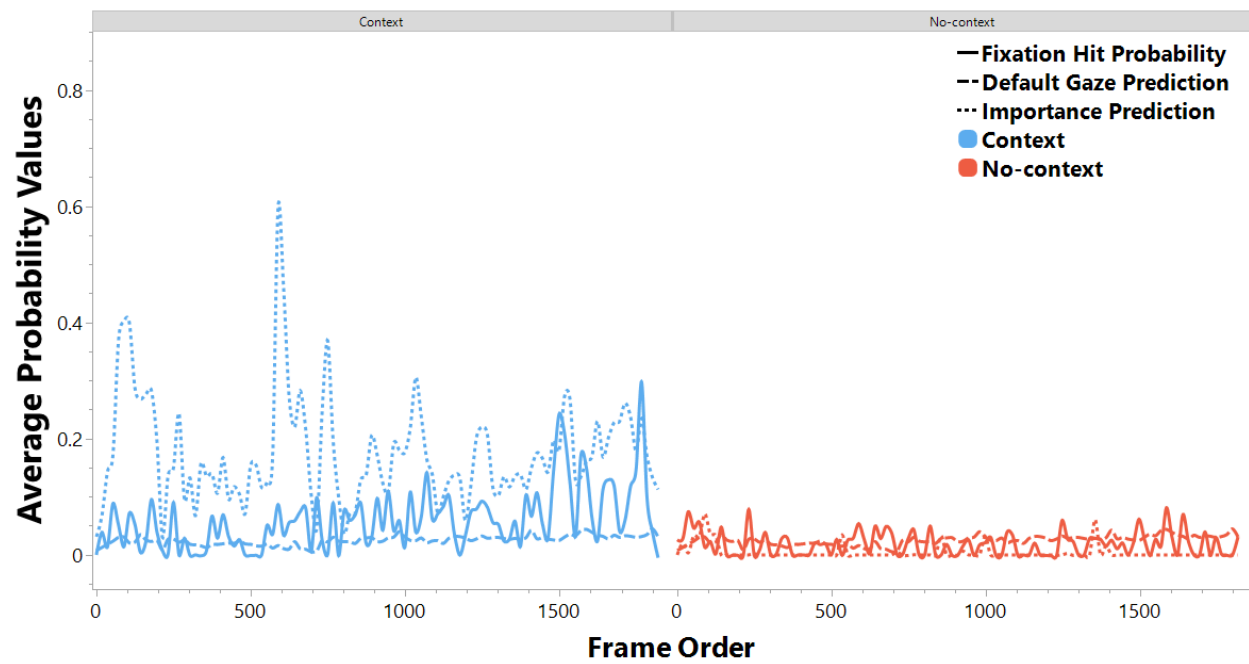


Figure 39

Frame by Frame changes in the raw average probability values of fixation hit probability and the two a priori predictions (Default Gaze Prediction and Importance Prediction) of the New Beggar Woman and Kid AOI in the Code Unknown clip for both Context and No-context viewing.

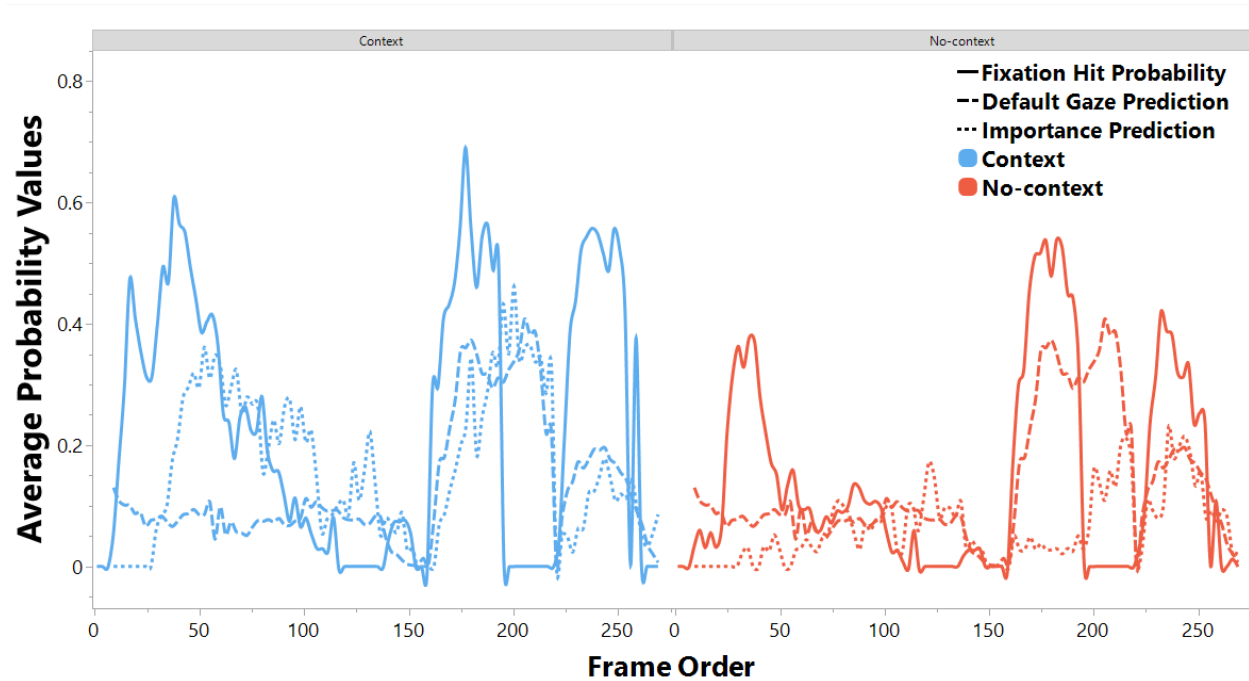


Figure 40

Frame by Frame changes in the raw average probability values of fixation hit probability and the two a priori predictions (Default Gaze Prediction and Importance Prediction) of the Window AOI in the Community clip for both Context and No-context viewing.

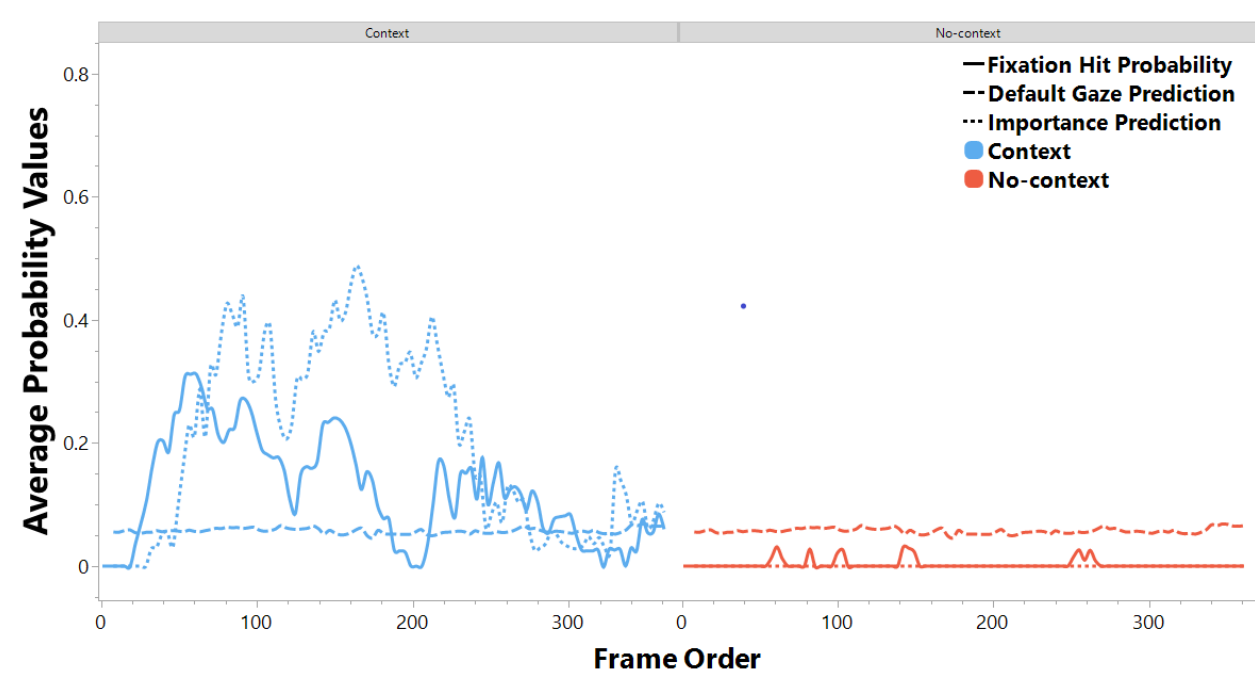


Figure 41

Frame by Frame changes in the raw average probability values of fixation hit probability and the two a priori predictions (Default Gaze Prediction and Importance Prediction) of the Car Trunk AOI in the Dogtooth clip for both Context and No-context viewing.

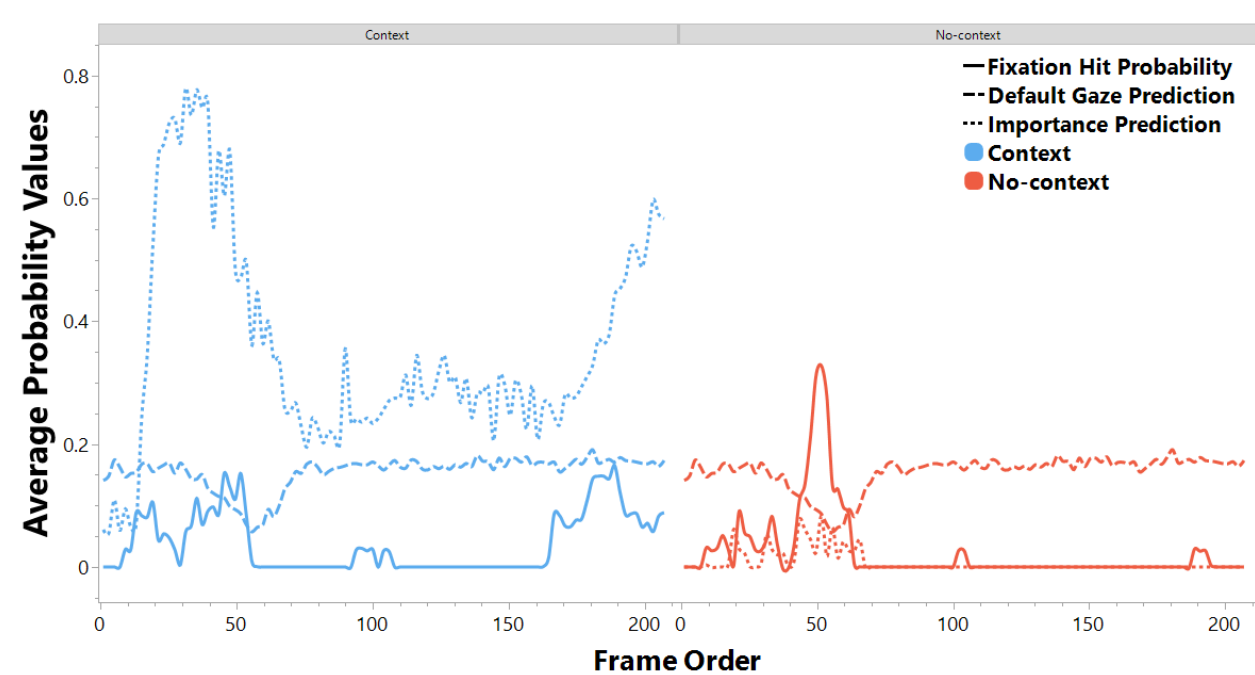


Figure 42

Frame by Frame changes in the raw average probability values of fixation hit probability and the two a priori predictions (Default Gaze Prediction and Importance Prediction) of the Suspicious Car AOI in the Drive clip for both Context and No-context viewing.

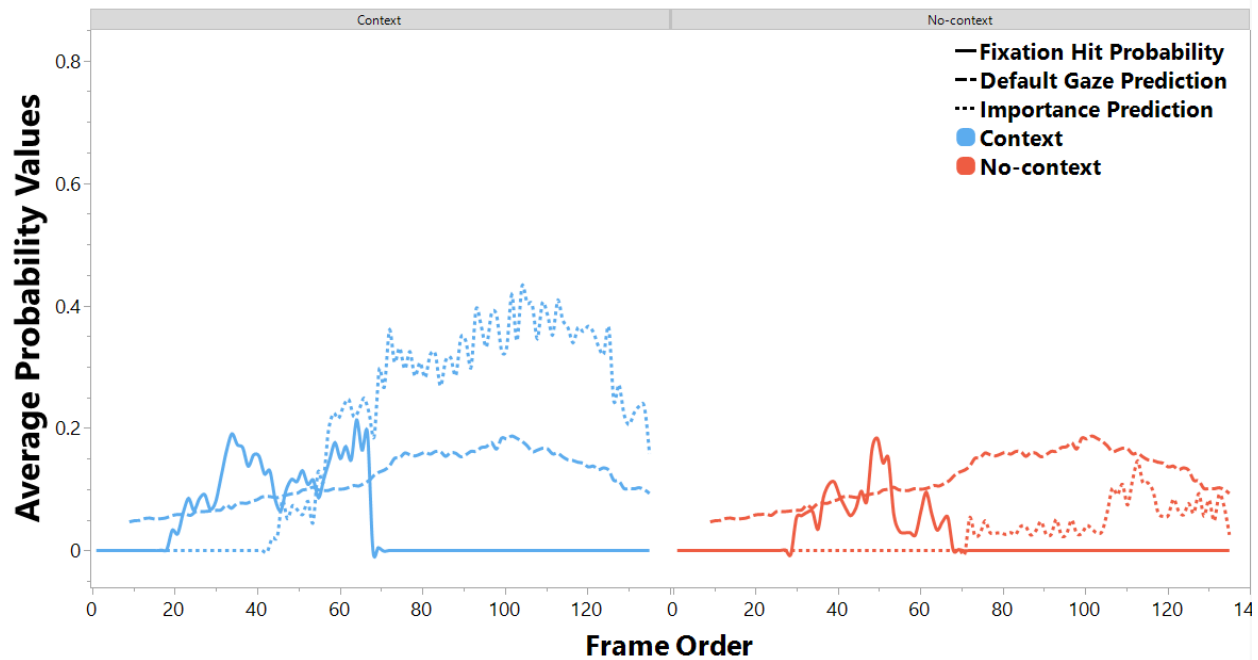


Figure 43

Frame by Frame changes in the raw average probability values of fixation hit probability and the two a priori predictions (Default Gaze Prediction and Importance Prediction) of the Security Personnell AOI in the Mr. Robot clip for both Context and No-context viewing.

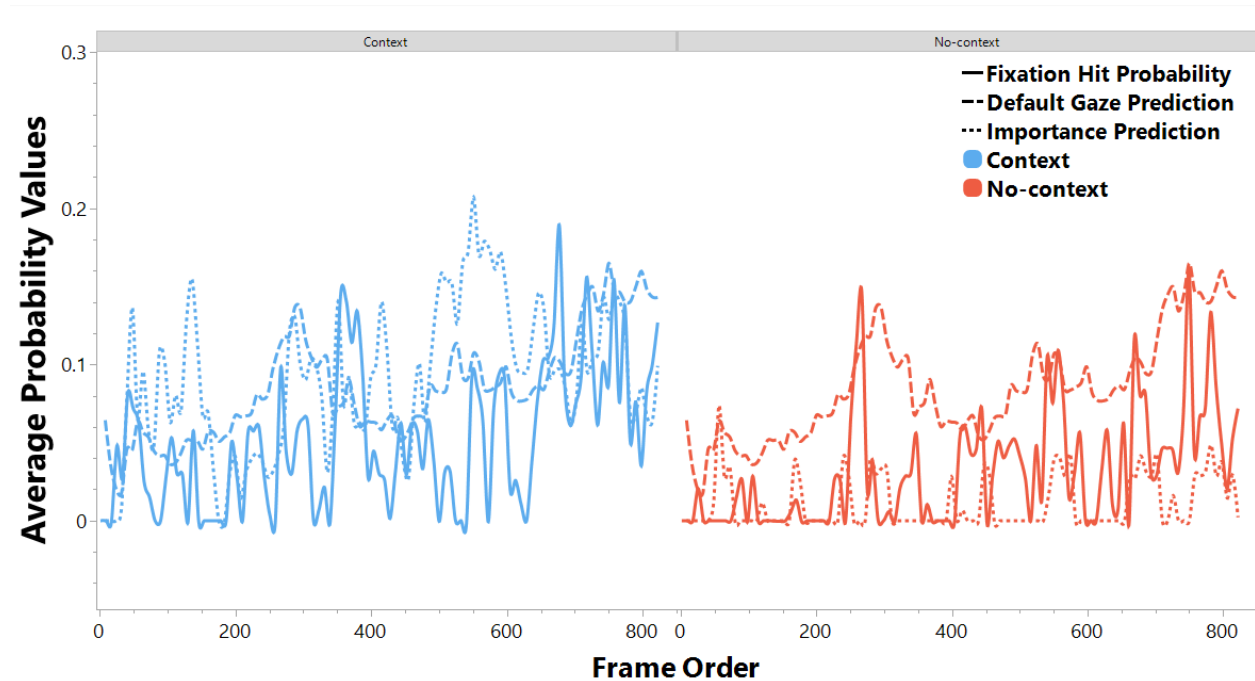


Figure 44

Frame by Frame changes in the raw average probability values of fixation hit probability and the two a priori predictions (Default Gaze Prediction and Importance Prediction) of the Security Personnel2 AOI in the Mr. Robot clip for both Context and No- context viewing.

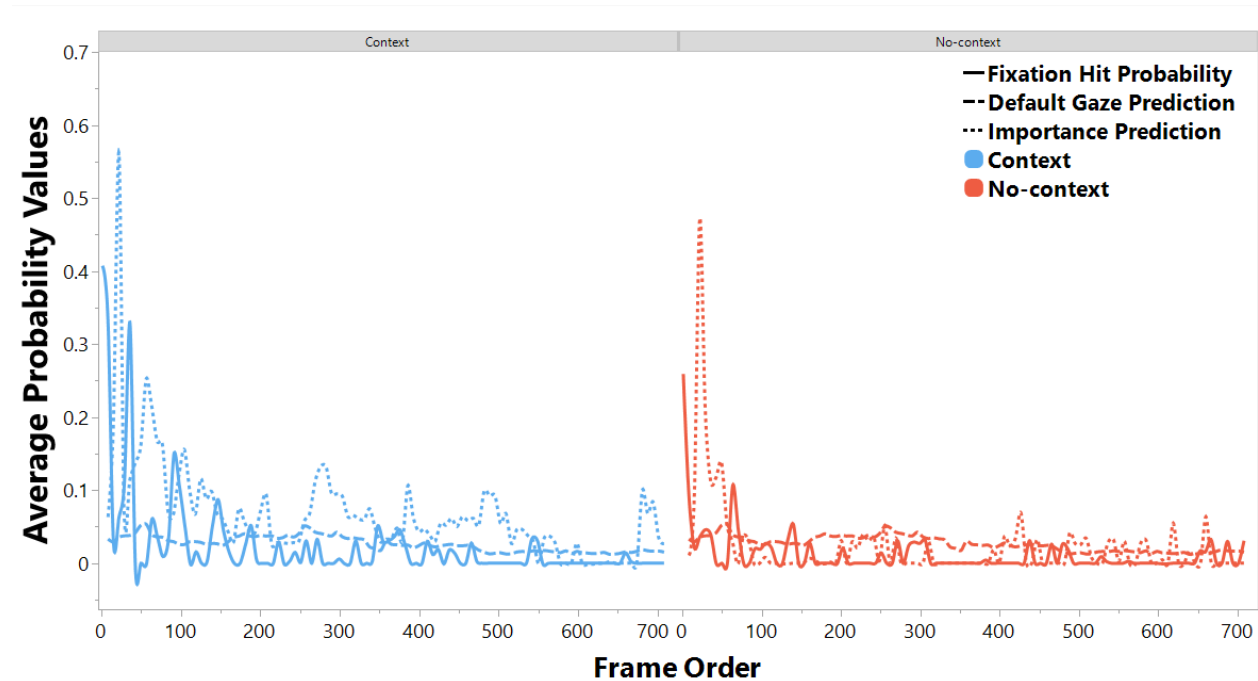
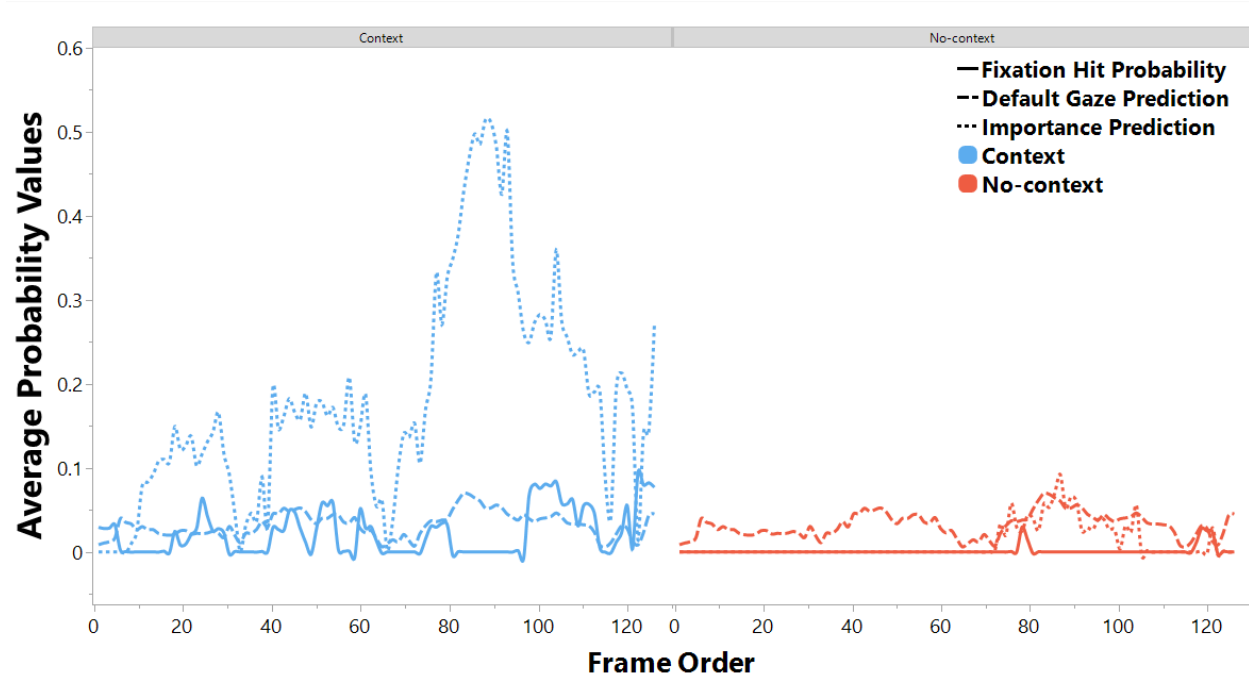


Figure 45

Frame by Frame changes in the raw average probability values of fixation hit probability and the two a priori predictions (Default Gaze Prediction and Importance Prediction) of the Rebel Sitting AOI in the The Passenger clip for both Context and No- context viewing.



Appendix N - Nested Model Comparison to Test the Relative Influence of Stimulus and Event Model

Model details

Base Model: 1 + (1 + Frame order|AOI) + (1 + Frame order|Participant)

Base + Stimulus Model: Default Gaze Prediction + (1 + Frame order|AOI) + (1 + Frame order|Participant)

Base + Stimulus + Event Model: Default Gaze Prediction + Importance Prediction + (1 + Frame order|AOI) + (1 + Frame order|Participant)

Table 14

Nested multilevel logistic regression model comparison for hypothesis testing and finding the best model fit for Fixation Hit Probability using Default Gaze Prediction and Importance Prediction as predictors.

Condition	Model Description	df	AIC	BIC	logLik	χ^2	Δdf	p
Context								
	Base Model	7	119440	119512	-59713			
	Base + Stimulus Model	8	118306	118389	-59145	1135.35	1	<.001
	Base + Stimulus + Event Model	9	117484	117577	-58733	824.15	1	<.001
No-context								
	Base Model	7	47494	47566	-23740			
	Base + Stimulus Model	8	46739	46822	-23362	757.13	1	<.001
	Base + Stimulus + Event Model	9	46727	46820	-23354	14.22	1	<.001

Note: The random-effect structure is the same across all models: random intercepts for each participant and AOI, along with random slopes for frame order varying by both participant and AOI.