Viewers Actively Extract and Maintain Spontaneous Theory of Mind Representations in Dramatic Irony Scenes

Cynthia Cabañas[1,2,3*], Atsushi Senju[1,4] and Tim J. Smith[3]

[1]Department of Communication Science, Vrije Universiteit Amsterdam, The Netherlands

[2]Department of Psychological Sciences, Birkbeck, University of London, UK

[3]Cognition in Naturalistic Environments (CINE) Lab, Creative Computing Institute, University of the Arts London, UK

[4]Research Center for Child Mental Development, Hamamatsu University School of Medicine, Hamamatsu, Japan

Author Note

Cynthia Cabañas [ORCID: https://orcid.org/0000-0002-2920-3907] is currently based at the Department of Communication Science, Vrije Universiteit Amsterdam.

Correspondence concerning this article should be addressed to:

Cynthia Cabañas
Department of Communication Science
Vrije Universiteit Amsterdam
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
Email: cynthiacabanas@gmail.com

Abstract

Understanding others' mental states is a cornerstone of human cognition, yet how spontaneous belief attribution operates in complex, naturalistic perception remains underexplored. Dramatic irony (DI)—a narrative device through which viewers are made aware of crucial information which is unknown to characters of the narrative—provides a controlled but naturalistic context to examine these processes. Using a curated DI film corpus (Cabañas et al., 2023), we combined eye-tracking with a Self-Paced Viewing (SPV) paradigm to investigate how SToM shapes real-time processing. In a pre-registered between-subjects study (https://osf.io/by56s), participants viewed edited Harold Lloyd film excerpts that either provided viewers with privileged information unknown to a character (Installation group) or withheld it to align viewer and character knowledge (Control group), reflecting false-belief and true-belief scenarios, respectively. We analysed narrative comprehension, reaction times, and eye-tracking measures indexing cognitive effort and attentional allocation. Results showed that the Installation group exhibited longer viewing times at moments of representational conflict and longer overall fixation durations, indicating greater cognitive effort. Gaze analyses revealed selective attention to belief-relevant characters and objects, suggesting top-down control guided by mental state representations. Extended dwell times on characters' mouths in the Installation group further reflected deliberate information seeking. These findings indicate that viewers spontaneously construct and maintain complex false-belief representations during DI comprehension, dynamically modulating cognitive processing and attention. By integrating naturalistic film stimuli with controlled manipulations, this study extends theories of scene comprehension to include spontaneous mentalizing, offering a tractable model of social reasoning in dynamic visual contexts.

Keywords: Dramatic Irony, Theory of Mind, Eye-Tracking, Film Narratives, Scene Processing, Cognitive Engagement, Attentional Strategies, Real-World Scene Perception

1. Introduction

When watching films, viewers are not just observing events unfold on screen—they make sense of characters' intentions and motives. To do so, they spontaneously consider what characters perceive and know, forming the basis for complex predictions about their behavior. This interpretation of characters' mental states may, in turn, shape how they process the unfolding narrative (e.g., (Kopatich et al., 2019; Levin et al., 2013; Magliano et al., 2024; O'Neill & Shultis, 2007; Persson, 2003; Rooney & Bálint, 2018). In film, as in real life, the understanding of others' perspectives draws on a key cognitive process known as Spontaneous Theory of Mind (SToM)—our tendency to attribute beliefs, desires, intentions, and emotions to others without explicit prompting (I. Apperly, 2010; Kovács et al., 2010; Meins et al., 2014). While some scholars distinguish between terms like Theory of Mind (ToM), mentalizing, and mindreading (Quesque et al., 2024), we use SToM here as a broad term for mental state attribution that occurs intuitively and continuously in dynamic social contexts, including narrative media. Unlike traditional Theory of Mind (ToM) tasks, which assess reasoning through explicit questions or external cues (e.g., Wimmer & Perner, 1983), SToM is self-initiated, arising from internal motivation or in response to contextual demands, without being prompted by external instruction.

Film narratives that feature Dramatic Irony (DI)—in which viewers hold privileged information unknown to certain characters—offer a particularly fruitful context for examining such spontaneous mental state reasoning. Although cinematic scenes are crafted stimuli, they offer rich, ecologically valid approximations of real-world social understanding (Smith, 2012), precisely because they embed SToM inferences within coherent narrative contexts. DI structurally parallels classic false belief paradigms, such as the Sally-Anne task (Baron-Cohen et al., 1985; Wimmer & Perner, 1983), by clearly delineating what characters know versus what they ignore. In these scenes, viewers must spontaneously attribute false belief to the "victims" of irony—characters who act based on incomplete or inaccurate information. This makes DI an ecologically valid yet systematically controlled way to study how SToM is initiated and how it shapes narrative processing.

Crucially, the experimental manipulation involved selectively presenting or omitting the installation segment, creating two conditions that differed only in viewers' access to key narrative information. In the Installation condition, this segment provided viewers with privileged knowledge, enabling them to attribute a false belief to characters who lacked crucial information. In the Control condition, its omission ensured that viewers shared the characters' limited perspective, producing a true-belief scenario. By maintaining a strict correspondence in visual and temporal characteristics across both versions and varying only the narrative context— an approach supported by studies in event cognition and discourse processing (e.g., Magliano et al., 2001; Rapp & Gerrig, 2006; Whitney et al., 2009)—this DI film corpus allows precise examination of how spontaneous false belief attribution influences viewers' cognitive and attentional processing in real-time scene comprehension. Using this film corpus and a free-recall task, Cabañas et al., (2023) found that participants who viewed the installation segment formed more elaborate event models of characters' mental states than those in the control condition. This elaboration reflected greater integration of the characters' ignorance and misconceptions, leading to increased focus on cognitive aspects like beliefs and intentions over affective responses. This was reflected in higher mental state reference frequency, indicating deeper cognitive engagement in their interpretation of character actions.

While these findings suggest that viewers spontaneously represent characters' false beliefs during dramatic irony, they leave open the question of how these SToM processes unfold in real time. In particular, Cabañas et al., (2023) did not examine whether mental state reasoning is reflected in moment-to-moment processing differences, such as visual attention. Understanding how mental state reasoning guides visual attention is key to developing more accurate models of real-time social cognition in naturalistic settings.

The present study builds on this by investigating whether SToM during dramatic irony manifests in viewers' gaze behavior. Research on visual attention in static scenes has shown that gaze is actively guided by viewers' internal representations and mental models (Castelhano et al., 2009; Henderson, 2003; Luke & Henderson, 2016), influencing fixation durations, saccade amplitudes, and object selection (Castelhano et al., 2009; Huettig et al., 2011; Luke & Henderson, 2016). The Scene Perception and Event Comprehension Theory (SPECT; Loschky et al., 2020) similarly proposes that attention in dynamic scenes is guided by evolving mental models. However, the extent to which these top-down influences emerge during film viewing appears to depend on stimulus and task conditions. Smith & Mital (2013) demonstrated that a viewing task prioritising static features of a dynamic scene (i.e. the background location) significantly altered gaze behaviour compared to free viewing. By comparison, Hutson et al., (2017) found minimal gaze differences despite comprehension variation, reinforcing the dominance of low-level stimulus features, such as editing or camera movement—a phenomenon they refer to as 'the tyranny of film' . But, similar to Smith & Mital (2013), when viewers were instructed to watch the scene in preparation for a map-drawing task that competed with narrative comprehension, gaze became less synchronized and more exploratory (longer saccades, wider dispersion). Likewise, removing motion revealed comprehension-driven gaze differences (Hutson et al., 2022), suggesting top-down effects emerge when bottom-up saliency is reduced. Further, research also suggests distinct visual attentional signatures depending on the type and availability of contextual knowledge. For instance, Pedziwiatr et al., (2023) found that prior event knowledge reduced visual exploration, while Liu et al., (2022) showed that higher cognitive load increased fixation durations, consistent with viewers dedicating more cognitive resources to the detailed processing of task-relevant stimuli.

How these mechanisms unfold during spontaneous ToM attribution in dynamic audiovisual scenes has been under researched. Although static ToM tasks such as the Director Task (Cane et al., 2017; Symeonidou et al., 2016) have revealed perspective-based gaze differences, they lack the contextual richness and ecological validity of narrative-based social reasoning (Kingstone et al., 2008). This limits their ecological validity and their ability to capture the internally driven nature of spontaneous ToM. To address these gaps, the current study examines how viewers construct event models in real time and whether different types of SToM representations embedded in those models shape visual attention during narrative comprehension.

A key challenge in real-time narrative comprehension is the continuous updating of event models—especially when the narrative introduces information that conflicts with characters' beliefs. Prior research has shown that this process is cognitively demanding, particularly when viewers must integrate new, incongruent information (e.g., Huff et al., 2017; Magliano & Zacks, 2011; Papenmeier et al., 2019; Zacks & Swallow, 2007), including in the context of sequential images like comic strips (Cohn & Paczynski, 2013).

False belief scenarios like those present in DI exemplify this difficulty, as it requires viewers to simultaneously represent two conflicting realities: the actual state of the world and the character's mistaken belief about it (Csibra & Gergely, 1998, 2007; Perner, 1991). This dual-representation process introduces representational conflict and is assumed to place greater demands on cognitive resources than scenarios involving true beliefs, in which character knowledge aligns with reality and thus requires no such reconciliation. Studies using visual perspective-taking tasks have consistently shown that judging another's mistaken belief increases reaction times and error rates (I. A. Apperly et al., 2006; Keysar et al., 2003; Samson et al., 2010), highlighting the cognitive demands of maintaining conflicting representations.

These findings support both Apperly's two-systems model (Apperly et al., 2010))—which distinguishes between fast, efficient ToM and slower, effortful ToM—and more recent single-system accounts like the Analogical Theory of Mind framework (Rabkina & McFate, 2022), which argue that both types of reasoning can arise from a unified cognitive architecture. In both frameworks, effortless and effortful ToM are context-dependent, shaped by factors such as mood, task demands, and internal or external motivation (Westra, 2017).

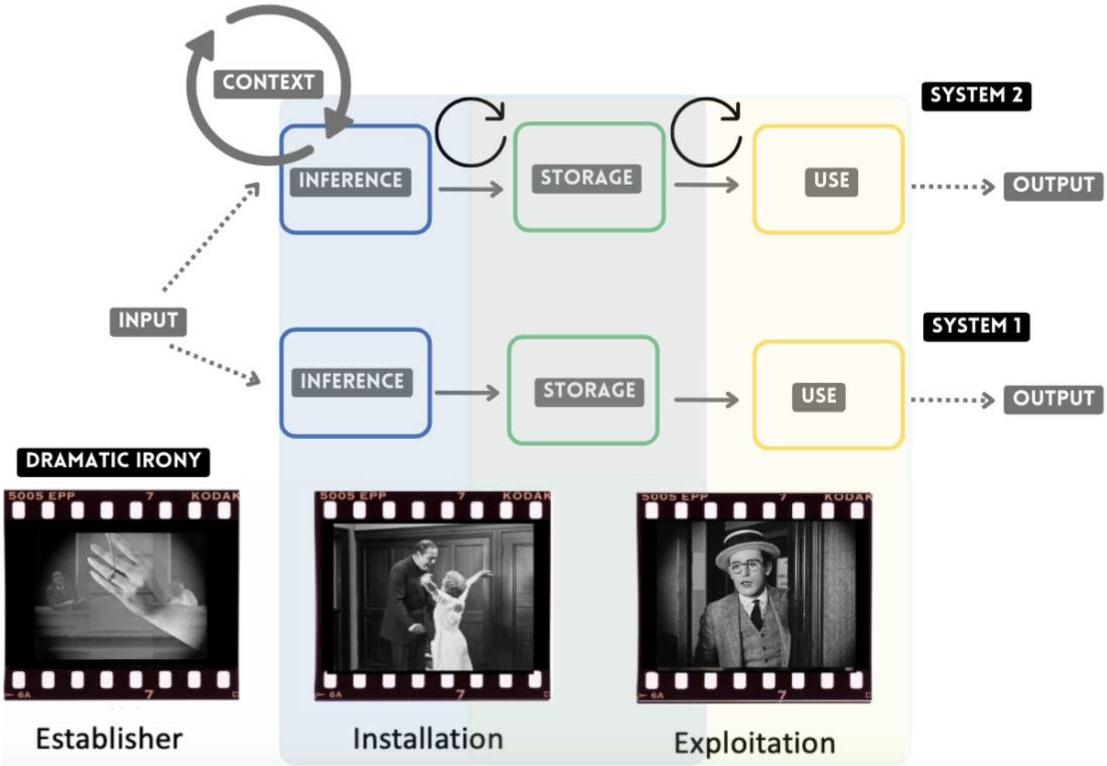


Figure 1. Cognitive processing of dramatic irony across narrative phases, adapted from Apperly's (2010) dual-system model of mindreading. The upper panel illustrates three cognitive stages—inference, storage, and use—mapped onto narrative segments in Harold Lloyd's silent film *Never Weaken* (Newmeyer, 1921), shown in the lower panel. *Establisher*: Harold proposes to his love interest, who accepts. *Installation*: She is later seen with another man—her brother and a minister—unbeknownst to Harold. *Exploitation*: Harold misinterprets the scene and, believing the man is a romantic rival, leaves in distress. Viewers in the Installation condition presumably infer Harold's false belief during the Installation (blue), maintain it across time (green), and apply it to interpret his actions in the Exploitation (yellow). The dual-pathway architecture highlights both fast, intuitive processing (System 1) and slower, effortful reasoning (System 2), with recursive arrows in System 2 indicating ongoing re-evaluation as new contextual information is integrated.

As visualized in Figure 1, this reasoning process is presumed to unfold across narrative phases in a dramatic irony sequence: the installation segment prompts spontaneous false belief inference, which must be maintained over time and applied during the exploitation to interpret the character's actions. However, this dual-system distinction has primarily been studied using explicit ToM tasks; whether similar dynamics apply in naturalistic, spontaneous ToM contexts—such as narrative film viewing—remains an open question.

In the current study, we extend this line of inquiry by testing whether dramatic irony systematically modulates visual attention in ways that reflect the cognitive demands of updating event models with information derived from SToM processing. Using the DI Film Corpus, we combined eye-tracking with a Self-Paced Viewing (SPV) paradigm to examine whether viewers in the false belief scenario (Installation group) display longer viewing times indicative of greater processing load than those who are in the true-belief scenario (Control group). By combining Self-Paced Viewing with eye tracking, this approach allows us to assess whether spontaneous false belief attribution is reflected in moment-to-moment visual attention and processing effort during naturalistic film viewing.

For this purpose, segments from the DI Film Corpus were presented as static frames, enabling participants to advance at their own pace. While this format sacrifices the natural continuity of film, the controlled presentation enhances the visibility of top-down attentional mechanisms by minimizing the influence of automatic gaze cues (such as motion), making it especially well-suited to capture the subtle dynamics of spontaneous Theory of Mind. By using SPV, we bridge the gap between tightly controlled false belief paradigms and rich, naturalistic film experiences. Crucially, this method offers a highly sensitive way to detect when and how viewers allocate attention in response to conflicting mental states—something traditional ToM tasks cannot capture. It allows us to pinpoint the exact moments where narrative complexity increases cognitive load, and to observe how false belief attribution actively shapes visual processing during real-time comprehension. This provides a powerful tool for studying the cognitive mechanics of SToM in ecologically valid, yet experimentally tractable, settings.

The current study investigates how spontaneous false belief attribution during DI influences real-time cognitive and attentional processing during film viewing. Building on prior work showing that viewers retrospectively form richer mental models when exposed to DI (Cabañas et al., 2023), we examine whether these differences are also reflected *during* scene comprehension—both in terms of cognitive effort and where viewers direct their gaze.

We address two core research questions.

RQ1 asks whether previously identified retrospective differences in comprehension manifest as real-time differences in cognitive effort during viewing. To address this question, we first include a manipulation check: we predict that viewers in the Installation group—those who viewed the installation segment providing information unknown to the character—will construct qualitatively different event models than those in the control condition, as reflected in higher DI comprehension scores. Our first main hypothesis (H1) posits that participants in the Installation group will exhibit greater cognitive effort than those in the Control group, reflecting spontaneous false-belief attribution. This difference should manifest in the SPV task as longer viewing durations (i.e., increased reaction times) for the Installation group at narrative moments where representational conflict is most salient. Because the clips vary in content and staging, we do not specify *a priori* the precise moments at which these effects will emerge.

RQ2 examines how spontaneous false-belief attribution during dramatic irony modulates visual attention strategies across viewing phases (Shared Clip vs SVP phase). Given the context-dependent effects summarized above, we advance two competing predictions. Our primary prediction (H2a) is that participants in the Installation group will exhibit longer fixation durations during the SPV phase compared to the Shared Clip phase, and relative to the Control group, reflecting increased cognitive load. As an exploratory alternative (H2b), participants in the Installation group may instead display more active narrative exploration during the SPV phase, evidenced by a higher number of fixations, longer saccades, and greater gaze dispersion. Finally, we hypothesize (H3) that viewers in the Installation group will selectively devote greater dwell time to narratively critical elements—such as the 'victims' of dramatic irony or objects central to the false-belief conflict—than those in the Control group, reflecting selective, goal-driven attention guided by spontaneous mental state reasoning.

2. Methods

2.1 Design

This study employed a mixed-design, with one between-subjects factor—condition (Installation vs. Control groups)—and one within-subjects factor—phase (Shared Clip vs. Self-Paced Viewing [SPV]). The primary dependent variables were: (1) *Dramatic Irony comprehension scores*, (2) *reaction times (RTs)* during SPV, and (3) a series of eye-tracking metrics indexing cognitive effort and attentional allocation. These included: fixation duration (reflecting processing load), fixation count (capturing the extent of information search), saccade amplitude (indicating visual exploration range), distance to screen centre (marking attentional deviation from default gaze), and proportion of dwell time in pre-defined Areas of Interest (AoIs), quantifying time spent on narratively relevant regions of the screen (see Table 2 for an overview of AoIs and their narrative definitions). Participants were randomly assigned to one of the two conditions, and the order of the six experimental clips was randomized for each participant.

2.2 Participants

Thirty participants (13 female; *M* age = 30, *SD* = 9.24) were recruited via the university's SONA system and randomly assigned to either the Installation or Control group (n = 15 per group). An a priori power analysis using G*Power (Faul et al., 2009) determined a minimum sample of 24 participants to achieve 0.80 power, based on a large effect size (Cohen's *d* = 1.092) observed in Cabañas et al. (2023) for mental state reference differences. To account for potential exclusions, a 25% buffer was added, leading to a total of 30 recruited participants.

Exclusion criteria included: prior familiarity with any film clips, insufficient English proficiency, diagnosis or ongoing assessment for Autism Spectrum Conditions (due to their documented association with atypical ToM processing; Chung et al., 2014; Happé, 1994; Senju, 2012), and low-effort responses (i.e., fewer than one sentence per scene in the free-recall task). Based on these criteria, six participants were excluded: four due to technical issues or incomplete data, one for prior clip familiarity, and one for undergoing Autism diagnosis.

The final sample consisted of 24 participants (12 per group). Ethical approval was granted by the Birkbeck, University of London Ethics Board (ID: 181949). All participants provided written informed consent prior to participation.

2.3 Stimuli

The study employed the Dramatic Irony Film Corpus (Cabañas et al., 2023), which features six excerpts from Harold Lloyd silent comedies (e.g., *Never Weaken*, 1921; *The Freshman*, 1925). Each excerpt was edited into two versions: a Control condition including only the establisher and exploitation scenes, and an Installation condition which additionally included the installation scene, thereby creating a knowledge disparity between viewer and character. All clips were silent, black-and-white, and selected for their capacity to elicit spontaneous false belief attribution. This design ensured that both groups viewed visually and temporally matched clips that differed solely in narrative information. Clips ranged from 44 to 318 seconds, preceded by a 5-second grey screen baseline to collect pupil data. Each clip was presented at 24 frames per second with a resolution of 720 × 480 pixels. Only the Exploitation scenes were shown in a SPV format. Slideshows were created by extracting every sixth frame (i.e., 1 frame per 250 ms) to approximate the natural pace of the original clip while minimizing participant fatigue. Resulting slideshows consisted of 156 to 247 frames per scene, depending on length. This extraction rate was informed by pilot testing, which showed that it preserved narrative continuity and allowed smooth, self-paced progression. For a visual overview of the DI structure, see Figure 1; for detailed scene breakdowns, see Table 1 and Supplementary Text 1. The corresponding clips are also available at the OSF repository (https://osf.io/q3pv7/).

Table 1. Summary of film clips with respective Control and Installation versions, including their duration and a brief description of the scenes.

Clip Title	Version	Duration	Short Description
Never Weaken I	Control	2m 21s	Harold aims to showcase an osteopathic clinic's effectiveness by healing a man on the street, attracting new clients to the clinic where his love interest works.
	Installation	3m 26s	Harold devises a plan with an acrobat to stake a fake injury recovery, drawing the attention of potential clients to the osteopathic clinic where his love interest works.
The Freshman	Control	3m 26s	Harold attempts to join the football team, successfully secures a spot, and enthusiastically heads to the field to play.
	Installation	3m 54s	Harold excitedly tells a girl he made the team and eagerly goes to the field, not knowing that his real role is the water boy.
Never Weaken II	Control	1m 09s	Harold proposes to a girl and overhears a conversation where he finds out that she is being proposed to by another man.
	Installation	1m 25s	Harold proposes to a girl who accepts, but later misunderstands her conversation with her brother, thinking she's being proposed to by another man.
Girl Shy	Control	2m 26s	Harold tries to publish his book but is rejected by the publisher and receives a rejection letter in the mail.
	Installation	3m 06s	Harold attempts to publish his book, and although initially rejected, the publisher reconsiders. Harold, believing the letter contains a rejection slip, tears up the unopened envelope containing a check.
For Heaven's Sake	Control	4m 11s	A missionary and his daughter write to Harold for help raising money for their mission. Harold comes across their mission cart and offers them a significant contribution.
	Installation	5m 14s	Harold accidentally burns a mission cart, writes a check to compensate, but is mistaken for a generous donor for the mission.
The Kid Brother	Control	1m 32s	Two men trick the sheriff into signing a permit for their traveling show.
	Installation	5m 06s	Harold, dressed as his sheriff father, is tricked into signing the permit for the two men.

2.4 Apparatus

The experiment was run using Experiment Builder software (SR Research) on a 17-inch CRT monitor with a resolution of 1024 × 768 pixels, viewed at a distance of 60.69 cm using a chin and forehead rest. This setup subtended a visual angle of 21.42° × 16.10°. Eye movements were recorded with an EyeLink 1000 desktop-mounted eye tracker (SR Research, Version 1.5.2), sampling at 1,000 Hz. Calibration used a nine-point grid, with a maximum allowed spatial error of 1° of visual angle and an average error of 0.5° or less.

2.5 Procedure

The study was conducted in person at the MERLiN lab (Birkbeck, University of London). Upon arrival, participants were seated at a desk-mounted eye-tracker and completed a 45–60 minute session. After calibration, participants viewed six film excerpts on a monitor (~90 cm from chinrest), depending on their assigned group (Installation or Control). Each excerpt was preceded by a 5-second gray screen (baseline). Participants first watched the establisher segment (and installation segment, if applicable) in full motion. Before the exploitation scene, a prompt in the top-right corner indicated the beginning of a self-paced slideshow. Participants advanced through the static frames by pressing the spacebar; reverse navigation was disabled, and no pacing instructions were given. Reaction times (RTs) per frame were recorded. After each clip, participants completed a free recall task to assess narrative comprehension, without specific mental state prompts. DI comprehension was coded from these reports following Cabañas et al. (2023), with inter-rater reliability (25% sample) assessed using Krippendorff's alpha (α = 0.938). After the six blocks, participants completed the Reading the Mind in the Eyes Test (RMET;(Baron-Cohen et al., 2001) to control for ToM ability, and answered open-ended debriefing questions (e.g., "Did you notice anything similar across the clips?"). No participants reported awareness of the DI pattern. Participants were then fully debriefed.

3. Results

Data Cleaning and Analysis.

The analysis plan for the RTs of the SPV task was prespecified in a preregistration document that can be found on the Open Science Framework [https://osf.io/by56s]. RMET scores were excluded from analysis due to their administration after the eye-tracking task, which may have influenced performance. However, scores were reviewed to ensure all participants met the threshold for adequate mentalizing ability.

To ensure valid responses within the self-paced viewing (SPV) paradigm, we first discarded SPV reaction times (RTs) above 2000 ms. This initial cutoff was informed by task parameters: since each frame was designed to require rapid sequential viewing, excessively long RTs likely reflected momentary disengagement or distraction rather than true cognitive processing. Thus, the 2000 ms threshold functioned as a safeguard against values incompatible with the task's real-time demands. Subsequently, we computed the mean and standard deviation of the remaining RTs and applied a 3SD trimming procedure. This is a standard approach to remove statistical outliers while preserving the majority of the data. We opted for a 3SD (rather than 2SD) criterion to avoid overly aggressive exclusion, thus retaining potentially informative responses at the longer end of the distribution. All data handling and statistical analyses were conducted in R using RStudio.

Manipulation Check: DI Comprehension Scores.

To verify the effectiveness of the experimental manipulation, we compared participants' DI comprehension scores between the Installation and Control groups. A Welch two-sample t-test revealed significantly higher DI comprehension scores in the Installation group (M = 8.75, SD = 2.30) compared to the control group (M = 0.83, SD = 1.03), $t(15.24) = -10.88$, $p < .001$, $d = 4.44$.

We also analyzed DI comprehension scores for each of the six clips separately. Welch's t-tests with Bonferroni correction showed significant differences across all clips: Clips 1, 5, and 6 showed moderate effects (all $p < .0062$), while Clips 2, 3, and 4 exhibited stronger differences ($p < .0001$). Clip 1 showed improved condition sensitivity relative to prior studies (Cabañas et al., 2023), while Clip 5 had the lowest comprehension scores overall—below partial understanding—indicating comprehension difficulties. These per-clip results are visualized in Figure 2.

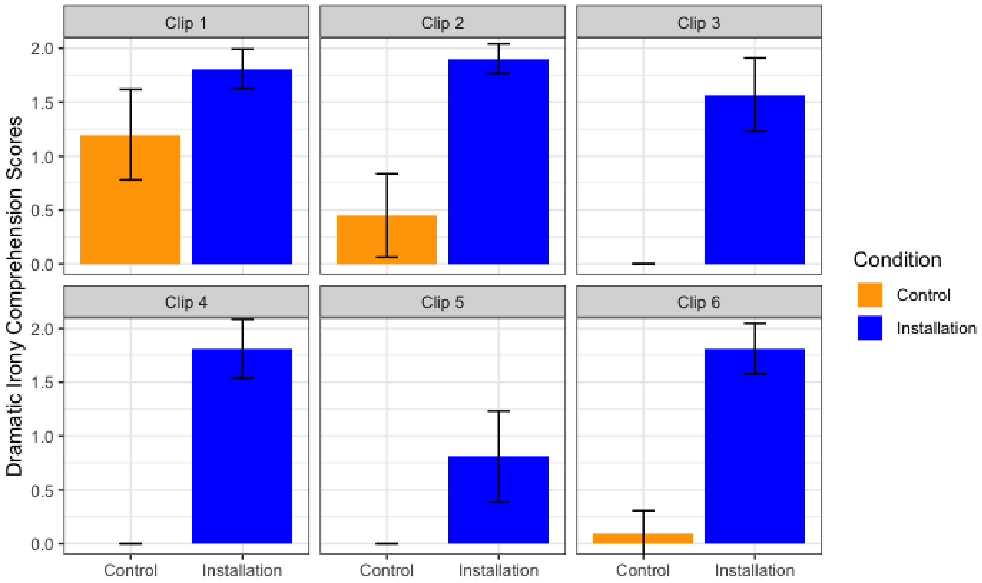


Figure 2. Bar plot depicting DI Comprehension scores averaged across clips distinguished by Condition group. Error bars represent the confidence intervals.

3.2 Increased Cognitive Effort During Exploitation Scenes: Self-Paced Viewing (SPV) Reaction Times

H1: Overall RT Analysis Across All Clips

To assess whether spontaneous false belief attribution increases cognitive effort, we analyzed SPV reaction times (RTs) during the exploitation scenes—identical across conditions but preceded by differing narrative contexts. A linear mixed-effects (LME) model approach was used on log-transformed RTs to address data skewness and account for both participant-level and clip-level variability. Stepwise comparisons showed that including random intercepts for participants and clips substantially improved model fit. Adding condition alone did not improve the model (L.Ratio = 0.634, p = .426). However, adding the condition × clip interaction significantly improved fit (AIC = 17465.33 vs. 17624.16, L.Ratio = 168.83, p < .0001), indicating that condition effects on RTs varied across clips. Full model outputs are presented in Table 1 in the Supplementary Materials.

Clip-Specific RT Analysis

To locate where effects occurred, separate LMEs were conducted for each clip. Clips 1, 2, 4, and 6 showed significantly longer RTs for the Installation group at specific frames (*p* < .05 for condition × frame interaction), indicating greater cognitive effort. No significant effects were found for Clips 3 and 5. Given Clip 5's poor DI comprehension (mean < 1), it was excluded from further analysis.

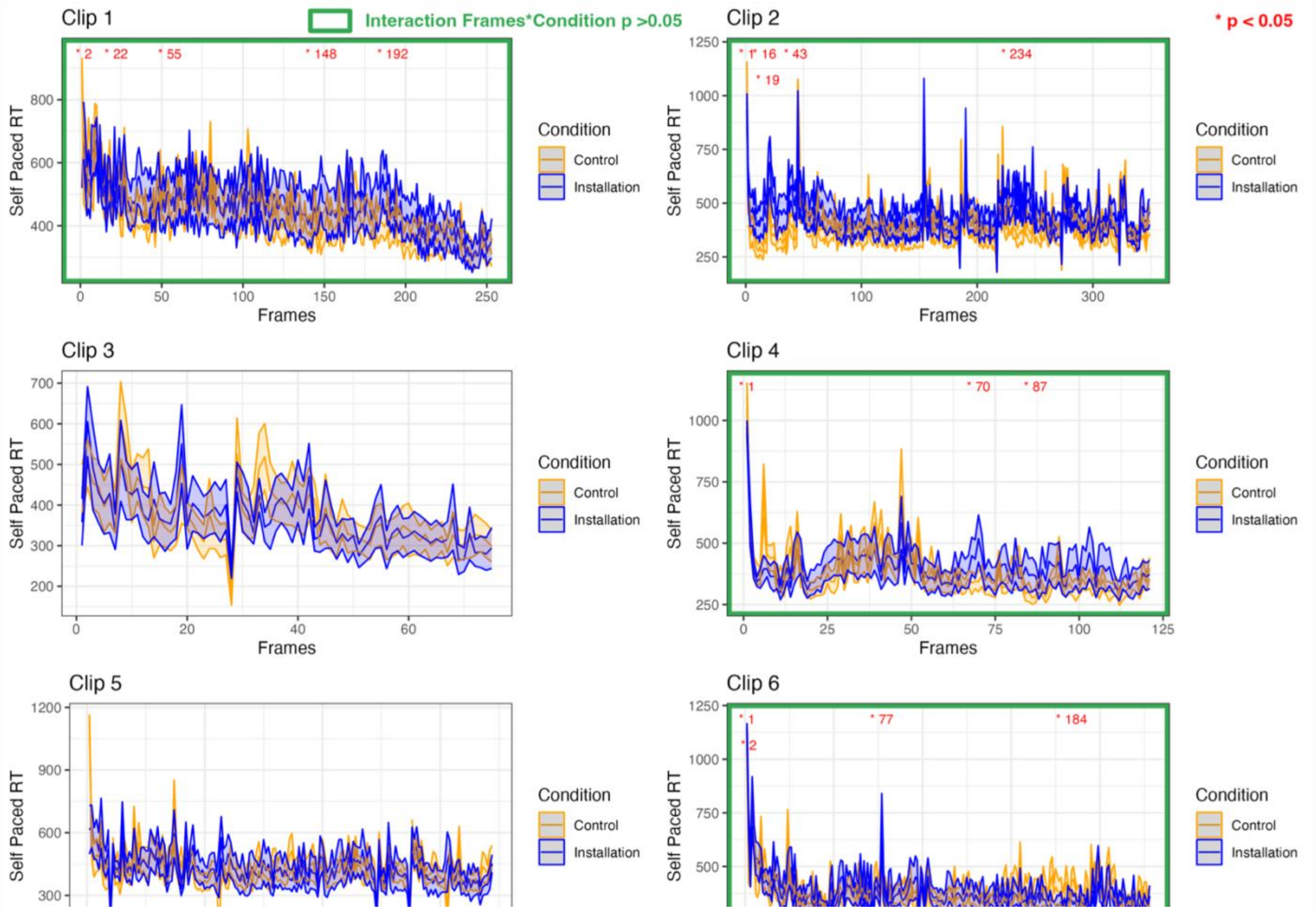


Figure 3. Graphical representation of self-paced reaction time data across film clips. The Y-axis represents the observed variability in reaction times over time (x-axis: frames). Colour-coded lines distinguish participants exposed to the installation (installation =blue; vs. control = orange). Green outlines indicate clips where significant interactions between condition and clip were identified in the LME models, reaching significance in frames marked with red asterisks. These frames show longer viewing times for the installation group, suggesting deeper cognitive processing of critical moments, compared to the control group. There were no significant findings in the reverse direction. Shaded areas represent SEM.

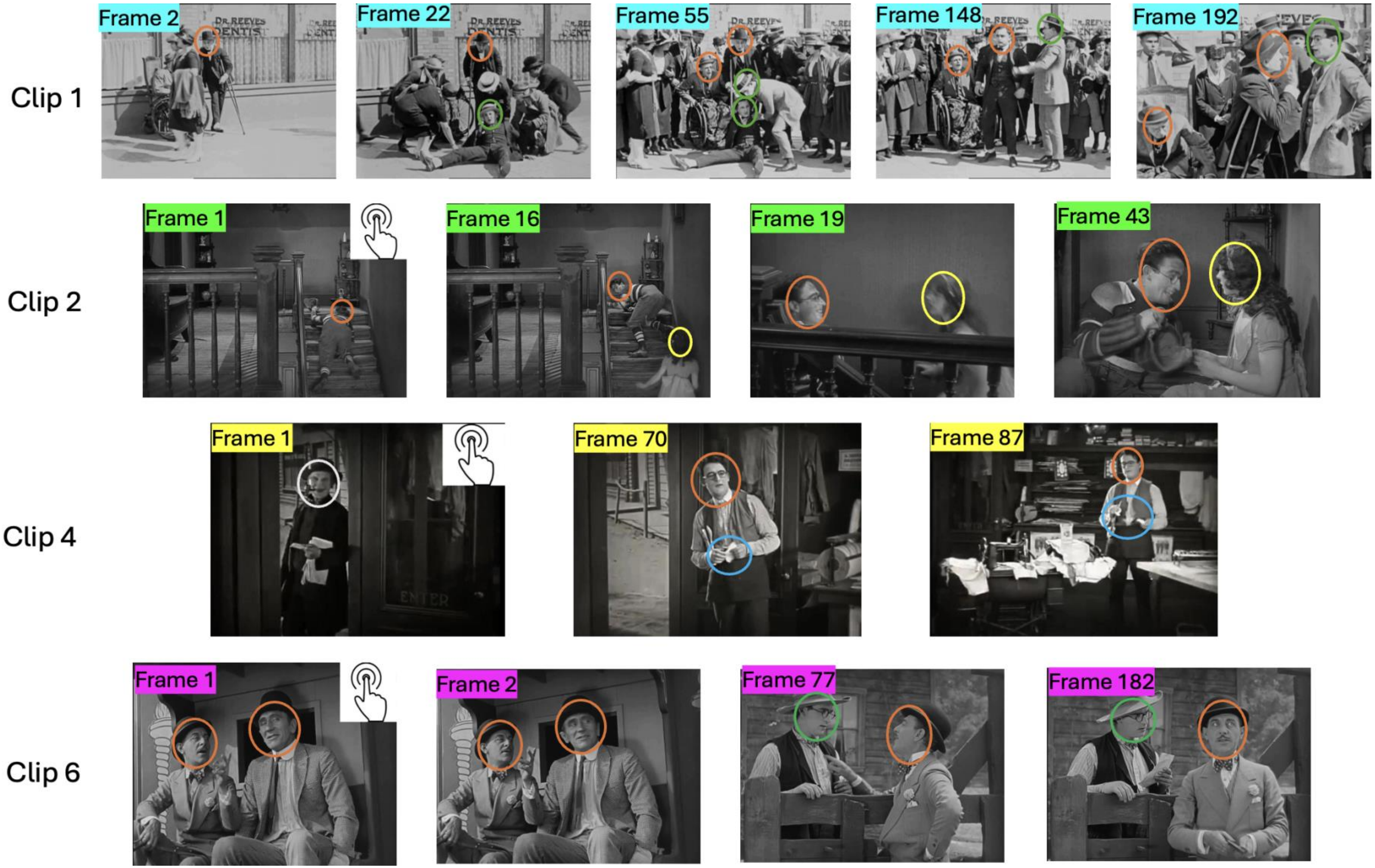


Figure 4. Comprehensive overview of the scenes at each significant frame, illustrating the critical moments of dramatic irony that elicited varied reaction times among participants. Coloured circle demarcations signify different mapping of AoIs: orange highlights AoIs linked to victims of dramatic irony; green marks the scene's protagonist or the overarching protagonist; yellow denotes witnesses to the dramatic irony; blue identifies objects significant to the narrative, such as the envelope with the cheque; and white outlines characters deemed irrelevant to the unfolding events.

For Clips 1, 2, 4, and 6, we identified specific frames with significantly higher RTs in the Installation group using Benjamini-Hochberg corrected contrasts. These frames and patterns are summarized in Figure 3 (RTs per frame) and Figure 4 (highlighted significant frames), with detailed statistics available in Supplementary Tables 1–2.

H2: Eye-Tracking Metrics: Attentional Effects of DI Exposure

To test whether prior exposure to the installation scene altered visual attention patterns during shared exploitation scenes, we analyzed four eye-tracking measures using mixed two-way ANOVAs: Fixation Duration, Fixation Count, Saccade Amplitude, and Distance to Screen Centre. Each analysis included phase (Shared Clip vs. SPV) as a within-subjects factor and condition (Installation vs. Control) as a between-subjects factor.

Fixation Duration

A significant main effect of phase was found ($F(1, 22) = 59.29$, $p < .001$, $\eta^2 = .374$), alongside a significant phase × condition interaction ($F(1, 22) = 9.17$, $p = .006$, $\eta^2 = .085$). Planned contrasts revealed longer fixation durations in the installation group during the SPV phase ($M = 253.13$ ms) than in the control group ($M = 216.95$ ms), $t = 2.34$, $p = .0252$, $d = 0.41$.

Fixation Count

Fixation count violated normality assumptions (Shapiro-Wilk $p < .001$), so we applied the Aligned Rank Transform (ART). Neither main effects nor interaction effects were significant (all $p = 1.000$).

Saccade Amplitude

A main effect of phase was observed ($F(1, 22) = 227.92$, $p < .001$, $\eta^2 = .652$), but no effects of condition or interaction emerged ($p > .5$).

Distance to Screen Centre

Similarly, a main effect of phase was significant ($F(1, 22) = 15.98$, $p < .001$, $\eta^2 = .077$), but there were no significant condition or interaction effects ($p > .47$).

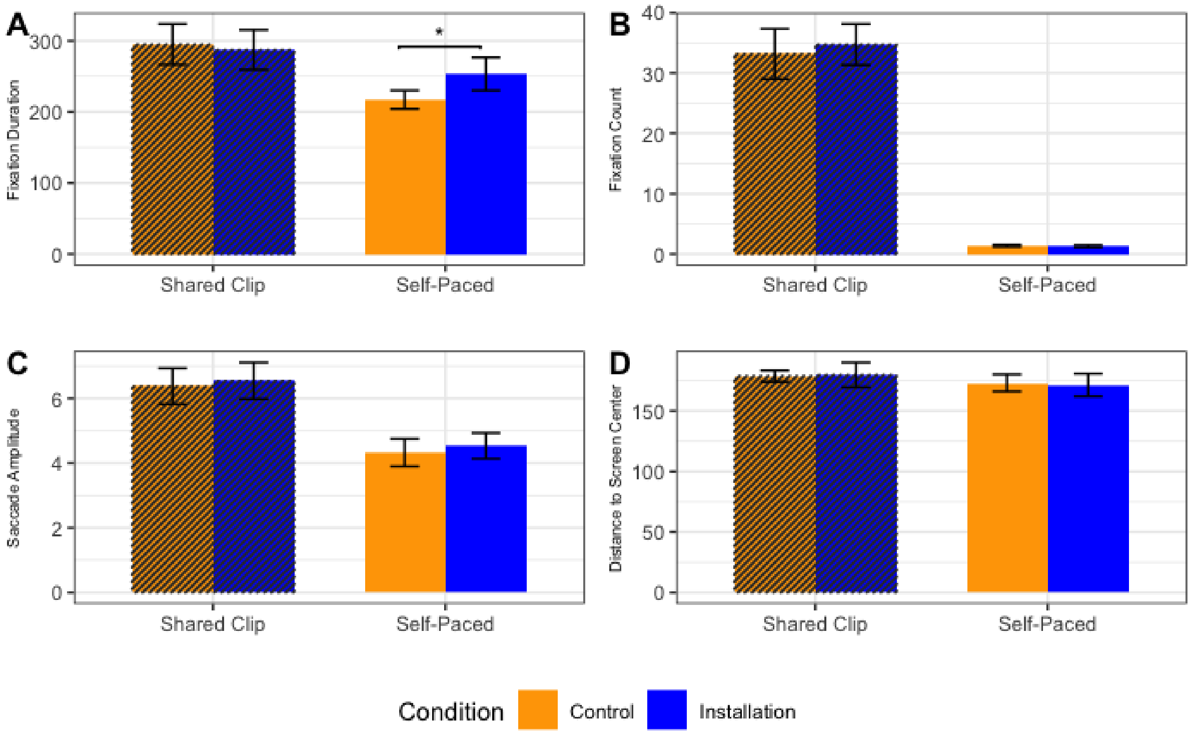


Figure 5. Comparative Visualisation of Eye-Tracking Metrics Across Different Viewing Conditions. This figure comprises four subplots (A-D) each representing a distinct eye-tracking metric: (A) Fixation Duration, (B) Fixation Count, (C) Saccadic Amplitude, and (D) Distance from Centre. All plots share a consistent colour scheme and x-axis categorisation, denoting different viewing conditions. The y-axis varies across the plots, representing the respective dependent variable for each metric. Error bars indicate the 95% confidence intervals, and significance asterisks (*) highlight statistically significant differences between conditions in the phase of interest (*SVP phase*).

H3: Dwell Time in Areas of Interest (AoIs): Spatiotemporal Attention and Narrative Relevance

To assess whether participants directed attention to elements relevant to the DI conflict, we analyzed gaze behavior in pre-defined Areas of Interest (AoIs) during the shared SPV phase of each clip. AoIs were selected based on narrative significance: they included characters or objects expected to be especially relevant for viewers who had access to the Installation scene and therefore could attribute a false belief. The focus was on Clips 1, 2, 4, and 6, as prior analyses indicated significant RT effects.

Mixed two-way ANOVAs (condition × AoI) were conducted using the Aligned Rank Transform (ART) due to violations of normality in dwell time distributions (Shapiro-Wilk p < .001 in all clips). A summary of the results of clip-specific AoI analysis for all clips which were previously identified as significant in the SPV analysis is depicted in Figure 4 (see also Table 2 for an overview of AoIs and narrative context across clips).

Table 2. Overview of AoIs and Narrative Elements Across Clips. The table outlines how each clip's narrative context and AoIs were hypothesised to influence viewer attention in situations of DI.

Clip	AoIs	Explanation of Critical Narrative Elements for AoIs
1	Face of protagonist vs. face of victim	**Exploitation**: An acrobat pretends to fall and get injured in front of a crowd of potential clients and Harold pretends to be an osteopath who can help him. When the acrobat seems to heal quickly, the crowd gets excited and asks for cards for the osteopathic clinic. **Victim of DI**: members of the crowd whose goal is to get a treatment for their illnesses or physical challenges. **Hyp:** Viewers will focus more on members of the crowd (victim of DI) near or shaking hands with Harold (protagonist), due to the staged recovery plot.
2	Face of character vs. face of victim/protagonist	**Exploitation**: Harold excitedly tells a girl that he made the team, unaware that he is just the water boy. **Victim of DI**: Harold whose goal is to be a football player. **Hyp**: Viewers will focus more on Harold (victim of DI and protagonist) than at the girl who is listening.
4	Face of victim/protagonist vs. letter	**Exploitation**: Harold, downhearted after rejection to publish his book and unaware of the content of the letter, tears it apart without opening and realising it is actually a cheque from the publuisher. **Victim of DI**: Harold whose goal is to make money with his book. **Hyp**: Viewers will focus more on the letter containing a check, which Harold (victim of DI and protagonist) mistakenly believes to be a rejection slip.
6	Face of protagonist, face of victim 1, face of victim 2	**Exploitation***:* Two men see Harold dressed up as sheriff and convince him to sign a permit for a traveling show. **Victim of DI**: The two men. **Hyp**: Viewers will focus more on the two men (victims of DI), who mistakenly believe Harold (protagonist) is the sheriff.

Note: Clip 3 is discussed separately as an exploratory analysis of gaze distribution over emotional facial regions.

Full results for all four clips are reported in Supplementary Tables 3–6. However, we highlight Clip 4 (which uniquely featured a key object-based belief conflict) and Clip 3 (an exploratory analysis on attention to facial regions across emotions) in the main text due to their theoretical relevance.

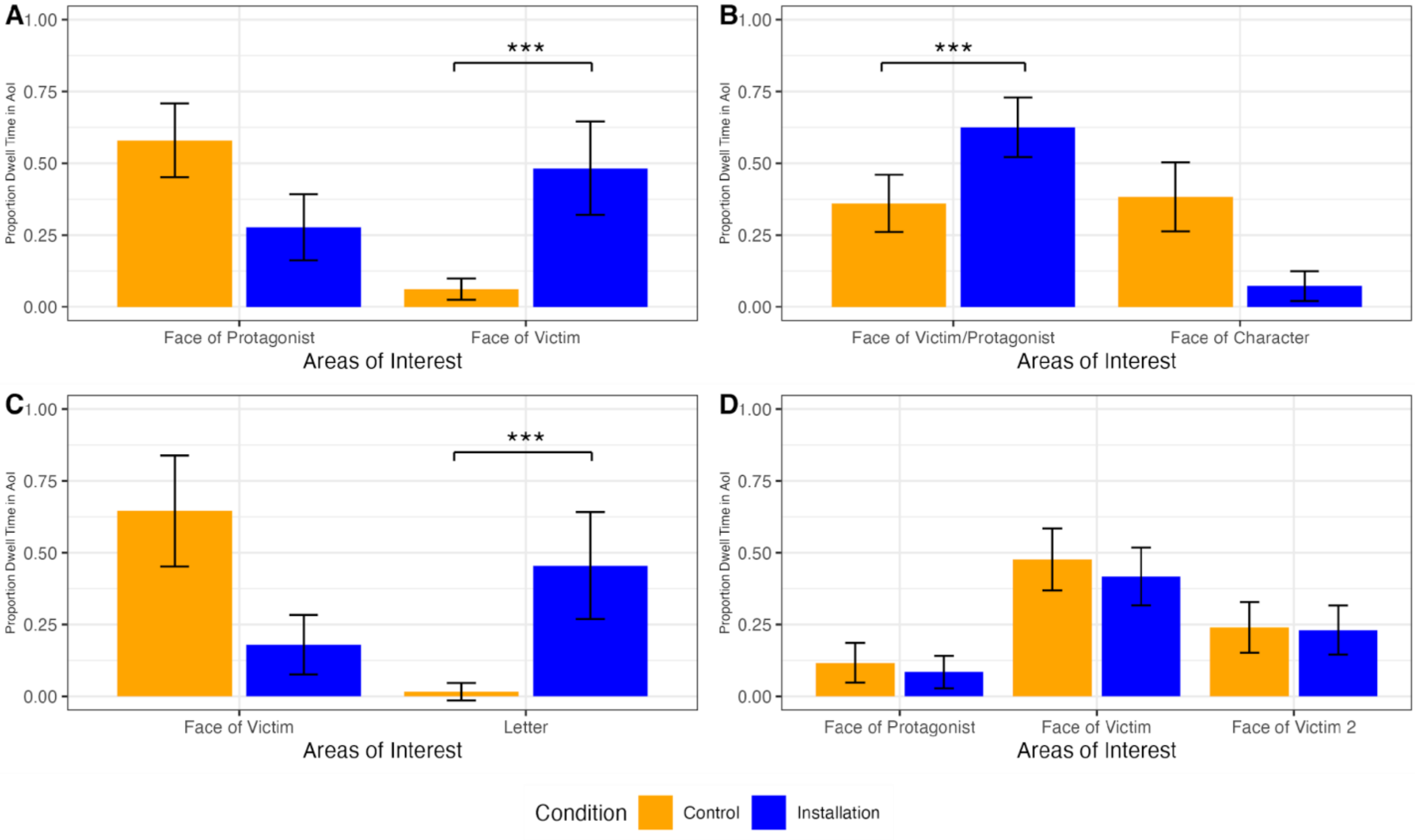


Figure 6. Proportions of Dwell Time Across AoIs and Condition. This figure includes four subplots (A-D), each showcasing the findings from: (A) Clip 1 - Never Weaken Part 1, (B) Clip 2 – The Freshman, (C) Clip 4 - Girl Shy, and (D) Clip 6 - The Kid Brother. Error bars represent 95% confidence intervals, with asterisks (***) marking statistically significant (($p < .001$)differences between conditions

Clip 4 – Girl Shy

A mixed ANOVA using the ART method indicated a significant interaction between condition and AoI: F(1, 93) = 4.698, p = .0328, η^2 = .048. No significant main effects were observed for condition (F(1, 93) = 0.000, p = .9937) or AoI (F(1, 93) = 0.496, p = .4832).

A Welch Two Sample t-test showed significantly greater dwell time on the letter in the Installation group (M = 0.4552) than in the Control group (M = 0.0162), t(24.19) = –4.8163, p < .001, d = 1.39 (see Figure 7 for a visualisation of group differences in dwell time).

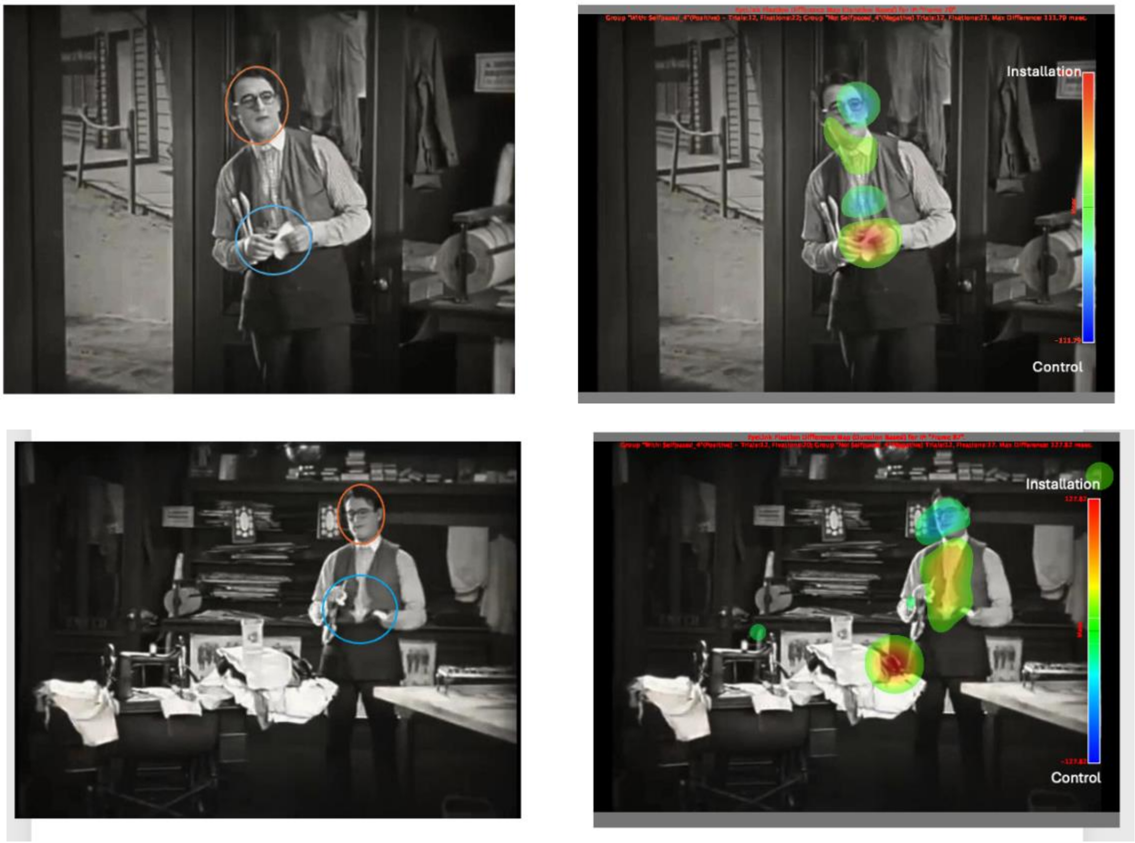


Figure 7. Comparative analysis of AoI mapping and difference fixation heat maps between installation and control groups for Frame 70 (top panel) and Frame 87 (bottom panel) of Clip 4. Colours show the fixation duration contrast: warmer colours indicate longer fixations by the Installation group (in comparison to the control group); cooler colours show longer fixations by the control group (in comparison the installation group).

Clip 3 (Never Weaken) – Exploratory Analysis

Employing a data-driven approach, similar to Haensel et al. (2020) and based on the precedent set by Võ et al. (2012), we visualised viewers' gaze patterns through heat maps to explore differences in attention to facial expressions. This observation led to a more focused analysis, where we focused on frames clearly portraying different emotions– initially happy when opening the door, shifting to surprise or shock upon seeing his fiancée with whom he believes to be her lover, and finally transitioning to sadness as he processes the situation.

An exploratory analysis was conducted on dwell time to the eyes vs. mouth regions in Clip 3 across three emotional expressions (happy, surprise, sad). Mixed ANOVAs revealed significant AoI × condition interactions for all three frames (Frame 5: $F(1, 44) = 12.59$, $p = .001$; Frame 48: $F(1, 44) = 12.77$, $p < .001$; Frame 62: $F(1, 44) = 12.43$, $p = .001$).

Post-hoc Wilcoxon tests indicated significant eye > mouth dwell time in the Control group across all expressions ($p < .01$), whereas no significant difference between regions in the Installation group. Full statistics are in Supplementary Table S4.

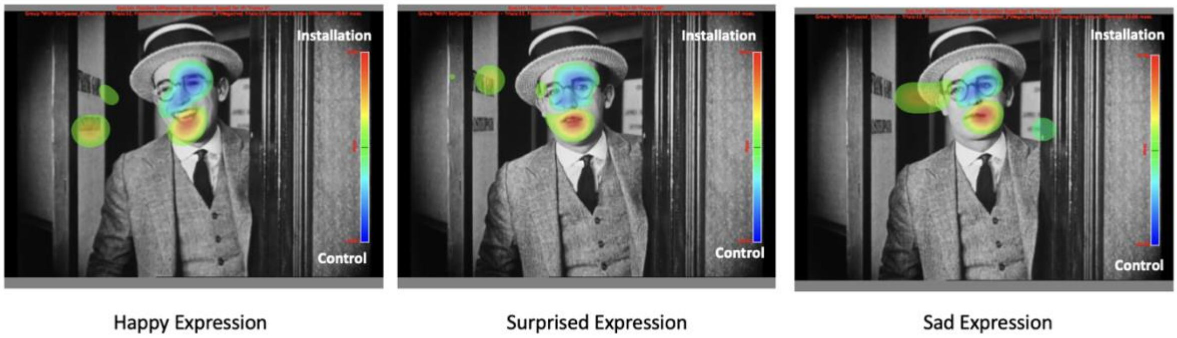


Figure 8. Difference Fixation Heat Maps corresponding to exploratory analyses conducted on the Exploitation Scene in Clip 4, where Harold undergoes a range of emotions due to a misunderstanding: happy (Frame 5), surprised or puzzled (Frame 48), and sad (Frame 62) respectively. Colours show the fixation duration contrast: warmer colours indicate longer fixations by the Installation group (in comparison to the *control* group); cooler colours show longer fixations by the *control* group (in comparison the *Installation* group).

4. Discussion

The current study examined whether exposure to narratively embedded false belief scenarios—via dramatic irony—elicits distinct patterns of cognitive processing compared to scenarios without such beliefs. We also investigated whether attention allocation reflects the construction and maintenance of SToM representations within viewers' event models. Using a self-paced viewing task and eye-tracking measures, we found that dramatic irony elicited both greater moment-specific cognitive effort and measurable shifts in visual attention. Participants exposed to false belief scenarios paused longer at specific frames and fixated longer on key elements, indicating that SToM attribution imposes representational demands that shape real-time engagement. These results demonstrate that spontaneous false reasoning during narrative comprehension is not only more cognitively taxing than true belief reasoning but also dynamically reflected in where viewers look and when. To our knowledge, this is the first study to demonstrate how spontaneous attribution of false beliefs dynamically shapes attentional allocation during scene perception, using naturalistic yet experimentally controlled stimuli.

Consistent with prior findings (Cabañas et al., 2023), viewers who saw the installation segment (Installation group) demonstrated significantly greater comprehension of DI scenarios than those in the control condition. These differences were used as a manipulation check that viewers formed qualitatively different event models based on the manipulated access to character knowledge.

RTs in the SPV task revealed that participants in the Installation group paused longer on specific frames within the exploitation scenes—despite all participants viewing the same static images. These increased dwell times support the notion that spontaneous attribution of false beliefs imposes higher cognitive demands at specific moments, as participants reconcile character ignorance with privileged narrative knowledge. In contrast, the control group, lacking such context, did not experience the same representational conflict. Notably, these effects were not uniformly distributed across all frames, as shown by the lack of main effect of the condition (Installation vs Control) but emerged specifically at moments requiring resolution of the belief conflict. This pattern suggests that it is not the mere presence of additional narrative information that increases cognitive effort, but rather the need to resolve the tension between conflicting representations at specific moments—highlighting the role of representational conflict in shaping moment-by-moment processing during narrative comprehension.

It is worth noting that the variation in RTs and eye-tracking differences across clips highlights the nuanced ways in which DI operates in film. Not all instances of DI impose the same cognitive demands; factors such as plot complexity or subtlety of cues influence how viewers process and understand DI. This variability underscores the need to consider specific narrative and cinematic features when assessing DI's cognitive impact. For example, Clip 3 relied heavily on medium close-ups, offering limited spatial and contextual information per frame—likely reducing the potential for differential processing between conditions during SPV.

Eye-tracking data further substantiated the heightened cognitive engagement in the Installation group, most likely as the results of SToM deployment. Across clips, longer fixation durations during the SPV phase were observed in this group, indicating greater processing load, while no differences emerged in exploratory metrics like fixation count, saccade amplitude, or gaze dispersion. This suggests viewers were using more cognitive resources to process information and integrate it into their mental models of the narrative ((Liu et al., 2022; Magliano & Zacks, 2011; Zacks & Swallow, 2007) found that perceptual load leads to shorter, more frequent fixations, while cognitive load increases fixation duration without affecting frequency. Our findings align with this, showing longer fixations but not more frequent ones—likely due to the SPV task. The mean SPV RT (379.41 ms, SD = 166.92) approximates a single fixation, potentially limiting additional fixations per frame. In contrast, (Payne et al., 2020) observed more exploration with static picture stories, which demand more visual search than our continuous-frame stimuli. This underscores how narrative format shapes processing and calls for further research into its cognitive effects.

Our findings build on those of Pedziwiatr et al., (2023), who showed that prior narrative context reduces exploratory gaze patterns in static film frames. While their work focused on general scene continuity, our study extends this effect to SToM contexts, demonstrating that belief-based discrepancies in character knowledge similarly shape real-time attention during narrative comprehension. Although we did not observe reduced exploratory behavior, our results also show that enriched narrative context—specifically, privileged knowledge in DI scenes—increases fixation durations, consistent with focal (detail-oriented) viewing. This supports event comprehension theories like SPECT (Loschky et al., 2020), which propose that event models guide moment-by-moment attention. Access to the installation segment enabled participants to form richer mental representations, which in turn directed gaze toward narratively critical elements. These results challenge the “Tyranny of Film” effect (Loschky et al., 2015) and align with Hutson et al., (2022), who found that reducing motion reveals comprehension-driven gaze differences.

Following the main analyses, we explored how viewers distributed attention to narratively relevant elements, especially the ‘victims’ of DI, by narrowing our analysis to specific frames flagged by significant differences in the SPV task. Results showed that the Installation group consistently prioritised these characters or objects in their gaze patterns. Clip 4 (‘Girl Shy’) illustrated this particularly well: viewers who knew the letter contained a cheque spent significantly more time fixating on it, unlike control participants, who misinterpreted its relevance. This object-focused attentional shift is especially notable, given that faces typically dominate gaze behavior in visual narratives (Smith, 2006). That viewers diverted attention away from facial features toward the letter strongly suggests that internal mental models, shaped by narrative context, actively guided their attentional strategy.

Importantly, not all clips yielded equally strong results. For instance, Clip 5 (For Heaven's Sake’) failed to elicit sufficient comprehension or RT effects and was excluded from further analysis. Clip 3 (‘Never Weaken’) also did not yield significant effects in the SPV analysis. However, given that this clip consisted entirely of medium close-ups, we considered whether the lack of variation in SPV might reflect limited scope for scene scanning, rather than an absence of attentional differences. Indeed, an exploratory heat map analysis revealed distinct patterns in gaze allocation. Control participants focused primarily on Harold’s eyes—an area typically associated with emotion recognition—whereas viewers in the Installation group distributed their gaze more evenly between the eyes and mouth. This broader scan pattern may reflect a shift toward inferring not just emotional, but also cognitive states such as misunderstanding or false belief. This interpretation aligns with prior findings by Cabañas et al., (2023), where participants in the Installation group made more references to characters’ beliefs and intentions in their narrative recall, emphasizing cognitive over affective mental state reasoning. It also resonates with Colombatto et al., (2019)'s concept of “mind contact,” which links increased attention to the mouth with the inference of communicative intent and goal-directed behavior. On the other hand, and consistent with these results, Sullivan et al., (2007) found that eye fixations are more strongly associated with decoding emotional states.

These findings contrast with prior research showing that the eyes and mouth serve distinct functions in emotion recognition: the eyes are typically fixated on in sad or angry expressions, whereas the mouth draws more attention in happy expressions (Beaudry et al., 2014; Calvo et al., 2018; Eisenbarth & Alpers, 2011; Guérin-Dugué et al., 2018; Neath-Tavares & Itier, 2016). However, most of these studies rely on static, decontextualised images. Work using naturalistic dynamic stimuli—such as film clips or face-to-face interactions—shows that gaze patterns flexibly shift depending on communicative context, including the presence or absence of speech (Haensel et al., 2020; Hessels, 2020; Võ et al., 2012). For instance, Võ et al., (2012) found that when dialogue was replaced by music, gaze shifted from the mouth to the eyes. In our silent clips, intertitles served as substitutes for speech. Control participants may have relied more on the eyes to integrate these cues, while Installation viewers—with their broader narrative knowledge—distributed gaze more widely, particularly to the mouth. This suggests that narrative context modulates facial processing, potentially enriching how viewers engage with both emotional and cognitive mental states.

This study offers new insight into how dramatic irony shapes real-time attention and cognitive effort, but several limitations should be acknowledged. First, the SPV paradigm, while effective in isolating top-down attentional processes, diverges from typical film viewing. Presenting static frames and allowing participants to control the pace may have altered natural gaze behavior. At the same time, this method reduced bottom-up visual saliency and helped reveal subtle differences in cognitive effort that might remain hidden in fully dynamic scenes. Future research should examine whether similar effects persist under more naturalistic continuous film viewing conditions. Second, while fixation duration served as an indicator of processing effort, it does not reveal the content or purpose of that effort. This highlights the importance of combining eye-tracking with recall and comprehension measures to triangulate cognitive processes and gain deeper insight into how viewers interpret and engage with narrative events. Lastly, the sample size limited our ability to explore individual differences in traits such as working memory or attentional control—an important direction for future studies.

In conclusion, our findings demonstrate that viewers spontaneously construct and maintain complex SToM models when engaging with dramatic irony in film. These mental models shape cognitive and attentional processing in real time, as reflected in increased viewing durations, longer fixations, and gaze shifts toward narrative-relevant characters and objects. These results contribute to a growing body of work on naturalistic social understanding and highlight the value of DI film structures in exploring spontaneous real-time mentalizing in richly contextualized, yet experimentally controlled, environments.

References

Apperly, I. (2010). *Mindreaders: The Cognitive Basis of 'Theory of Mind'*. Psychology Press.

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217. https://doi.org/10.1080/17470210903281582

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is Belief Reasoning Automatic? *Psychological Science*, *17*(10), 841–844. https://doi.org/10.1111/j.1467-9280.2006.01791.x

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251.

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., & Tapp, R. (2014). Featural processing in recognition of emotional facial expressions. *Cognition and Emotion*, *28*(3), 416–432. https://doi.org/10.1080/02699931.2013.833500

Cabañas, C., Senju, A., & Smith, T. J. (2023). The audience who knew too much: Investigating the role of spontaneous theory of mind on the processing of dramatic irony scenes in film. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1183660

Calvo, M. G., Gutiérrez-García, A., & Del Líbano, M. (2018). What makes a smiling face look happy? Visual saliency, distinctiveness, and affect. *Psychological Research*, *82*(2), 296–309. https://doi.org/10.1007/s00426-016-0829-3

Cane, J. E., Ferguson, H. J., & Apperly, I. A. (2017). Using perspective to resolve reference: The impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 591–610. https://doi.org/10.1037/xlm0000345

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3), 6–6. https://doi.org/10.1167/9.3.6

Chung, Y. S., Barch, D., & Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophrenia Bulletin*, *40*(3), 602–616.

Cohn, N. (2016). From Visual Narrative Grammar to Filmic Narrative Grammar: The narrative structure of static and moving images. In *Film Text Analysis* (pp. 94–117). Routledge.

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. *Cognitive Psychology*, *67*(3), 73–97. https://doi.org/10.1016/j.cogpsych.2013.07.002

Colombatto, C., van Buren, B., & Scholl, B. J. (2019). Intentionally distracting: Working memory is disrupted by the perception of other agents attending to you — even without eye-gaze cues. *Psychonomic Bulletin & Review*, *26*(3), 951–957. https://doi.org/10.3758/s13423-018-1530-x

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, *1*(2), 255–259. https://doi.org/10.1111/1467-7687.00039

Csibra, G., & Gergely, G. (2007). ‘Obsessed with goals’: Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*(1), 60–78.

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion*, *11*(4), 860.

Guérin-Dugué, A., Roy, R. N., Kristensen, E., Rivet, B., Vercueil, L., & Tcherkassof, A. (2018). Temporal dynamics of natural static emotional facial expressions decoding: A study using event-and eye fixation-related potentials. *Frontiers in Psychology*, *9*, 371926.

Haensel, J. X., Danvers, M., Ishikawa, M., Itakura, S., Tucciarelli, R., Smith, T. J., & Senju, A. (2020). Culture modulates face scanning during dyadic social interactions. *Scientific Reports*, *10*(1), 1958. https://doi.org/10.1038/s41598-020-58802-0

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. https://doi.org/10.1016/j.tics.2003.09.006

Hessels, R. S. (2020). How does gaze to faces support face-to-face interaction? A review and perspective. *Psychonomic Bulletin & Review*, *27*(5), 856–881. https://doi.org/10.3758/s13423-020-01715-w

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Huff, M., Maurer, A. E., Brich, I. R., Pagenkopf, A., Wickelmaier, F., & Papenmeier, F. (2017). Construction and Updating of Event Models in Auditory Event Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*. https://doi.org/10.1037/xlm0000482

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. https://doi.org/10.1111/cogs.13131

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), 46. https://doi.org/10.1186/s41235-017-0080-5

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41. https://doi.org/10.1016/s0010-0277(03)00064-7

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology (London, England: 1953)*, *99*(Pt 3), 317–340. https://doi.org/10.1348/000712607X251243

Kopatich, R. D., Feller, D. P., Kurby, C. A., & Magliano, J. P. (2019). The role of character goals and changes in body position in the processing of events in visual narratives. *Cognitive Research: Principles and Implications*, *4*(1), 22. https://doi.org/10.1186/s41235-019-0176-1

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science (New York, N.Y.)*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Lavandier, Y. (2005). *Writing drama: A comprehensive guide for playwrights and scriptwriters*. Le Clown & l'Enfant.

Levin, D. T., Hymel, A. M., & Baker, L. (2013). Belief, desire, action, and other stuff: Theory of mind in movies. In *Psychocinematics: Exploring cognition at the movies* (pp. 244–266). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199862139.003.0013

Liu, J.-C., Li, K.-A., Yeh, S.-L., & Chien, S.-Y. (2022). Assessing perceptual load and cognitive load by fixation-related information of eye movements. *Sensors*, *22*(3), 1187.

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. https://doi.org/10.1371/journal.pone.0142474

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.

Luke, S. G., & Henderson, J. M. (2016). The Influence of Content Meaningfulness on Eye Movements across Tasks: Evidence from Scene Viewing and Reading. *Frontiers in Psychology*, *7*, 257. https://doi.org/10.3389/fpsyg.2016.00257

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, *15*(5), 533–545. https://doi.org/10.1002/acp.724

Magliano, J. P., Yan, E. F., Ackerman, T., Mccarthy, K. S., & Kurby, C. A. (2024). *Understanding the Role of Cinematic Features on the Experience of Filmed Events*. https://doi.org/10.3167/proj.2024.180302

Magliano, J. P., & Zacks, J. M. (2011). The Impact of Continuity Editing in Narrative Film on Event Segmentation: Cognitive Science. *Cognitive Science*, *35*(8), 1489–1517. https://doi.org/10.1111/j.1551-6709.2011.01202.x

Meins, E., Fernyhough, C., & Harris-Waller, J. (2014). Is mind-mindedness trait-like or a quality of close relationships? Evidence from descriptions of significant others, famous people, and works of art. *Cognition*, *130*(3), 417–427. https://doi.org/10.1016/j.cognition.2013.11.009

Neath-Tavares, K. N., & Itier, R. J. (2016). Neural processing of fearful and happy facial expressions during emotion-relevant and emotion-irrelevant tasks: A fixation-to-feature approach. *Biological Psychology*, *119*, 122–140. https://doi.org/10.1016/j.biopsycho.2016.07.013

O'Neill, D. K., & Shultis, R. M. (2007). The emergence of the ability to track a character's mental perspective in narrative. *Developmental Psychology*, *43*(4), 1032–1037. https://doi.org/10.1037/0012-1649.43.4.1032

Papenmeier, F., Boss, A., & Mahlke, A.-K. (2019). Action goal changes caused by agents and patients both induce global updating of event models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(8), 1441–1454. https://doi.org/10.1037/xlm0000651

Payne, K. B., Smith, M. E., Hutson, J. P., Magliano, J. P., & Loschky, L. C. (2020). Eye movements reveal event understanding in visual narratives. *Journal of Vision*, *20*(11), 1645–1645.

Pedziwiatr, M. A., Heer, S., Coutrot, A., Bex, P., & Mareschal, I. (2023). Prior knowledge about events depicted in scenes decreases oculomotor exploration. *Cognition*, *238*, 105544. https://doi.org/10.1016/j.cognition.2023.105544

Perner, J. (1991). *Understanding the representational mind* (pp. xiv, 348). The MIT Press.

Persson, P. (2003). *Understanding Cinema: A Psychological Theory of Moving Imagery*. Cambridge University Press.

Quesque, F., Apperly, I., Baillargeon, R., Baron-Cohen, S., Becchio, C., Bekkering, H., Bernstein, D., Bertoux, M., Bird, G., Bukowski, H., Burgmer, P., Carruthers, P., Catmur, C., Dziobek, I., Epley, N., Erle, T. M., Frith, C., Frith, U., Galang, C. M., … Brass, M. (2024). Defining key concepts for mental state attribution. *Communications Psychology*, *2*(1), 29. https://doi.org/10.1038/s44271-024-00077-6

Rabkina, I., & McFate, C. J. (2022). Efficient and Effortful Theory of Mind Reasoning in the AToM Cognitive Model. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://escholarship.org/uc/item/8tx0r50x

Rapp, D. N., & Gerrig, R. J. (2006). Predilections for narrative outcomes: The impact of story contexts and reader preferences. *Journal of Memory and Language*, *54*(1), 54–67. https://doi.org/10.1016/j.jml.2005.04.003

Rooney, B., & Bálint, K. E. (2018). Watching More Closely: Shot Scale Affects Film Viewers' Theory of Mind Tendency But Not Ability. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02349

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Senju, A. (2012). Spontaneous Theory of Mind and Its Absence in Autism Spectrum Disorders. *The Neuroscientist*, *18*(2), 108–113. https://doi.org/10.1177/1073858410397208

Smith, T. J. (2006). *An Attentional Theory of Continuity Editing*. https://era.ed.ac.uk/handle/1842/1076

Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, *6*(1), Article 1.

Smith, T. J. and Mital, P. K (2013) Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision, 13* (8). ISSN 1534-7362

Sullivan, S., Ruffman, T., & Hutton, S. B. (2007). Age differences in emotion recognition skills and the visual scanning of emotion faces. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *62*(1), P53–P60.

Symeonidou, I., Dumontheil, I., Chow, W.-Y., & Breheny, R. (2016). Development of online use of theory of mind during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology*, *149*, 81–97.

Võ, M. L.-H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, *12*(13), 3. https://doi.org/10.1167/12.13.3

Westra, E. (2017). Spontaneous mindreading: A problem for the two-systems account. *Synthese*, *194*(11), 4559–4581. https://doi.org/10.1007/s11229-016-1159-0

Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., & Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. *NeuroImage*, *47*(1), 360–366. https://doi.org/10.1016/j.neuroimage.2009.04.037

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition*, *13*(1), 103–128.

Zacks, J. M., & Swallow, K. M. (2007). Event Segmentation. *Current Directions in Psychological Science*, *16*(2), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x

Supplementary Materials

Supplementary Text 1

Description of Clips

Clip 1 - Never Weaken Part 1:

Establisher: Harold wants to help his friend or love interest, who will get fired from the osteopathic clinic where she works if they do not get more clients.

Installation: Harold runs into an acrobat and they come up a plan together.

Exploitation: The acrobat pretends to fall and get injured in front of a crowd of potential clients and Harold pretends to be an osteopath who can help him. When the acrobat seems to heal quickly, the crowd gets excited and asks for business cards for the osteopathic clinic.

Victim of dramatic irony: the people in the crowd whose goal is to get a treatment for their illnesses or disabilities.

Clip 2 - The Freshman:

Establisher: Harold out for the team to be a football player

Installation: After tryouts, Harold is unaware that he did not make the team. The coach and teammate trick him into thinking he made the team but assign him the role of water boy.

Exploitation: Harold excitedly tells a girl that he made the team and eagerly goes to the field to play because he thinks he is on the team, unaware that he is really just the water boy.

Victim of dramatic irony: Harold whose goal is to be a football player.

Clip 3 - Never Weaken Part 2:

Establisher: Harold proposes to a girl, and she accepts.

Installation: In another scene, the girl is talking to a man who is actually her brother, and he's an ordained minister.

Exploitation: Harold overhears the man offering to marry the girl, not realizing that he's actually her brother. Harold thinks the man wants to marry her, not to officiate the wedding, so he's visibly upset and walks away.

Victim of dramatic irony: Harold whose goal is to marry the girl.

Clip 4 - Girl Shy:

Establisher: Harold visits a publishing house to inquire about the possibility of publishing his book. However, the publisher finds his book to be extremely comical, so they reject it and inform Harold that he will receive a rejection letter in the mail.

Installation: When Harold leaves, a senior employee convinces the editor to reconsider and publish the manuscript as a comedy. He then instructs the employee to send a check to Harold instead of the rejection letter.

Exploitation: Harold, downhearted and unaware of the content of the letter, tears it apart without opening it.

Victim of dramatic irony: Harold whose goal is to publish his book and make money with it.

Clip 5 - For Heaven's Sake:

Establisher: A missionary and his daughter need to raise money for a homeless mission and decide to write a letter to Harold, who is portrayed as a wealthy man.

Installation: Harold accidentally burns down a cart that belongs to the missionary.

Exploitation: Harold wants to pay for the burnt cart, so he writes a check for the missionary, who doesn't know that Harold caused the accident. The missionary and his daughter are grateful and think that Harold made the donation on purpose to help the mission.

Victim of dramatic irony: the missionary and his daughter, whose goal is to raise money for the mission.

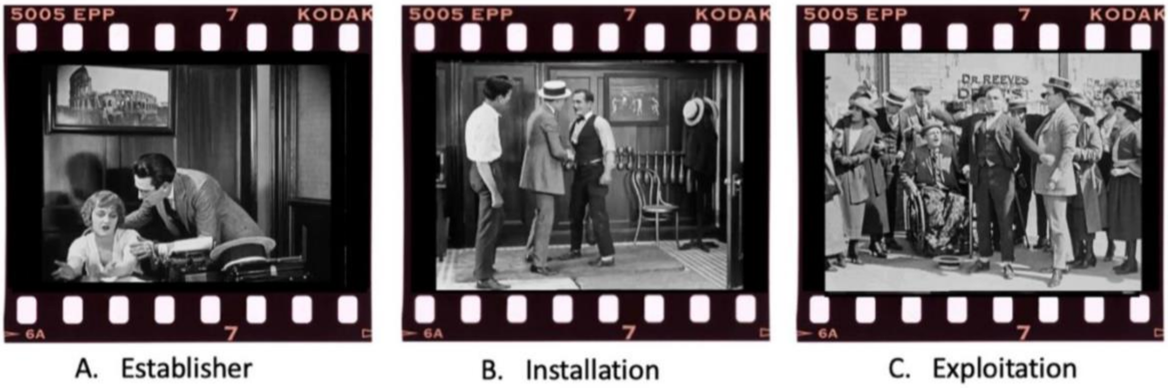
Clip 6 - The Kid Brother:

Establisher: Two men running a traveling show are seeking a permit from the sheriff.

Installation: Harold feels inferior to his brothers and is not allowed to go to town with them. He stays at home and dresses up as his father, who is the sheriff.

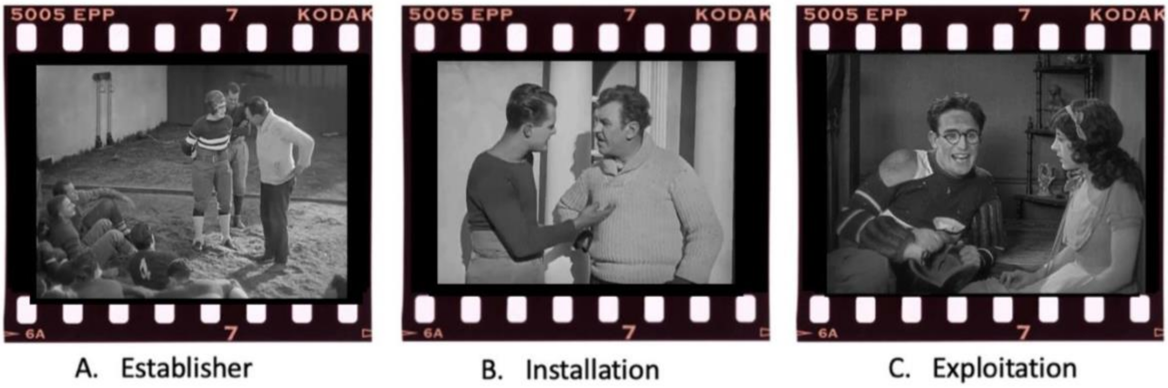
Exploitation The two men see Harold dressed as sheriff and convince him to sign the permit.

Victim of dramatic irony: The two men, whose goal is to get the permit for their show.



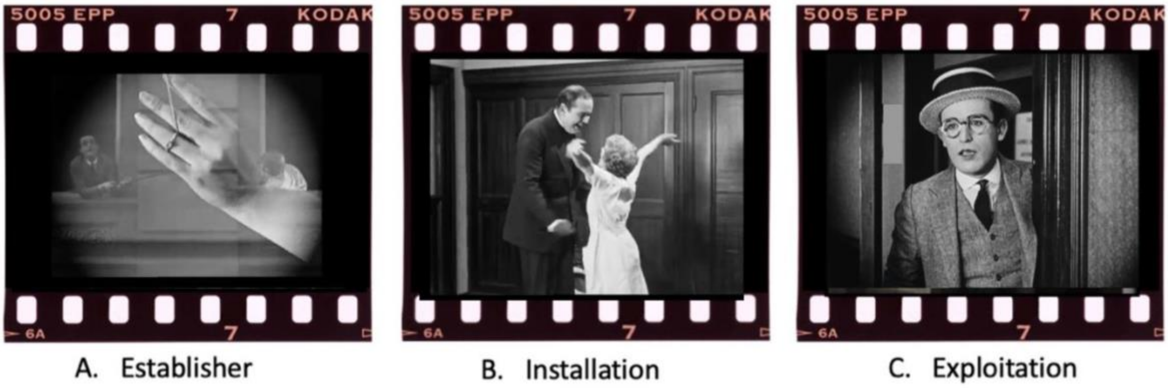
A. Establisher B. Installation C. Exploitation

Supplementary Figure 1. Example stills from Clip 1. Stills taken with permission from Never Weaken (Copyright of the Harold Lloyd Trust, 1921).



A. Establisher B. Installation C. Exploitation

Supplementary Figure 2. Example stills from Clip 2. Stills taken with permission from The Freshman (Copyright of the Harold Lloyd Trust, 1925).



A. Establisher B. Installation C. Exploitation

Supplementary Figure 3. Example stills from Clip 3. Stills taken with permission from Never Weaken (Copyright of the Harold Lloyd Trust, 1921).

A. Establisher B. Installation C. Exploitation

Supplementary Figure 5. Example stills from Clip 5. Stills taken with permission from For Heaven's Sake (Copyright of the Harold Lloyd Trust, 1926).

A. Establisher B. Installation C. Exploitation

Supplementary Figure 6. Example stills from Clip 6. Stills taken with permission from The Kid Brother (Copyright of the Harold Lloyd Trust, 1927).

Supplementary Text 2

Deviations from Pre-registered Analysis

The analysis plan for the reaction times of the SPV task was prespecified in a preregistration document that can be found on the Open Science Framework [https://osf.io/by56s]. In order to investigate how deeper cognitive processing is reflected in participants' gaze and pupil size, we also collected eye-tracking and pupillometry data. We pre-specified that the results from the SPV paradigm's reaction time analysis would inform the secondary analysis of eye-tracking data. However, due to the complexities of individual and frame-specific varying reaction times and slow changes in pupil size, pupillometry data was not analysed in this study. Moreover, motion stimuli (film viewing) can induce pupillary constriction as a confounding variable, making it advisable to use shorter length clips than those available in our film clip corpus (Mathôt & Vilotijević, 2022). Moreover, controlling for these factors becomes more challenging in the SPV paradigm, where the length of each trial is variable.

In accordance with the pre-registration and the outlined procedural plan for the current study, the Reading the Mind in the Eyes Test (RMET) was administered. However, for logistic reasons it had to be administered following the eye-tracking task. This decision was methodologically strategic; conducting the RMET after the primary task aimed to mitigate potential priming effects that might influence participants' responses during the eye-tracking session. Nevertheless, this procedural configuration presented an interpretative challenge for the RMET scores. The cognitive and emotional engagement with the eye-tracking task might have impacted participants' performance on the subsequent RMET, leading to results that reflect the residual effects of the task rather than baseline mentalizing abilities. Given this interpretive ambiguity, the decision was made not to analyse the RMET responses. Nonetheless, to ensure that the sample did not include individuals with poor mentalizing capabilities—which could confound the primary analysis—RMET scores were reviewed to confirm that all participants exceeded the established threshold for adequate mentalizing abilities. This precautionary measure allowed us to maintain confidence in our sample's representativeness regarding typical mentalizing skills. Data management and statistical analyses were performed using the statistical programming language R in R-studio.

Supplementary Table 1. Summary of Step-Wise Linear Mixed Effects Model Analyses Across All Clips.

Clip Nº.	Final Model	AIC	BIC	logLik	L.Ratio	p-value
Clip 1	random_intercept_ interaction_clip_1	3531.052	6947.176	- 1256.526	294.121	0.0351*
Clip 2	random_intercept_ interaction_clip_2	2135.644	6586.264	-426.822	402.394	0.0009***
Clip 3	random_intercept_ interaction_clip_3	1030.415	1800.985	- 373.2075	81.4741	0.1446
Clip 4	random_intercept_ interaction_clip_4	1590.056	2983.992	- 560.0278	203.0367	<.0001***
Clip 5	random_intercept_ interaction_clip_5	2056.135	3816.559	- 741.0676	142.4028	0.4511
Clip 6	random_intercept_ interaction_clip_5	2364.051	5094.084	-763.026	241.721	0.0493*

Note. The table details the model comparisons, degrees of freedom (df), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood (logLik), test comparisons, likelihood ratio (L.Ratio), and p-values.

Supplementary Table 2. Summary of Stepwise Linear Mixed Effects Model Analyses for Each Clip.

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
baseline	2	35079.71	35096.01	-17537.857			
random_intercept	3	18638.62	18663.07	-9316.312	1 vs 2	16443.090	<.0001
random intercept_clip	4	9799.26	9831.85	-4895.632	2 vs 3	8841.360	<.0001
random_intercept condition	5	9800.63	9841.37	-4895.315	3 vs 4	0.634	0.4257
random_intercept + condition_clip	10	17624.16	17705.64	-8802.08	4 vs 5	7813.53	<.0001
random_intercept + condition_interaction	15	17465.33	17587.54	-8717.67	5 vs 6	168.83	<.0001

Note. The table presents the final model for each clip, alongside key statistical metrics including Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood (logLik), likelihood ratio (L.Ratio), and p-values. Significance levels are denoted with asterisks: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Supplementary Table 3

Contrast Analysis for Clip 1

Frame	Estimate	SE	df	t-ratio	p-value
2	-0.64357	0.213	22	-3.026	0.0062*
22	-0.47060	0.213	22	-2.213	0.0376*
55	-0.61968	0.213	22	-2.914	0.0080*
148	-0.46862	0.213	22	-2.203	0.0383*
192	-0.52373	0.213	22	-2.463	0.0221*

Supplementary Table 4

Contrast Analysis for Clip 2

Frame	Estimate	SE	df	t-ratio	p-value
1	-0.536746	0.201	22	-2.668	0.0140*
16	-0.422440	0.201	22	-2.100	0.0474*
19	-0.439594	0.201	22	-2.185	0.0398*
43	-0.453654	0.201	22	-2.255	0.0344*

Supplementary Table 5

Contrast Analysis for Clip 4

Frame	Estimate	SE	df	t-ratio	p-value
1	-0.76376	0.204	22	-3.739	0.0011***
70	-0.52290	0.204	22	-2.565	0.0177*
87	-0.48507	0.204	22	-2.375	0.0267*

Supplementary Table 6

Contrast Analysis for Clip 6

Frame	Estimate	SE	df	t-ratio	p-value
1	-0.489428	0.204	22	-2.396	0.0255*
2	-0.633976	0.204	22	-3.103	0.0052*
77	-0.514155	0.204	22	-2.517	0.0196*
184	0.495020	0.204	22	2.423	0.0241*

*p < .05, **p < .01, ***p < .001