frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 28 March 2024
DOI 10.3389/fcomm.2024.1347788

OPEN ACCESS

EDITED BY
Claudia Lehmann,
University of Potsdam, Germany

REVIEWED BY
Karl-Heinrich Schmidt,
University of Wuppertal, Germany
Assimakis Tseronis,
Örebro University, Sweden

*CORRESPONDENCE
Dayana Markhabayeva
✉ dayana1@uni-bremen.de

RECEIVED 01 December 2023
ACCEPTED 08 March 2024
PUBLISHED 28 March 2024

CITATION
Markhabayeva D and Tseng C-I (2024) Multimodal cohesion and viewers' comprehension of scene transitions in film: an empirical investigation.
Front. Commun. 9:1347788.
doi: 10.3389/fcomm.2024.1347788

Multimodal cohesion and viewers' comprehension of scene transitions in film: an empirical investigation

Dayana Markhabayeva[1]* and Chiao-I Tseng[2]

[1]Faculty of Linguistics and Literary Studies, University of Bremen, Bremen, Germany, [2]Department of Applied Information Technology, University of Gothenburg, Gothenburg, Sweden

This paper presents three empirical studies that unravel how the devices of multimodal cohesion support viewers' narrative interpretation of scene transitions in film. The linguistics-informed method of cohesion analysis in film uncovers the establishment of cohesive ties between characters, objects, settings and characters' actions. Previous studies using eye-tracking and comprehension tests already indicate the significance of multimodal cohesion in people's comprehension of background settings within a continuous scene. The present paper investigates further whether film cohesion impacts viewers' story comprehension across different scenes and settings. Moreover, it also explores whether the spatio-temporal relations between scenes is a significant factor, along with cohesive devices, in viewers' scene comprehension. Methodologically, we create contrasting film situations by manipulating cohesion structures and spatio-temporal orders of scenes. Our comparative analyses of viewers' comprehension of these different film situations reveal that the presence of cohesive cues significantly can influence viewers' accurate scene comprehension. Through testing the inter-relation of cohesion, spatio-temporal order, characters' intention and viewers' time perception, this paper offers new avenues for further exploration of space, time and coherence in film.

KEYWORDS

film, cohesion, scene transition, multimodal discourse analysis, narrative comprehension, spatio-temporal relations

1 Introduction

In the history of film studies, the ways of how the viewers are carried from shot to shot and scene to scene as well as the effects of different types of scene transitions have been frequently investigated. Eisenstein (1969) explored how filmmakers' combinations of different shots and scenes can lead the viewers to interpret meaning in particular ways . In the 1970s, transitions of shots and scenes were systematically analyzed by the semiotician Christian Metz. Metz (1974) *Grande Syntagmatique* proposes eight types of cinematic syntax, namely, the transitions between film shots and scenes. Following the pursuit of Metz to create a generalized modeling of scene and shot transitions, the recent study of Bateman (2007) and Bateman and Schmidt (2012) proposed a *Grande Paradigmatique*, which maps out a more comprehensive set of semantic relations between different shots such as different types of spatio-temporal and logical relations.

While the semiotic theories by Bateman focus on shot-based semantic relations, other scholars have explored another type of mechanism, namely, *cohesion*, addressing how verbal, visual and audio elements *within* film shots are tied together to signal the coherent flow of film narratives across scene changes.

Cohesion is originally a linguistic concept. It refers to a set of semantic relations in text which enable the interpretation of meaning coherence. In text linguistics (Halliday and Hasan, 1976), text as a coherent whole is the result of cohesive devices at work. Halliday and Hasan (1976, p.12) posits that "cohesion is a relational concept; it is not the presence of a particular class of item that is cohesive, but the relation between one item and another".

In the context of cinema, the film theorist Bordwell (2008) provides an exploratory account of how different types of audiovisual cohesion function to carry viewers across scene transition and how patterns of film cohesion unravel viewers cognitive comprehension activities. Bordwell exemplifies, for instance, how cohesive relation is established when the re-occurrence of a same object or characters in two different scenes cues the viewer to interpret the coherent narrative flow. The recent empirical studies of scene comprehension (Loschky et al., 2015b) also investigate the elements within and across shots that lead to coherent scene perception. This paper, employing Bordwell's definition of *scene* and insights from cognitive studies, examines scene transition as a disruption in space and time, namely, a change of event location and the break of continuous events.

Systematically employing the linguistic concept of textual cohesion, the two works of Tseng (2012, 2013) extend Bordwell's attempt to identify internal cohesive structure and propose a more systematic framework of film cohesion. In linguistic analysis, the analytical tools of cohesion are used to describe the "repetition" and "re-occurrences" of linguistic patterns, with which a text holds itself together as a unit of communication. Along the same lines, the multimodal cohesion analysis unravels how the "repetition" and "re-occurrences" of narrative elements such as people, places, objects and actions, whether identified in the visualtrack (e.g., visible figures or as written names on the screen) or in the audiotrack (e.g., spoken names or sounds and music that represent certain identities), are cued to the viewers for interpreting the narrative coherence within and across shots.

While the framework of multimodal cohesion has been applied to other media such as TV series, comics, other graphic novels (Tseng and Bateman, 2018; Tseng et al., 2018; Drummond and Wildfeuer, 2020) and interactive narratives (Tseng and Thiele, 2022), there has not been sufficient empirical investigations of just which verbal, visual and audiovisual cohesive cues in film play the dominant, pivotal role of facilitating the seamless connection between the storytelling units.

One of the first empirical attempts for triangulating the multimodal cohesion framework and the viewer's cognition and memory is conducted by Tseng et al. (2021). The authors use comprehension tests and eye-tracking experiments to compare viewers' attention and narrative interpretation of film sequences, either with or without cohesive cues crucial for the viewers' comprehension of the specific settings within a scene. They use film sequences extracted from *The Birds* and a *Monty Python* sketch for the experiments and their findings indicate the significant role of cohesive devices for the viewers' narrative comprehension and gaze-behavior. The findings open up more questions as to whether cohesion still plays a role in the more complex scene transitions, whether cohesion in film influences the way the viewers interpret the continuity of spatio-temporal and logical relations as those theorized by Bateman (2007) and whether cohesion and spatio-temporal relations are related to viewers' perception of intention and time. This paper will precisely extend the previous empirical endeavor to address these open questions.

The paper is structured as follows: Section 2 exemplifies the analysis of multimodal cohesion and how the cohesive structures, termed *cohesive chains*, reflect the viewers understanding of the presentation and re-occurrences of characters, objects and settings. Drawing on the multimodal cohesion analysis, Section 3 presents the empirical studies we conduct to triangulate the multimodal linguistic framework with the viewer's cognitive process. Several cognitive studies have endeavored to address how the audiences' coherent narrative comprehension is steered by film narrative and technical features such as continuity editing (Smith, 2012), event recognition (Zacks, 2015) and scene construction (Loschky et al., 2015a). Our studies of multimodal cohesion complement the previous cognitive studies through providing a semiotically formulated model of interpretation. As we will see in the next section, this semiotic-textual level of analysis offers a more fine-grained yet systematic investigation interconnecting the functions of film technical features, narrative elements, semantic structures and the overall contextual coherence.

2 Analysing multimodal cohesion in film

The framework of multimodal cohesion (Tseng, 2013) provides a powerful discourse semantics for examining cohesive ties between film elements within and across shots and scenes. It was formulated drawing on the discourse semantic model of identification, which was developed for the analysis of natural language (Martin, 1992). In the linguistic analysis, the choices of the identification system realize the identity presentation and re-occurrence of people, places and things throughout a text. The structures of identification, namely, how relevant people, places and things are actually tracked, then highlight the textually constructed unity of any particular text. Tseng (2013) applied the discourse semantic framework to film. In this way, the framework captures not only the area of semiotic work shared across language and film but also the differentiation of the filmic cohesion analysis from the linguistic analysis.

The multimodal cohesion system developed for film is shown in Figure 1 represented as a system network. System networks are used in systemic functional linguistics to show the abstract paradigmatic "choices" available for language users drawn from the meaning potential of their language (Halliday and Matthiessen, 2013). In the film system, the network in Figure 1 shows the functional mechanisms for cuing identities of characters, objects and settings as a film unfolds. In the system network, contrasting options are collected together into individual systems of choice: for instance, in the system of [presenting/presuming], only one of the two features may be selected at a time. Certain feature selections then also lead on to finer classifications. For example, in the case of the choice [gradual], the system leads on to a further dependent, i.e., finer, choice between [dynamic] and [static]. It is also possible for several dimensions of classification to be pursued in *parallel*: such systems are called simultaneous systems and are grouped with a curly right-facing bracket. In Figure 1, for example, choices need to be made

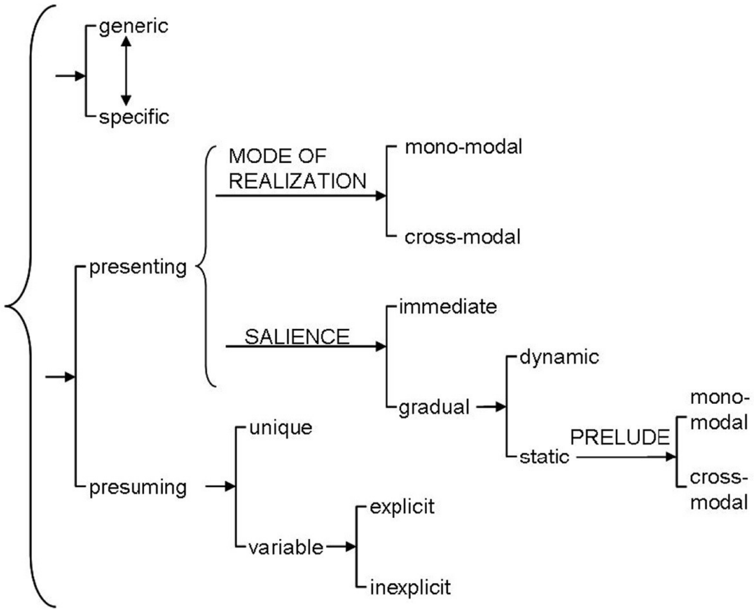


FIGURE 1
The system of multimodal cohesion in film developed by Tseng (2013).

from the features presented by *both* the systems [generic/specific] and [presenting/presuming] for a complete description.

We exemplify the process of constructing a multimodal cohesive analysis of film using a scene of Nolan's (2000) *Memento*. The film is a thriller, depicting the main character, Leonard Shelby, an insurance investigator, suffers from short-term memory loss and uses notes and tattoos to hunt for the man he thinks killed his wife, which is the last thing he remembers. The segment we exemplify here is a scene when Leonard goes into a tattoo shop. We also employ this segment below in one of our experimental studies.

Figure 2 shows the scene transition from street view to the indoor setting of a tattoo shop. It includes selected shots that can best depict the location transition across the outdoor and indoor scenes. Shot 1a shows the front door of a shop. A small orange sign at the bottom left shows it is a tattoo shop. Within the same shot, a car is seen and heard squeaking and stopping abruptly in front of the shop (shot 1b). It is then cut to shot 2, the closeup of Leonard, who is seen looking at a white object. The point-of-view shot in shot 3 then shows the note he is reading, with *Tattoo Fact 6: car license* written on it. A closeup in shot 4 shows the shop's name, *Emma's TATTOO*. It is followed by the transition into the indoor scene of the shop. Shot 5 shows the closeup of the tattooist's hands tattooing the same written text of the note seen in shot 3 on someone's skin. Shot 6 and shot 7 reveal that Leonard is the one being tattooed. Throughout shots 5 to 7, the audience can hear the continuous tattooing sounds. The second character, Teddy, entered the room in shot 8, greeting Leonard: "Lenny!". While in shot 9, Leonard lifts his head, seeming not remembering who the person is, in shot 10, the tattooist yells at Teddy: "It is private back here. Wait out there". In shot 11, Teddy looks frustrated but goes out to wait in another room in the same tattoo shop, shown in shot 12. The same tattoo setting is suggested by the background tattoo images and symbols on the walls of the room. In shot 13, Leonard and the tattooist both came out. Shots 14 and 15 construct shot/reverse shot showing the conversation between the two characters in the room.

Focusing for the purposes of illustration on the setting of *Tattoo shop*, we can describe the cohesive devices for presenting and tracking the tattoo shop based on the instantiation of features from the system network of Figure 1.

In shot 1, the front door of a shop is seen from a street view. For the viewers who notice the *Tattoo* sign written on the orange board at the left corner, the specific identity of the tattoo shop is immediately established. This is therefore a case of [presenting] rather than [presuming]. As the shop is specified as Tattoo shop right at the outset, the cohesive devices at work are therefore [specific] from the continuum [generic - specific] and [immediate] salience.

In shot 3, the written text *tattoo* on the paper held in Leonard's hands is the multimodal re-occurrence of the tattoo shop. Although *tattoo* does not directly refer to the shop setting, it is a *hyponym* of tattoo shop, cohesively related to the previously seen tattoo setting. Hence, this is the case of [presuming] that track the same identity of tattoo shop.

Similarly, in shot 4, the front door of the shop with the shop name *Emma's Tattoo* cohesively cues the viewers back to the tattoo note. Here the cohesive devices [presuming] and [explicit] reappearance are at work to track of the tattoo shop.

From shot 5 onwards, re-occurrences of the theme *Tattoo* and *Tattoo shop* are visualized through more sets of multimodal elements: the female tattooist's tattooing Leonard's thigh, the continuing tattooing sounds and the background tattoo pictures in the room from shots 12 to 15.

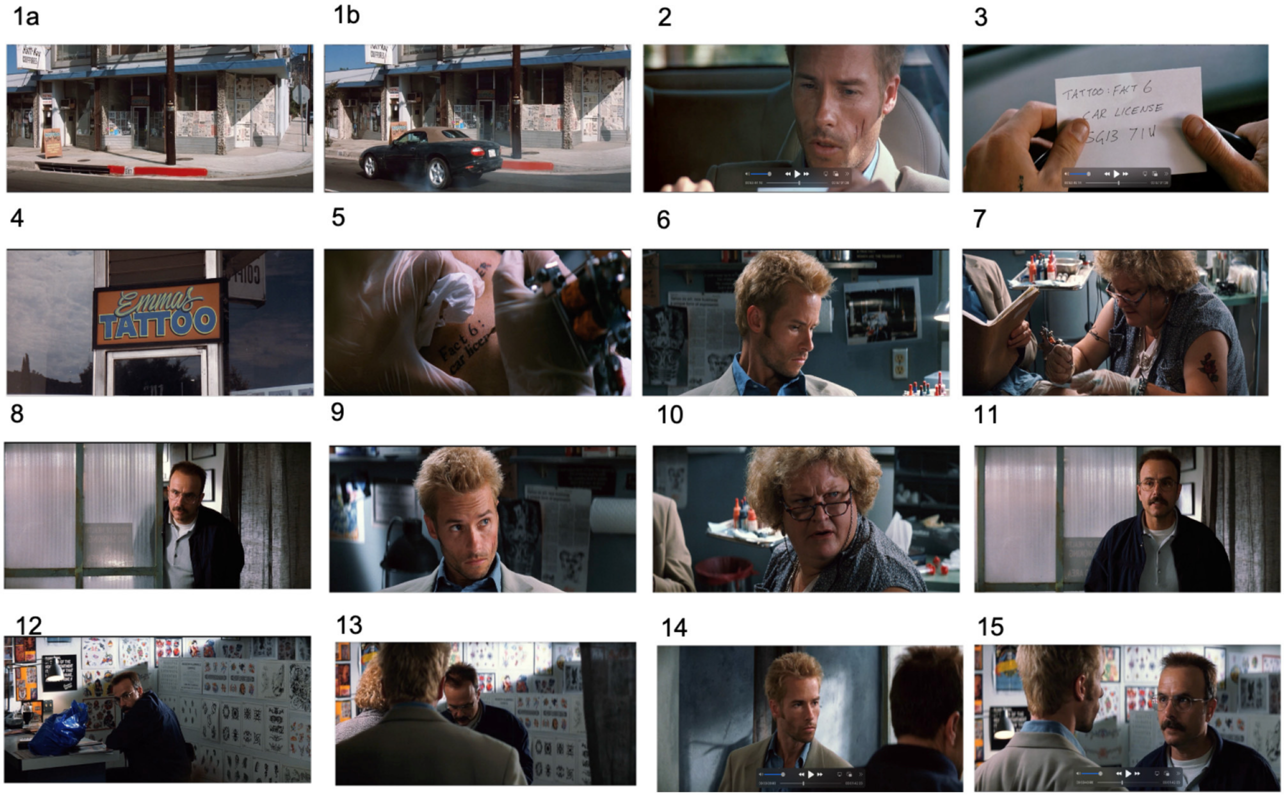


FIGURE 2
Selected shots of the tattoo shop scene in *Memento*.

In other words, all these verbal, visual and audio cues are cohesively tied together to signal the concept of tattoo/tattoo place; and each re-occurrence of a cohesive element is related to preceding occurrences by specifically labeled cohesive ties showing the tracking strategy involved.

Whereas cohesive ties relate pairs of cohesive elements, sequences of element re-occurrences and the classified cohesive ties between those occurrences are structured into *cohesive chains*, which show textual development of narratively significant characters, objects and settings across larger portions of film sequences. The overall cohesive chains structured from the tattoo shop scene is displayed in Figure 3. Here we can see that the tattoo shop chain interlinks the multimodal realization of the elements we discussed above. This chain starts with *Tattoo* in written text and the visual figure of the tattoo shop, annotated as [V]. It is then linked by an upward pointing arrow from both written text *Tattoo* in shot 3. The arrows refer to the semantic relation of anaphora, which ties the pairs of cohesive narrative elements together. Along the same lines, the continuing multimodal cohesive chain links together the written text *Emma's Tattoo*, audio (tattooing sounds) and visual (indoor shop) elements of the setting *tattoo shop*.

Similar to the cohesive chain of *tattoo shop*, the chain of Leonard shows how this character is visually presented (in visual image [v] in shot 2) but reoccurred multimodally in the following shot—in shot 8, his identity is realized in spoken text when Teddy called his name. Moreover, the cohesive chain of *car* also shows a multimodal presentation of the object (shot 1b) – it is not only seen but also heard when the car breaks with a squeaking sound. Hence, the first element of the *car* chain is annotated as [v](squeaking).

Moreover, in research work on verbal texts (Hasan, 1984) as well as film (Tseng, 2008), it has been observed that such chains and, in particular chain *interactions*, appear to be more revealing of a text's organization than elements that occur in relative isolation. Interactions between chains occur whenever elements of distinct chains are brought together within the depiction of a single action or event. Thus, although any element in a textual artifact typically enters into a large number of cohesive links with other elements, it is the elements participating in chain interactions that are constructed as being textually "significant". This constructs a useful method for selecting from all the cohesive ties potentially available in a text just those collections of ties that are hypothesized to be most likely to play a role in guiding the viewers' narrative interpretation. That is, a viewer does not need to attend to "everything" that is audio-visually on offer, but rather will be guided to attend to those elements that contribute to interacting chains. For example, before the scene transition across shot 4 and 5, namely, from the outdoor to the indoor scenes, the cohesive chains of *street* and *car combine/interact* with the chains of character Leonard and setting *tattoo shop* to construct a coherent event that might be glossed in natural language as follows:

> On a *street* in front of a *tattoo shop*, a *car* driven by *Leonard* stops in front of the shop door.

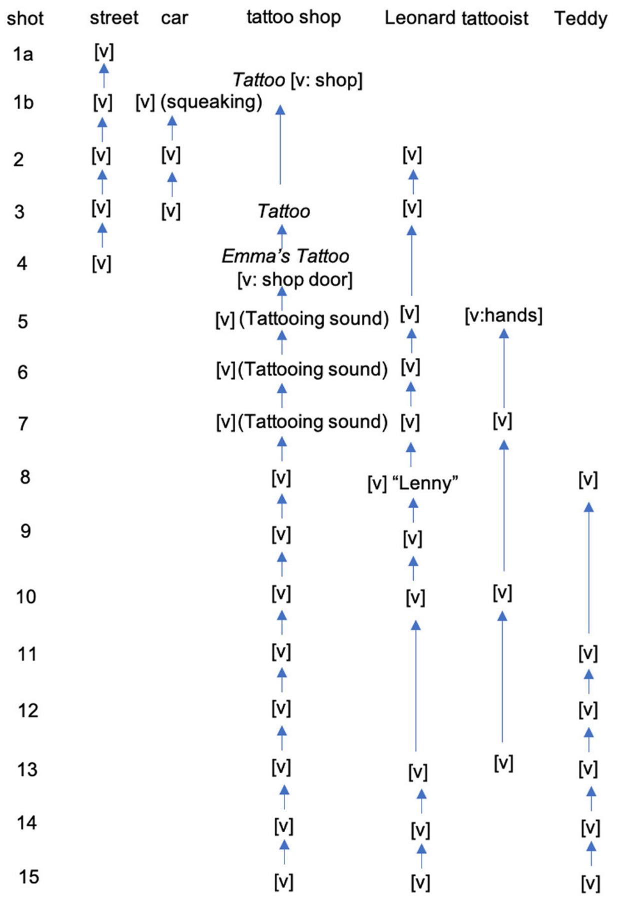


FIGURE 3
Cohesive chains of the tattoo shop scene. [v]= visual realization of people, places, and objects. ()= audio elements, *italics*= written text, ""= spoken text.

After the scene transition into the indoor setting, the cohesive chains allow the construction of sequences of three further events:

> While Leonard in the *tattoo shop* is being tattooed by a *tattoist*, *Teddy* comes in and interacts with him.

We predict that viewers are likely to see such generalized events across the scene transition based on the audiovisual material they engage with. Therefore, the deployment of the material possibilities of film itself serves a central role in guiding a film's reception. Hence, we predict that the *interaction* of cohesive chains (Hasan, 1984), namely, how elements in cohesive chains are combined/co-occur in each shot, should lead to the interpretation path of the generalized events across scene transition.

We explain how we experimentally investigated this in the following questions.

3 Toward experimental investigation

We have predicted that multimodal cohesion analysis can reveal how the filmic elements presented and maintained in a film sequence lead the viewers to interpret particular 'events' across scene transitions on the basis of the available cohesive cues.

In the remainder of this paper, we investigate the empirical support for such a close association of cohesive patterning and narrative interpretation. In this pursuit, we employ the methodology of selecting film sequences and systematically modifying those sequences so that different patterns of cohesion are established. The two sequences, original and manipulated, are then shown to different groups of viewers. We then measure and compare the comprehension and engagement by the two groups of participants.

The measurement was conducted by providing participants with questionnaires designed to evaluate their understanding of the observed events. Three studies testing the functions of cohesive cues in events across scene transitions were performed. The first uses the sequence of the tattoo shop scene in *Memento* analyzed above. The second study employs the same method but factors in the aspect of spatio-temporal cues to test the viewers scene transition in the beginning sequence of *Memento*.

While the results of the second study revealed little effect for the spatial-temporal order, we speculate that this is because it is a puzzle film which begins with loosely connected scenes. This kind of challenging patterns of scene transitions in the film beginning is typical of puzzle film genres (Bateman and Tseng, 2013). his motivated us to conduct a third study using a different film with a structure distinct from the complex, puzzle structure of *Memento*. We chose Ephron's (2009) *Julie* & *Julia*. The movie has simple, linear structure. It portrays the lives of two women, Julie Powell and Julia Child. Julie finds herself in a career rut and decides to challenge herself by embarking on a journey to cook all the recipes from Julia Child's cookbook. She documents her experiences in a blog and discovers a new passion for cooking. The segment we selected for the experiment is a scene in which Julie goes to a butcher's shop to buy ingredients after her failed attempt to cook a dish from Julia's book. In the third study, in addition to selecting a film with a different structure, we also expanded the scope of the measures. We used more fine-grained scales for comparing degrees of correctness of participants' responses and included the measures of participants' confidence level for their responses. We also tested viewers' interpretation of the main character's intention and the length of event time.

In other words, the presentation of the three studies demonstrates the process of our step by step investigation into the complex configuration of cohesive and structural factors in the viewers' narrative comprehension of events and scene comprehension.

3.1 Study 1: "tattoo shop" scene in *Memento*

For the manipulation of the sequence analyzed above, we focused on the scene transition from the street view to the specific tattoo shop. As suggested in Figure 3, the original sequence encompasses sufficient multimodal cohesive cues for explicitly identifying just what kind of shop Leonard is in after the scene transition. For testing the functions of these

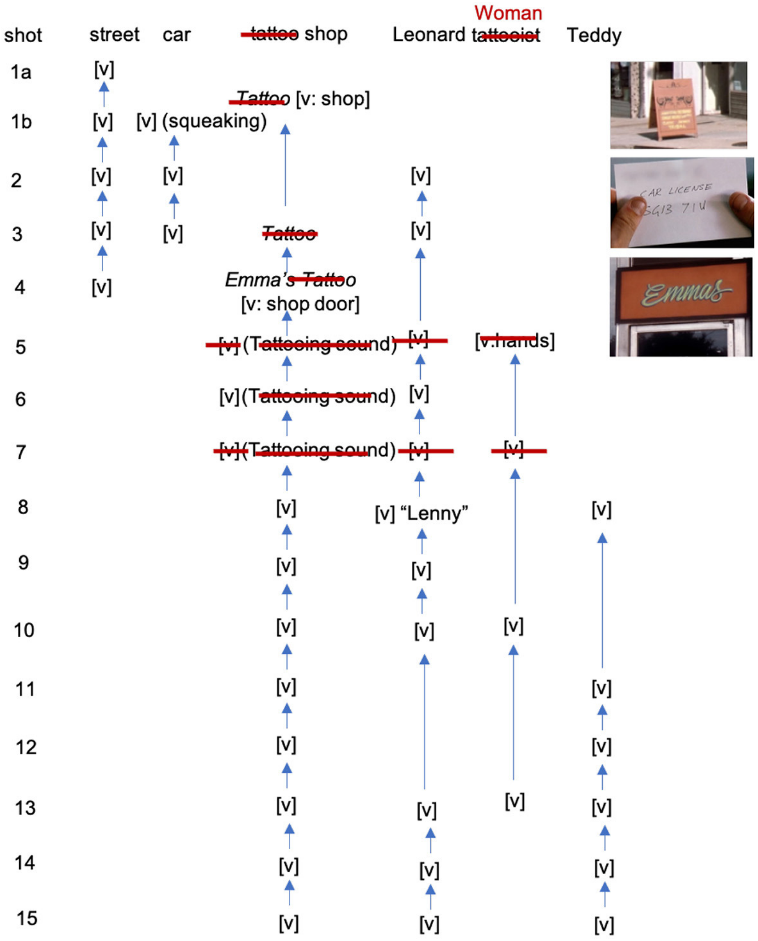


FIGURE 4
Changes in the chain patterns of the manipulated version of the tattoo shop sequence.

multimodal cohesive cues, we subtly removed the specific cues that indicate the identity of *tattoo* shop. That is, we blurred the written signs (on the orange board in shot 1, on the paper in shot 3 and on shop door in shot 4) identifying it as a tattoo shop. We replaced the tattooing sound with some generic background music and we cut out shot 5 and shot 7, the closeups of the tattooing actions. As the two shots removed for the experimental version are close-ups of the tattooing actions, the manipulation does not lead to the loss of other significant story information.

Except for the texts, sounds and actions about the tattoo, the manipulated sequence is identical to the original sequence. Our hypothesis is that removing the cohesive connections in this way should nevertheless disrupt the viewers' narrative comprehension of scene transition from street view to the specific indoor location.

Figure 4 shows the audiovisual cohesive chain analysis of the modified sequence. The removal of the tattoo text in shots 1, 3 and 4 results in the change of the setting from a specific named tattoo shop to a generic indoor space. It could still be recognized as a shop due to some visual elements such as the orange board in shot 1, *Emmas* in shot 4, which are usually recognized as signs for a shop. In terms of multimodal cohesion and the classification system of Figure 1, therefore, the modification undertaken at the discourse level was a change in presentation strategy for the shop scene from a [specific] to [generic] shop. The manipulation also resulted in the change of the tattooist chain - as we cut away shot 5 and shot 7 when the woman is seen specifically as a tattooist, the woman shown in shots 10 and 13 then changes to a [generic] woman in the indoor space.

3.1.1 Hypothesis

To assess whether the manipulations indeed disrupted participants' comprehension of the excerpt, we tested the specific hypothesis:

- Viewers of the manipulated versions will be less certain about the specific identities of the shop, even though the relevant visual elements inside the shop (i.e. tattoo pictures on the wall) are still readily accessible on screen.

TABLE 1

	Cued	Uncued	Total
Comprehended	23	13	36
Not comprehended	0	9	9
Total	23	22	45

Memento tattoo shop study: number of participants with correct or incorrect answer to the question in group 1 (cued version) and group 2 (uncued version).

3.1.2 Experiments

This hypothesis was investigated by having participants answer the following questions immediately following their viewing of the *Tattoo shop* segments:

- "Where is the setting of the indoor place"?

The comprehension test was conducted at the University of Bremen, and participants ($n = 45$) were undergraduate students who had not seen the film before the experiment. The participants were divided into two groups. Group 1 ($n = 23$) watched the original versions (i.e., the cohesively "cued" versions) of the two sequences, while Group 2 ($n = 22$, "uncued") viewed the manipulated versions with cohesive cues removed. Fisher's exact test was used to evaluate statistical significance of dependencies between the cohesion status (cued *vs.* uncued) and viewers' interpretations of the location (correct *vs.* incorrect), with $p < 0.05$ considered significant.

3.1.3 Results

Table 1 presents the test results. All 23 participants in Group 1 who watched the cued version were aware of the specific identity of the tattoo shop, while only 13 participants from Group 2, who watched the uncued version (without cohesive cues), answered the question correctly. The 6 participants who were not certain about the location gave answers varying from a generic room to an office. Fisher's exact test shows a significant association between the independent variable "cohesion" (cued/uncued) and the dependent variable "establishment of the setting's identity" (correct/incorrect) ($p = 0.0006$). Thus, although it is certainly the case that viewers of the uncued version might be able to guess the kind of shop involved correctly based on the pictures of tattoo patterns in the background (in shots 12–15) , the question interrogated here is whether the manipulation makes a difference. The results demonstrate that the cued and uncued versions indeed differ significantly in comprehension.

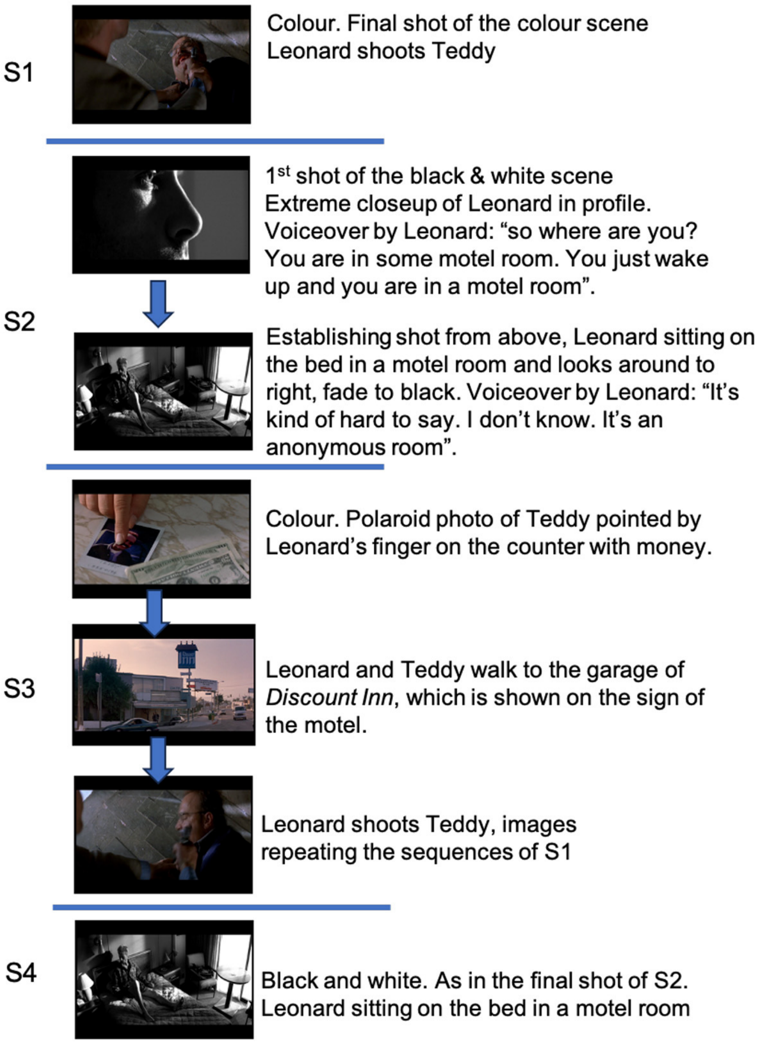
While the previous empirical study of multimodal cohesion by Tseng et al. (2021) focuses on setting interpretation within one continuous scene, our result above re-endorses the empirical ground of cohesive setting across a scene change. The test design was further expanded in the second study to include the factor of spatio-temporal cues between scenes.

3.2 Study 2: the beginning four scenes of *Memento*

In order to show how the deployment of cohesive and spatio-temporal relation can interact and guide narrative interpretation, in the second study, we applied the same method of multimodal cohesive analysis to the beginning sequence of *Memento.* It is the first seven minutes of the film composed of a two-track alternating sequence of four scenes. The detailed cohesion analysis of the four scenes are provided by Tseng (2013). Here we focused on the comprehension tests of the transition of the four scenes, which we simply label S1, S2, S3 and S4, respectively, in order to emphasize their location and inter-relations in the film. Figure 5 shows the transition of the four scenes and the shots before and after the transition points. The changes of these four scenes are very clear for viewers in that their boundaries are signaled through fade-outs and fade-ins, which give the viewer explicit cues for recognizing that a new narrative segment may be beginning.

The first scene, S1, is presented in color and runs behind the opening credits. It depicts events in which Leonard shoots Teddy dead. This scene runs in reverse: i.e. the film is actually played backwards (although the sound runs forward to avoid overly disturbing interpretative possibilities). The second scene, S2, is a black and white scene depicting Leonard sitting in a motel room looking and feeling confused. His confusion is depicted through his voiceover narration. The third scene, S3, then returns to a color scene. It starts with Leonard pointing at Teddy's picture to the receptionist at the motel counter, before Teddy shows up at the reception and walks to the motel garage with Leonard. The middle image of S3 in Figure 5 shows the long shot which depicts their walking from reception to garage. The long shot clearly shows the motel name *Discount Inn* on a big sign seen on the upper part of the screen. Leonard drives Teddy to an abandoned building where Teddy is then shot dead by Leonard. The narrative in this scene therefore directly precedes and overlaps with that of the first color sequence (S1). Finally, the second black-and-white scene (S4) continues Leonard voice-over narration from the previous black-and-white scene in the same motel room.

Drawing on the detailed cohesion analysis by (Tseng and Bateman, 2012) and (Tseng, 2013), the beginning four scenes of Memento are non-linear and have no clear spatio-temporal or logical relations across the color and the black and white scenes. Nevertheless, there are sufficient cohesive cues to interpret the characters and settings across the four scenes. Figure 6A summarizes the straightforward pattern of cohesive chains of the main characters and settings across the four scenes. As the chains show, Leonard and Teddy are both presented visually in S1. The *Leonard* chain shows that the reappearance of Leonard's face is tracked in S2 and S3, while his name as "Lenny" was explicitly identified by Teddy in S3. The *Teddy* chain shows that Teddy is visually presented in S1, and his name is also explicitly written on his photo seen in the beginning shot in S3, before he appears in the motel counter. The chain of the first setting, the building, connects the visual repetition of the same setting in S1 and S3. This re-occurrence is further endorsed through the repetition of the same actions in the images where Leonard shoots Teddy. The chain of the second setting, the motel room, shows that the setting is



S1

Colour. Final shot of the colour scene Leonard shoots Teddy

S2

1st shot of the black & white scene Extreme closeup of Leonard in profile. Voiceover by Leonard: "so where are you? You are in some motel room. You just wake up and you are in a motel room".

Establishing shot from above, Leonard sitting on the bed in a motel room and looks around to right, fade to black. Voiceover by Leonard: "It's kind of hard to say. I don't know. It's an anonymous room".

S3

Colour. Polaroid photo of Teddy pointed by Leonard's finger on the counter with money.

Leonard and Teddy walk to the garage of *Discount Inn*, which is shown on the sign of the motel.

Leonard shoots Teddy, images repeating the sequences of S1

S4

Black and white. As in the final shot of S2. Leonard sitting on the bed in a motel room

FIGURE 5
Shots before and after scene transitions S1 to S4 in the opening sequence of *Memento*.

first explicitly identified as "motel room" in S2 by Leonard's spoken text "So you are in a motel room". In S3, the sign of *Discount Inn* and the visual image of motel reception then cohesively link back to the motel room in S2. There is no clear cue whether the room in S2 is in *Discount Inn* in S3 but the cohesive cue is nevertheless established through a *hyponymous* relation (i.e. motel room, motel reception and motel garage). The hyponymy could possibly lead the viewer to interpret the same setting. The same motel room explicitly reappears in S4.

Along the same lines of the previous study, we created the second version for comparison and manipulated the original sequence by removing the cohesive cues that direct the viewers to the specific interpretation of the second setting, the motel room. To this purpose, we wiped out the two lines of the spoken text by Leonard in S2, which explicitly refer the room as motel room: "You are in some motel room. You just wake up and you are in a motel room". That means, the viewers only hear Leonard saying "so where are you?" In S3, we also manipulated the motel sign, wiping out *Inn* from *Discount Inn*.

Figure 6B shows the cohesive analysis across the four scenes in the manipulated version. The main difference lies in the chain *motel room*. The original *motel room* chain now connects the generic room in S1 and S4. The scene setting in S3 is disconnected from the original motel chain to form an independent generic *Discount* shop counter chain, as any cue indicating the link between a shop/counter and a room is missing here.

Apart from manipulating cohesive cues for the two comparative sequences, namely, cued and uncued versions, we also manipulated the spatio-temporal order of the four scenes. As described above, the chronological sequence of the film story actually runs as S2-S4-S3-S1. In this chronological sequence, Leonard is in a motel room contemplating and plotting the murder of the assumed killer of his wife. He then emerges to the reception, encounters Teddy, and subsequently murders Teddy after driving from the motel to the building. Hence, we also prepared two versions of the scene orders, an original film version and the re-edited version with the S2-S4-S3-S1 order. Our hypothesis was that the re-edited sequence with chronological order might untangle the narrative complexity and

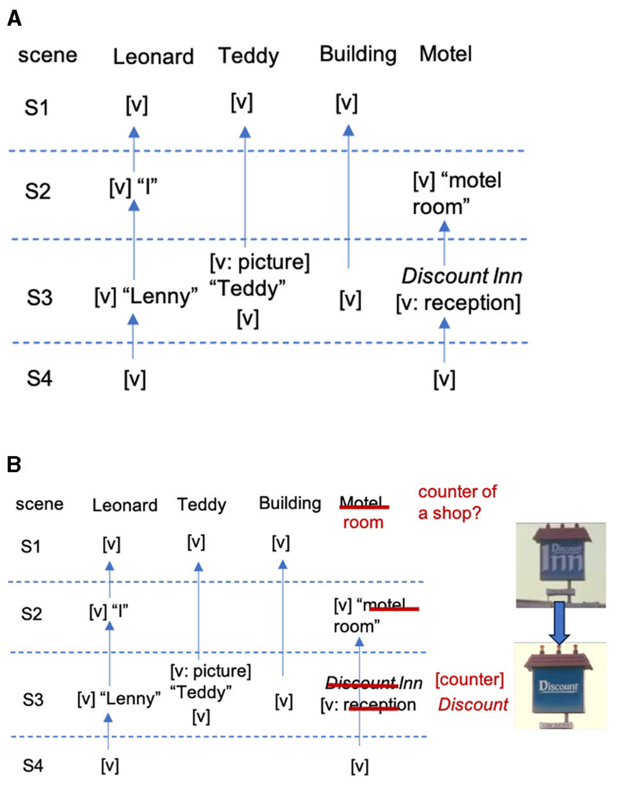


FIGURE 6
The cohesive chains of the original and manipulated versions of the four scenes of the opening sequence of *Memento*. **(A)** The cohesive chains of the two characters, *Leonard*, *Teddy* and the locations *building* and *motel*, are established from S1 to S4. **(B)** The cohesive chains of the manipulated version. The two images on the right show the removal of *Inn* from *Discount Inn*.

lead the viewers to interpret the correct event development, namely, Leonard was first in motel room in black and white scene, which precedes the color scenes. We also predicted that chronological order and cohesive cues impact each other in directing the viewers' scene comprehension. Cohesive cues may help the viewers to interpret scene order and vice versa, temporal cues of the scenes may improve the viewers' identification of the motel room setting, because the color scenes start with the location of reception counter and garage of a motel, which might increase the viewers' inferences of black and white scene as a motel room.

Hence, this study follows a 2x2 design, with cohesive cues and chronological scene order as two independent variables. The sequences, original and manipulated, were then presented to four different groups of participants and differences in their comprehension were measured.

3.2.1 Hypotheses

To assess the predicted effects, we test the specific hypotheses:

- Viewers of the manipulated, *uncued*, versions will be less certain about the specific identities of the motel room, even though the relevant visual elements inside the room are still readily accessible on screen.
- Viewers of the original, *achronological* version will be less certain about the event order of the black and white and the color scenes.

3.2.2 Experiments

The hypotheses were investigated by having participants evaluate one question and one statement immediately following their viewing of the beginning sequence of *Memento*:

1. Where is the setting of the black and white scene?
2. From the perspective of the characters, the black and white scenes happens before the color scenes.

As the previous study, the first one is an open question, while the second question was designed as a Likert scale. The viewer needs to select a response from the 5 points: 1 (totally disagree) to 5 (totally agree).

The comprehension test was conducted at the University of Bremen, and the participants ($n = 74$) were undergraduate students who had not seen the film before the experiment. As the study had a 2x2 design, the participants were divided into four groups: Group 1 ($n = 21$) watched the original achronological version without cohesive cues (with motel cues removed), group 2 ($n = 17$) viewed the chronological version (edited s2-s4-s3-s1 sequence) with cohesive cues removed, group 3 ($n = 17$) viewed the original achronological version with the cohesive motel cues, group 4 ($n = 17$) watched the chronological version with cohesive cues.

3.2.3 Results

Question 1—Comprehension of the motel setting.

For analysing the open answers of the first question about the motel setting (*Where is the setting of the black and white scene?*), we coded the accurate answer (motel/hotel room) as 1 and all other answers as 0. Most inaccurate answers included "a room" or "sleeping room". In this study, we used dichotomous coding and treated any answer without mentioning "motel room" as incorrect. This indeed revealed a clear impact of cohesive cues. Nevertheless, as we will see in Study 3, we decided to refine the coding of the answer about the setting to finer gradations, which then uncovered more nuanced differences of participants' interpretation.

For the statistical analysis, we used logistic regression, suitable for modeling binary responses, to analyse the relationship between cohesive cues, temporal order and the viewers' comprehension. In general, the results show a significant effect of cohesive cues on the viewers' ability to establish the identity of the motel room ($p = 0.0199$). The other independent variable, temporal cue, did not have a significant effect on the viewers' scene comprehension ($p = 0.92763$).

More importantly, logistic regression analysis shows the relationship between cohesive cues and chronological order on the probability of correct comprehension of the motel room. This is demonstrated in terms of odds ratio—it was found that, holding chronological order constant, the odds of accurate comprehension *decrease* by 87% for the viewers who watch the sequence without cohesive cues, compared to the viewers who watch the sequence with cohesive cues. It was also found that, holding cohesive cues constant, the odds of correct comprehension increase only by 5.3% for the viewers who watch the sequence with chronological temporal order, compared to the viewers who watch the sequence with the original, complex achronological order.

The above comparative result of odds ratio is visualized in the Figure 7. Here we can see the impacts of the two factors ("with" vs "without" cohesive cues, "chronological" vs "achronological" sequencing) to the correctness of participants' answers.

A significant decrease of correct comprehension of motel room (between the probabilities of 1 and 0) if the cohesive cues are removed (namely, when data points move from "with" to "without" variable). In terms of "chronological and achronological" variable, there is no significant difference in the probabilities of correct comprehension. The two lines are nearly merged.

Question 2—Temporal relation between black-white and color scenes

For analysing the Likert scale results of question 2 (*From the perspective of the characters, the black and white scenes happens before the color scenes*), namely, about the temporal order of color and black and white scenes, we used the Align-and-Rank transform (ART) test. The results show a significance of cohesive cues ($p = 0.0156$) in the viewers' inferences of event orders across the scene transitions. The violin plot in Figure 8 shows the main difference of the two conditions (with and without cohesive cues). The distributions of the Likert scale score (1–5) for the two conditions are demonstrated through density curves - here we can see that in the original version with cohesive cues, a significant portion of participants is related to the score of 5 (totally agree), while the responses of the participants who watch the version without cohesive cues substantially vary, with more responses toward 1 (totally disagree).

However, the variable of temporal order did not have a significant effect on the viewers' scene connections ($p = 0.1379$). Moreover, no significant interaction effect between cohesive cues and temporal order was revealed ($p = 0.7566$).

In summary, in this study, cohesive cues remain significant in leading the viewers' interpretation of both the specific setting of the scenes (Hypothesis 1) and temporal order of event sequences (Hypothesis 2).

However, the second factor that we considered, the variable of chronological order of the scenes, does not have significant effects both on the comprehension of motel setting and on the interpretation of event sequence orders.

The reason for the weak effect of the second factor, the chronological order of scenes, might be attributed to the fact that *Memento* is a puzzle film characterized by Nolan's signature complex, non-linear film structure. Although in the manipulated version, we tried re-ordering the four scenes to match its general story order (S2-S4-S3-S1), the scene transition between S4 and S3, namely, the black and white scene of Leonard in a motel room and the next color scene of Leonard at the motel reception, still exhibit a substantial ellipsis. This deliberate narrative gap in Nolan's famous puzzle film might be the reason why the effect of the variable chronological order is diluted.

To refine our test design in order to further investigate the significance of chronological orders of scene relations theorized by (Bateman, 2007) and (Bateman and Schmidt, 2012), we conducted a third study, using a sequence of a more straightforward drama film, *Julie* & *Julia*. With this film material, in the next study, we were able to refine our experimental measures and broaden our questions to include the viewers' confidence level of their comprehension and their interpretation of the main characters' intention. The rationale behind testing participants' level of confidence in their own inferences was to test whether the manipulation has resulted in any uncertainties among participants as to their own judgements regarding the setting and the goal of the main character. Testing participants' self-rated level of confidence could provide insight into whether there was a discernible difference in the perceived confidence influenced by the manipulation.

3.3 Study 3: Julie & Julia

As described above, in the third study, we tested the same independent variables, cohesive cues and temporal orders, but we refined our measures and used a sequence extracted from a non-puzzle film. The sequence also deals with three transitions across four scenes.

Figure 9 presents four representative shots from each scene. Shot 1 depicts the first scene in the living room of the main character, Julie Powell, while she is seen typing on her laptop and reading aloud to her husband a passage from her blog, wherein she recounts her unsuccessful cooking attempt from the previous day. Shot 2 shows the second scene, in which she walks on the street before entering a butcher's shop. In shot 3, Julie has entered the shop. Inside the shop, Julie is seen purchasing ingredients for one of the recipes she is attempting from Child's cookbook. In this scene, the shop setting is filled with conventional visual cues of a butcher's shop, e.g. meat displayed behind the counter. The viewers can also hear Julie off-screen voiceover depicting her cooking plan throughout shot 2 and 3. This scene is then cut to the kitchen setting, shown in shot 4, where Julie is already back from the butcher's shop and is cooking using the ingredients she just purchased from the shop.

While creating film materials for all experimental conditions, we decided to first remove the overall spoken text of the entire sequence (including the dialogues between Julie and her husband and Julie's voiceover). The reason is that the spoken text is highly indicative for Julie's plan to go to the butcher's shop and purchase meat in the shop. We wiped out the entire spoken text to remove verbal cohesive cues leading to the setting of the butcher's shop. Nevertheless, removing the entire verbal text for the version without cohesive cues could lead to the substantially loose control of two conditions because we also wiped out other verbal information that is relevant to the overall narrative interpretation. Hence, to secure clean effects through the experimental control, we decided to remove and replace the spoken text with the film's soundtrack music for all conditions first and then manipulate visual cohesive cues and temporal orders based on these sequences already without verbal text.

Figure 10A illustrates the analysis of cohesive chains of the version with visual cohesive cues. As we can see in the chain pattern here, no spoken verbal cues are included, as they were all removed.

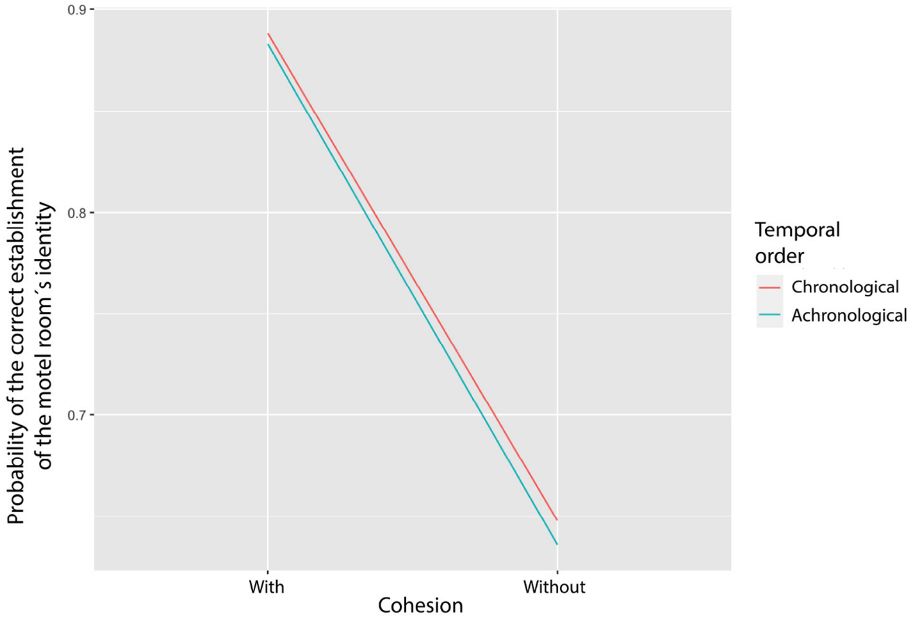


FIGURE 7
The probability of the correct establishment of the motel room identity across the groups with/without cohesion and chronological/achronological order of the scenes.

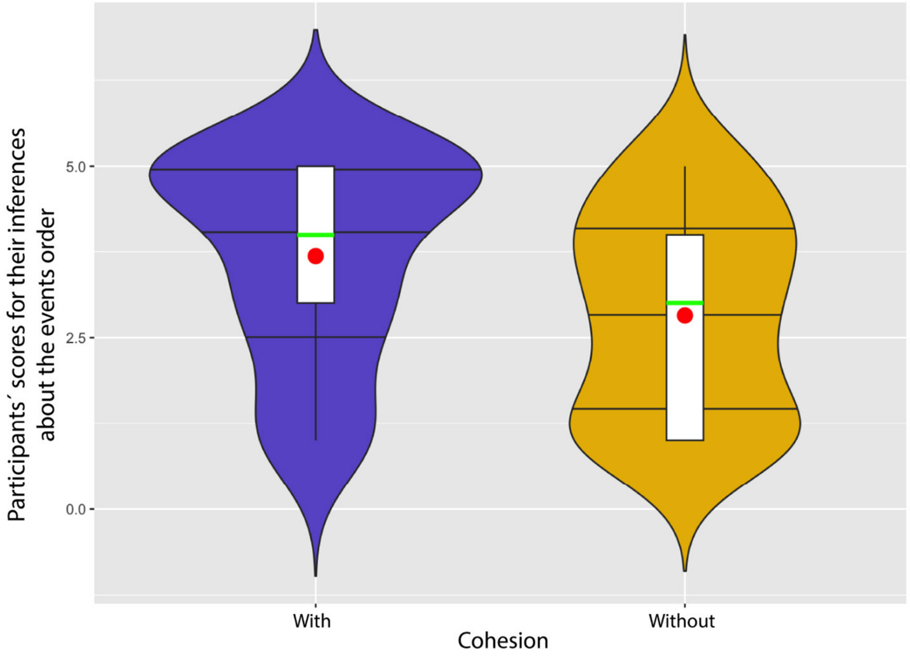


FIGURE 8
Main effect of cohesive cues (with and without) on participants' inferences of the events ordering.

Across the four scenes, three reoccurring narrative elements are tracked. The first *Julie* chain shows her reoccurring appearance throughout the entire segment. The second chain is the setting of Julie's *home*, it is first presented with a living room setting and is cohesively linked through a hyponymous relation by another home setting, the kitchen. The third chain is a butcher's shop. It is presented in shot 2, the specific identity of the *butcher's shop* chain is introduced through textual visual cues such as *K&T Quality Meats*, *Meat and Poultry* and the associated price tags on the window. In shot 3, visual cues such as butcher's outfit, meat and cheese products inside the shop cohesively link the setting to the *butcher's shop* chain introduced in the previous shot.

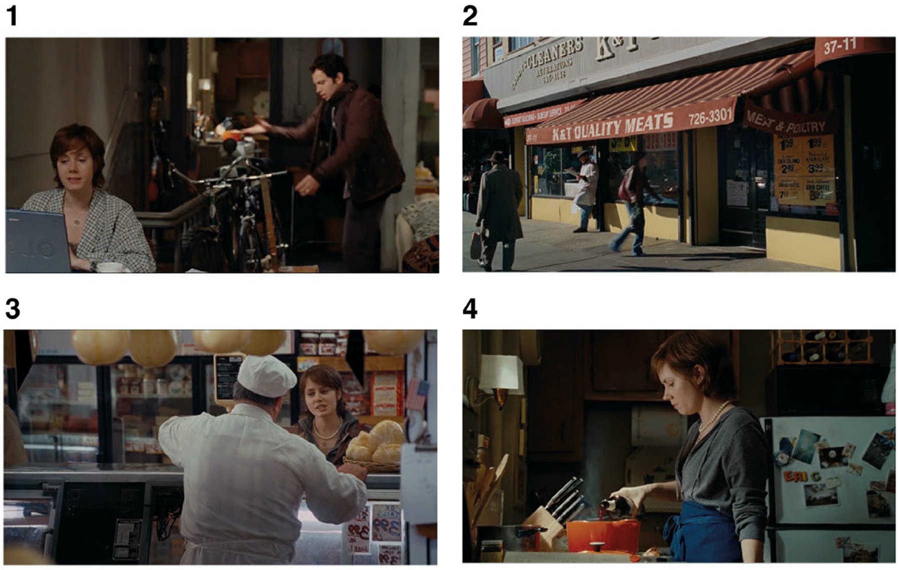


FIGURE 9
Selected shots in *Julie & Julia*.

The events across the scene transitions can then be depicted based on the chain pattern as follows:

> *Julie* is first at *home* and then she goes to the *butcher's shop*, before returning *home* again.

To manipulate visual cohesive cues, we targeted the setting of the butcher's shop. We removed all visual cohesive cues from the butcher's setting (shop front door in shot 2 and indoor setting in shot 3) that reveals the specific identity of the shop. Figure 11 shows exactly what visual cues were removed from the original scenes. Here we can see that the text *K&T Quality Meats*, *Meat and Poultry* and the price tags for the meat products are written on the roof and the windows of the shop that Julie is entering in shot 2. In the uncued version, these texts have been removed and we can see that Julie is entering an indoor space with no written indication of its identity (Figure 11A). Moreover, in the version with cohesive visual cues we can see Julie talking to the butcher who is attired in a traditional white butcher's costume (Figure 11B), placing and weighing meat pieces on the scale. There are also refrigerators stocked with jars, meat and cheese, big chunks of cheese hanging off the ceiling and price tags on the counter's glass. Contrasting this, in the uncued version (bottom image of Figure 11B) we turned the butcher's conventional outfit into a blue shirt and a red hat. We have also removed all food products, price tags, the scale and the meat pieces in the butcher's hand, so that it is no longer identifiable what he is putting on the counter.

Figure 10B illustrates the cohesive analysis across the four scenes in the version without cohesive cues. Similar to the previous two *Memento* studies, the removal of cohesive cues transforms the specific shop identity (here the butcher's shop) into a generic, non-specified indoor space. Hence, the same setting chain now includes only visual elements of a generic indoor space, such as a door and a counter.

In addition to the removal of the cohesive visual cues and for testing the second factor, namely, the temporal scene order, we re-edited the temporal relation between the four scenes. That means, based on the original temporal order (S1-S2-S3-S4), we edited the order into an alternative one, S4-S1-S2-S3. In this alternative version, viewers first see Julie cooking in the kitchen, followed by the living room scene and subsequently, Julie's visit to the butcher's shop. This was done in order to temporally change the logical relation between the scene in the butcher's shop and the scenes at home. Our hypothesis following the scene order modification is that, the home cooking scene (S4) directly followed by Julie's visit to a shop (S3) might lead the viewers to infer the setting as a food-related shop, even in the absence of the cohesive visual cues of butcher's shop.

3.4 Hypotheses

To assess the predicted effects of cohesive cues and temporal orders, we tested the specific hypotheses:

- Viewers of the uncued (without cohesive visual cues) and alternative temporal order version will be less accurate at establishing the identity of the butcher's shop.
- Viewers of the uncued (without cohesive visual cues) and alternative temporal order will be less accurate at identifying the goal and intention of Julie's actions in the story.
- Viewers of the uncued (without cohesive visual cues) and alternative temporal order version will be less confident in their inferences about the butcher's shop identity and the goal of the female character's actions.

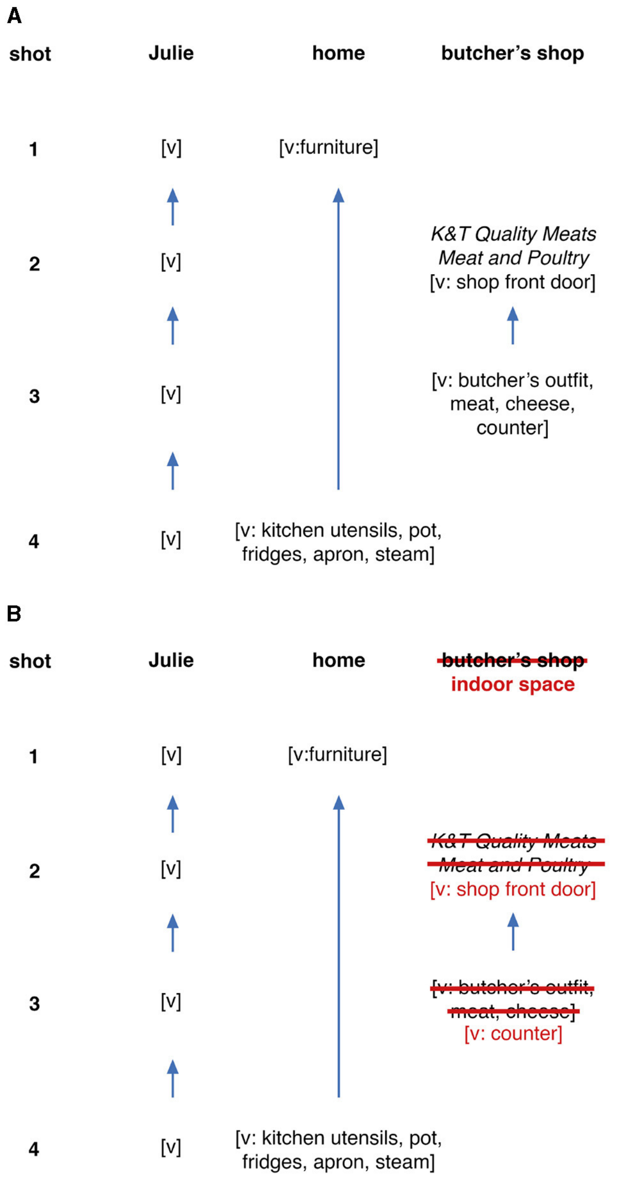


FIGURE 10
The cohesive chains of the original and manipulated versions of the selected sequence of *Julie & Julia*. **(A)** The cohesive chains of the character Julie and the locations home and butcher's shop are established from S1 to S4. **(B)** The cohesive chains of the manipulated version.

- There will be a difference in participants' time perception of the story events between the original and a orders of the segment.

3.5 Experiments

The experiment was conducted at the University of Bremen. All participants ($n = 76$) were students or employees of the university who had not seen the movie before. Each participant was allocated to one of the four experimental groups: Group 1 ($n = 19$) watched the original temporal order version with cohesive cues, Group 2 ($n = 19$) watched the original temporal order version without cohesive cues, Group 3 ($n = 19$) watched the alternative temporal order version with cohesive cues, Group 4 ($n = 19$) watched alternative temporal order version without cohesive cues.

For a general comprehension check, that is, to make sure that participants were also able to follow the non-manipulated part of the excerpt, we also asked them some basic comprehension questions to reveal whether they noticed that there were 3 characters in total, that Julie was typing and talking in S1 and that the husband was holding a bike. Hence, we first asked the following questions:

- How many characters are in the video clip (both main and secondary)?
- What was the female character doing when she was talking to the man in the living room?
- What was the man holding when he was leaving the living room?

The questions of comprehension check were immediately followed by the questions listed below to address the hypotheses of cohesive cues and temporal order:

1. In what kind of place was the female character when she was talking to the other man?
2. Why do you think she went there? Be specific.
3. How sure are you about your answer to the previous question?
4. Estimate approximate time period shown in the video clip.

To avoid leading language, we used the wording "the other man" when referring to the butcher. This decision was based on the arrangement of questions in the questionnaire provided to participants. The question directly followed two questions that mentioned "the man in the living room".

For analysing the answers, participants' responses to question 1 and question 2 were converted into points that ranged from 0 to 3, where 0 indicated the least accurate answer, and 3 represented the right establishment of the butcher's shop or the goal of the female character going there. Participants who explicitly stated that the female character went to the designated place to buy ingredients received a score of 3. Examples of such responses include *"She was buying ingredients for cooking"* and *"She probably wanted to buy some ingredients for dinner"*. Those who inferred that she was buying food were awarded a score of 2, as can be seen in the following response *"She went there to get lunch"*. Participants who mentioned shopping, without specifying ingredients or food, received a score of 1, for example, as in the response *"womöglich um sich irgendwas zu kaufen"*(perhaps to buy something). Those who did not mention any of the above received a score of 0, as indicated in the response *"to pick up some parcel"*.

Question 3 is about participants' confidence in their previous responses. It is estimated using a five-point Likert-Scale question, ranging from 1 (not sure) to 5 (very sure).

Question 4 addresses participants' time perception of the story events. Participants were given the following options: two hours, four hours, one day, more than one day.

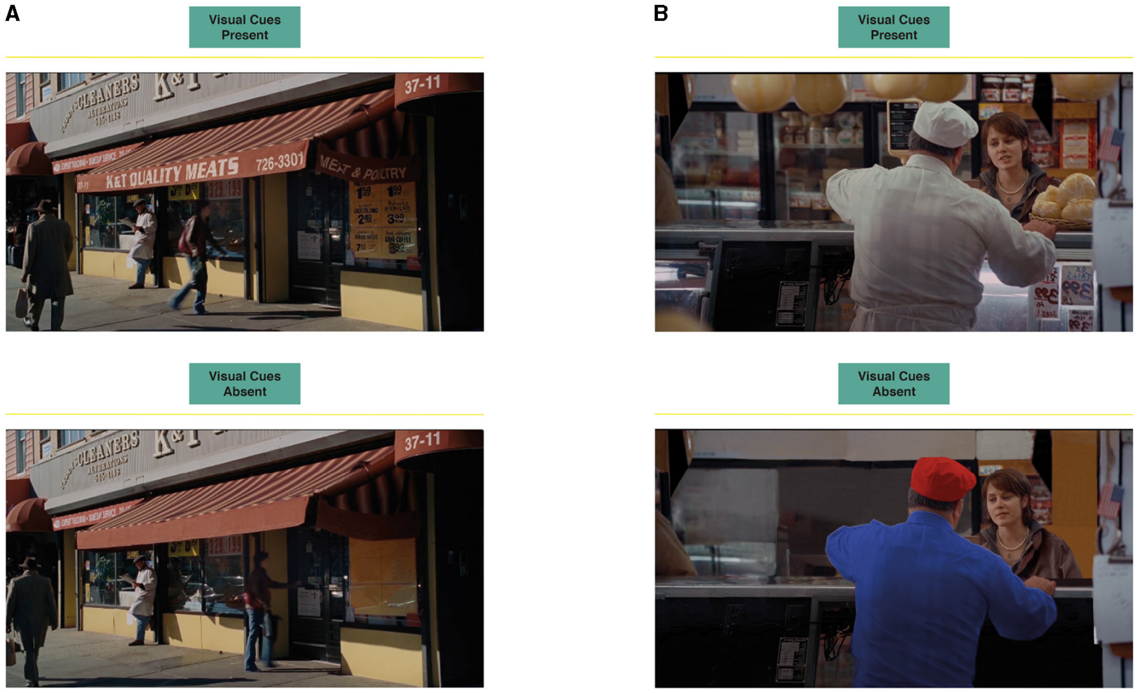


FIGURE 11
Screenshots from the original and manipulated versions of the butcher's shop scene in *Julie & Julia*. **(A)** Scene outside of the butcher shop: The version with cohesive visual cues (top) versus the version without cohesive visual cues (bottom) in *Julie & Julia*. **(B)** Scene inside of the butcher shop: The version with cohesive visual cues (top) versus the version without cohesive visual cues (bottom) in *Julie & Julia*.

3.6 Results

General comprehension check

The general comprehension check shows that participants across the four groups understood the overall, non-manipulated part of the sequence.

Question 1 - Comprehension of the setting identity of the butcher's shop

For analysing responses of this question, we used the ART test. In general, the main effects of both independent variables, cohesive cues and temporal order, were significant.

In terms of the variable of cohesive cue, the analysis results show differences in participants' comprehension of the segment between the version with and without cohesive cues. In the conditions where cohesive cues were present, participants were significantly more accurate at establishing the identity of the butcher's shop. This result is visualized in Figure 12A. Here one can see that participants from the condition with cohesive cues on average received higher scores than participants in the condition without cohesive cues. [Mean (M) = 2.553, Interquartile Range (IQR) = 1] compared to the uncued conditions, where visual cues were absent (M = 1.684, IQR = 0), as revealed by the Align-and-Rank transform (ART) test ($p < 0.05$). The plot shape indicates greater variation in participants' responses in the condition without cohesive cues, while those in the conditions with cohesive cues exhibited higher agreement.

Unlike the previous study on the puzzle film *Memento*, we found that the temporal order in this study played a significant role in comprehending the butcher's shop setting. This result is shown in Figure 12B in which we can tell the difference that participants who watched the manipulated, alternative temporal order version on average received higher scores (M = 2.237, IQR = 1) than those who watched original temporal order version (M = 2, IQR = 0), as shown by the ART test (p = 0.038) (Figure 12B). As described above, we predict that the alternative temporal order version which brings the home cooking scene before the butcher shop scene enhances the inferences of the shop as a food/cooking relevant shop. The plot shape also reveals that in the original temporal order conditions, the majority of participants received the score of 2. In contrast, the alternative temporal order conditions exhibited a more wide spread distribution between scores 2 and 3, with more participants receiving a score of 3.

With regard to the interaction effect between the visual cues and the temporal order, our results show no significance, as the interaction effect between the two factors tested in Study 2.

Question 2 - Main character's intention in the story

To analyse participants' responses of question 2, namely, "*Why do you think she went there?*", the ART test revealed that the removal of cohesive cues indeed led to a deterioration in participants' ability to comprehend the goal of Julie in the segment (p = 0.013). As shown in Figure 13, the average score is significantly higher in the conditions with cohesive cues (M = 2.132, IQR = 0.75) compared to the conditions without cohesive cues (M = 1.632, IQR = 1). The IQR illustrated in the plot indicates that, in the conditions without cohesive cues, the middle 50% of participants received scores below 2, while in cued conditions, the middle 50% received scores above 2.

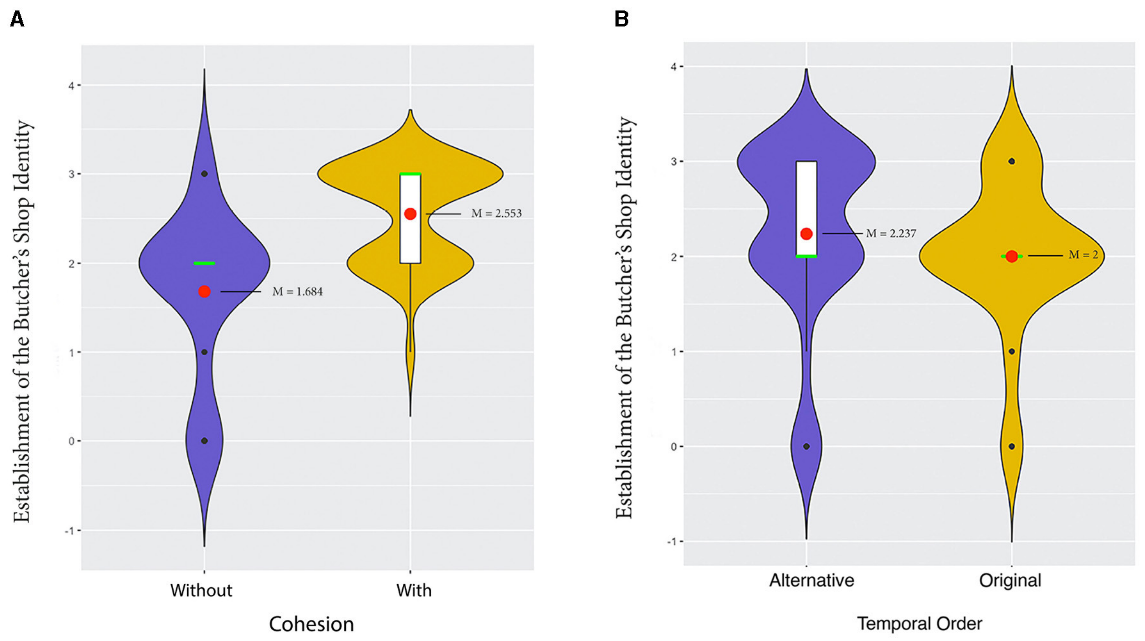


FIGURE 12
Participants' scores for establishing the identity of the butcher's shop based on the presence of cohesive cues (without/with) and temporal order (alternative/original) in *Julie & Julia*. **(A)** Main effect of cohesive cues on participants' ability to establish the butcher's shop identity. **(B)** Main effect of temporal order on participants' ability to establish the butcher's shop identity.

In terms of the second factor, the temporal order, the effect was not significant ($p > 0.05$). Our analysis also reveals that there was no significant interaction effect between cohesive visual cues and temporal order.

Question 3 - Confidence level of viewers in their own responses

At first glance at the results, we note that none of the participants who watched the extract without cohesive cues rated their confidence at level 5 (representing "very sure"). Similarly, none of the participants who viewed the extract with visual cues rated their confidence at level 1 (indicating "very unsure") in both the establishment of the butcher's shop identity and the establishment of the causal relation.

We then used two-way ANOVA to analyse responses to question 3. Our results show that the main effect of cohesive visual cues on confidence level was highly significant ($p < 0.001$). As illustrated in Figure 14, the comparative findings indicate that participants in conditions with cohesive cues displayed significantly higher confidence in their inferences (M = 3.5, SD = 1.033) compared to participants in conditions without cohesive cues (M = 2.474, SD = 0.893).

As in the analysis of the questions 1 and 2, the effect of temporal order on confidence level was not statistically significant ($p > 0.05$) and there was no significant interaction effect between visual cohesive cues and temporal order either.

Question 4 - Viewers' time perception of story events

For analysing the responses of question 4, we used the ART test again to test the effect of cohesive cues and temporal scene order on the time perception of the story events.

A significant main effect of temporal order was observed ($p = 0.024$), indicating that participants estimated the approximate duration of the events taking place in the story to be *longer* in the conditions with original temporal order (M = 2.105, IQR = 0.75) compared to conditions with alternative temporal order (M = 1.737, IQR = 1), as illustrated in Figure 15. The IQR illustrated in the plot indicates that, in the alternative order conditions, the middle 50% of participants rated that the events in the sequence took less than four hours. In contrast, in the original temporal order conditions, the middle 50% of participants estimated the events took more than four hours.

There was no significant interaction effect between the independent variables on participants' temporal perception of the segment ($p > 0.05$). The results show that the main effect of cohesive cues on time perception was not statistically significant ($p > 0.05$).

4 Discussion and conclusion

The results of our three studies have further empirically supported our hypothesis that cohesion in film is highly relevant and significant in people's comprehension of scenes and settings whether during a continuous scene or transition across different scenes and whether in a complex puzzle film or in a narratively straightforward film. We also provide results showing that cohesion is significant in viewers' inferences of character's intention in the story. Moreover, we also show the significance of temporal order of scenes in viewers' inferences of both scenes and settings and the length of event time.

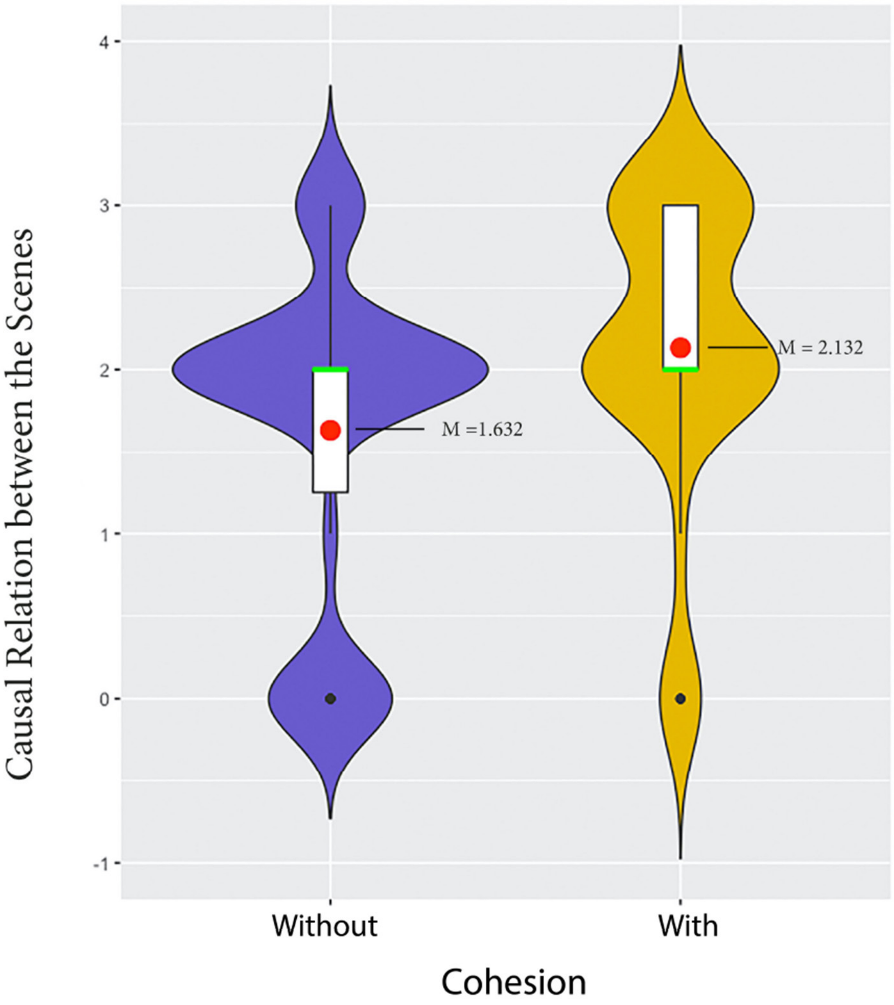


FIGURE 13
Main effect of cohesive cues on participants' ability to establish the goal of the main character in *Julie & Julia*.

The empirical results of our study lead to more questions and hypotheses for future investigation.

First of all, the cultural background of viewers might impact their understanding of a film. For instance, two participants in our Study 3 noted after the experiment that such butcher's shops were more prevalent in the countries of their origin than in Germany. This observation is similar to the previous research on the role of cultural background in narrative comprehension (Horiba, 1990), which demonstrated that, when reading about scenes taking place in Japan, native Japanese readers focus on more intricate details and utilize their cultural knowledge about the local details to infer the protagonists' actions. In contrast, non-native Japanese readers did not exhibit the same degree of event details in their narrative comprehension. Hence, we believe that cultural origin of viewers could be a relevant factor for different ways of establishing cohesion within a scene and could be a crucial variable to investigate empirically.

Another question to dive deeper into is what narrative features impact viewers' interpretation of intentions and goals of characters. The event comprehension model (Zacks, 2007) proposes that the changes of space, time and intention all lead to the comprehension of event change. However, there has not been sufficient research indicating whether these factors actually interact—our study 3 shows that space (cohesive cues of setting) is significant in viewers' comprehension of character's intention, while time is not a significant factor for story intention. Hence, more empirical tests are thus required to untangle the inter-relation of these factors for event comprehension.

Our study 3 also explores the intriguing issue of time perception and its relation to space in film. Our results indicate that the difference in the perception of the temporal length of events in the two experimental groups (original and alternative versions) actually rests on the different narrative spatial structures. For instance, in the non-chronological version, in which S4 in Julie's kitchen is edited directly before S1 in Julie's living room, the story event time seem shortened for the viewers due to the contiguous space relation between S4 (kitchen) and S1 (living room). We hence further hypothesize that event perception of two contiguous spaces

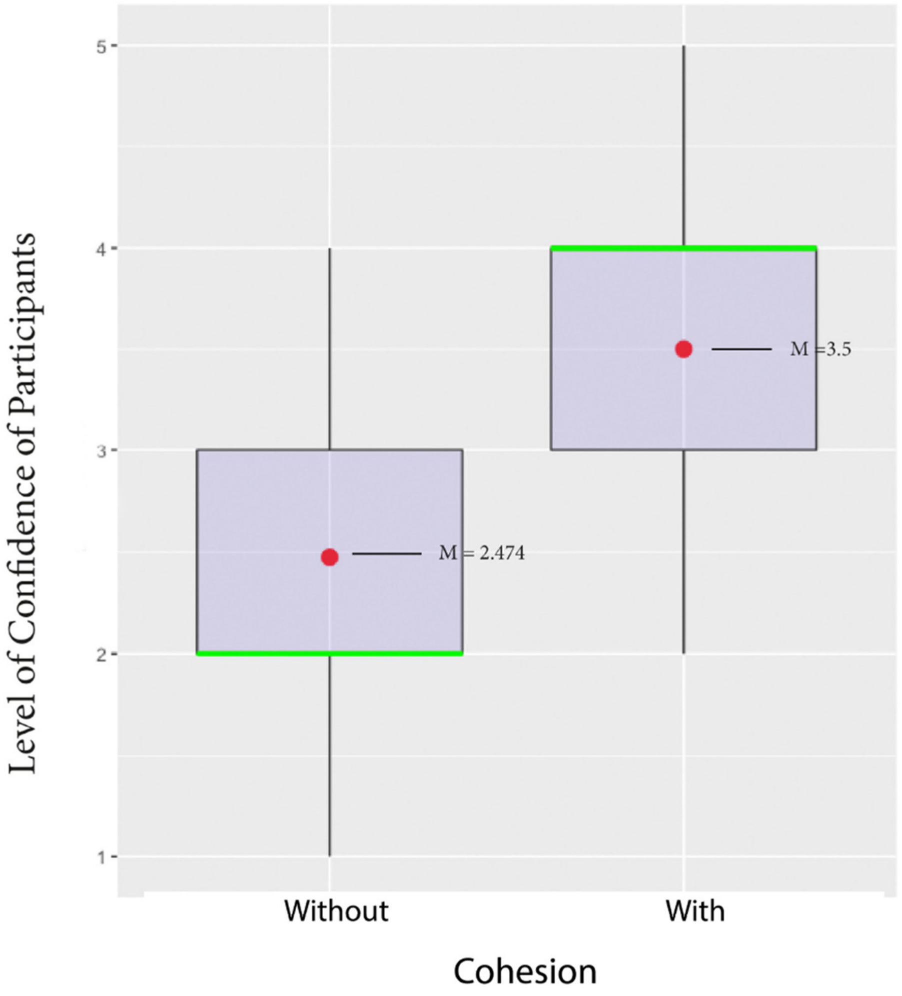


FIGURE 14
Main effect of cohesive cues on participants' level of confidence in their inferences in *Julie & Julia*.

could also lead to the event interpretation of closeness in event time. The hypothesis will require further empirical studies.

In this paper, we have presented results of three empirical studies conducted with the aim of investigating how multimodal cohesion in film influences viewers' narrative comprehension of events across scene transitions. While the previous research (Tseng et al., 2021) has indicated that the absence of cohesive cues leads to an uncertainty about the setting within a continuous scene, we have broadened the test scope about multimodal cohesion in three aspects: (1) we tested how cohesive cues function to carry viewers across scene transitions (study 1), (2) we added another factor, namely, temporal order of scenes theorized by Bateman (2007), to investigate how these two factors impact narrative comprehension independently and interactively, (3) apart from testing viewers' correct understanding of setting identities, we also tested the viewers' confidential level about their answers, whether their understanding of character's intention and event time perception are related to the two factors, cohesive cues and temporal scene orders.

We also identify limitations of conducting experiments using cinematic materials. It is challenging to predict if a film material offers enough control of stimuli. The refinement of our experimental from Study 1 to Study 3 shows our endeavor to shift from *Memento* to *Julie* & *Julia* in order to extend the questions that can be addressed in a more controlled fashion.

We hope our empirical studies on multimodal cohesion demonstrate a valuable combination of empirical methods and multimodal discourse analysis, which is a robust, textual-based model highly valuable for investigating people's cognitive processes through uncovering how people maximize coherence when perceiving multimodal artifacts. Finally, we also hope to have shown how the multimodal film research endeavors in the last decade by Bateman (2007), Bateman and Schmidt (2012), Tseng and Bateman (2012), and Tseng et al. (2021) continue to develop and shed light on significant aspects of human perception and meaning interpretation of film narratives.

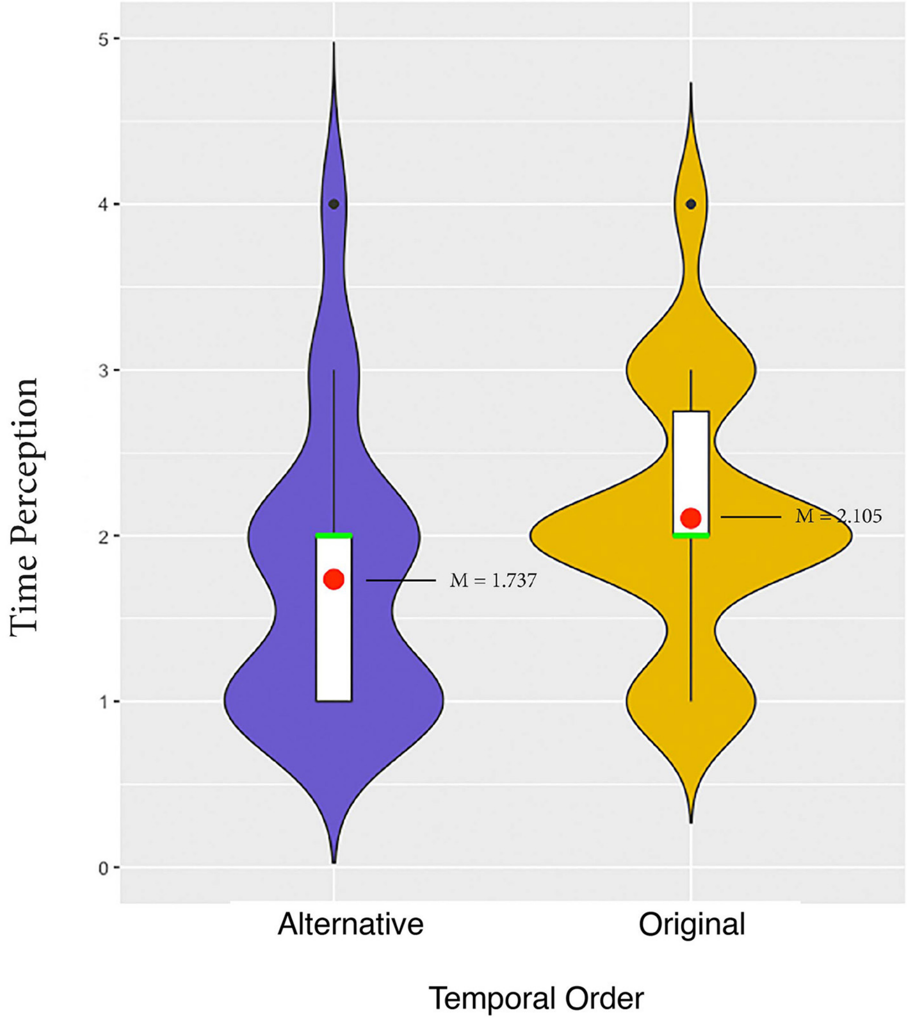


FIGURE 15
Main effect of temporal order on participants' time perception in *Julie & Julia*.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the studies involving humans because in our studies, participants watched movie clips and then completed a questionnaire to measure their understanding. No personal data was collected. Participants were informed about any potential risks and their participation was strictly voluntary, with the option to withdraw at any time. All collected data was kept strictly confidential. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of data included in this article.

Author contributions

DM: Writing – original draft. C-IT: Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The research in this paper was conducted with the support of the University of Bremen.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bateman, J. A. (2007). Towards a grande paradigmatique of film: Christian Metz reloaded. *Semiotica* 167, 13–64

Bateman, J. A., and Schmidt, K.-H. (2012). *Multimodal Film Analysis: How Films Mean. Routledge Studies in Multimodality*. London: Routledge.

Bateman, J. A., and Tseng, C. (2013). The establishment of interpretative expectations in film. Rev. Cognit. Linguist. 11, 353–368. doi: 10.1075/rcl.11.2.09bat

Bordwell, D. (2008). *The Hook: Scene Transitions in Classical Cinema. Online Essay (David Bordwell's Website on Cinema)* Available online at: http://www.davidbordwell.net/essays/hook.php (accessed March 15, 2024).

Drummond, T., and Wildfeuer, J. (2020). "The multimodal annotation of gender differences in contemporary tv series. Annotations in scholarly editions and research," in *Functions, Differentiations, Systematization*, 35–58.

Eisenstein, S. (1969). *Film Form: Essays in Film Theory*. Eugene: Harvest Book.

Ephron, N. (2009). "Julie & Julia," in *Columbia Pictures, USA (film)*.

Halliday, M. A. K., and Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M. A. K., and Matthiessen, C. M. I. M. (2013). *Halliday's Introduction to Functional Grammar*. London and New York: Routledge.

Hasan, R. (1984). "Coherence and cohesive harmony," in *Understanding Reading Comprehensi on: Cognition, Language, and the Structure of Prose*, ed. J. Flood (Newark, Delaware: International Reading Association), 181–219.

Horiba, Y. (1990). Narrative comprehension processes: a study of native and non-native readers of japanese. *Modern Lang J.* 74, 188–202.

Loschky, L. C., Larson, A. M., Magliano, J. P., and Smith, T. J. (2015a). What would jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLoS ONE.* 10, e142474. doi: 10.1371/journal.pone.0142474

Loschky, L. C., Ringer, R. V., Ellis, K., and Hansen, B. C. (2015b). Comparing rapid scene categorization of aerial and terrestrial views: a new perspective on scene gist. *J. Vision* 15, 1–29. doi: 10.1167/15.6.11

Martin, J. R. (1992). *English Text: Systems and Structure*. Amsterdam: Benjamins.

Metz, C. (1974). *Film Language: a Semiotics of the Cinema*. Oxford and Chicago: Oxford University Press and Chicago University Press.

Nolan, C. E. (2000). "Memento," in *Summit Entertainment, USA (film)*.

Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*. 6, 1–27. doi: 10.3167/proj.2012.060102

Tseng, C. (2008). "Cohesive harmony in filmic text," in *Multimodal Semiotics: Functional Analysis in Contexts of Education*, ed. L. Unsworth (London: Continuum), 87–104.

Tseng, C. (2012). Audiovisual texture in scene transition. *Semiotica* 192, 123–160

Tseng, C. (2013). *Cohesion in Film: Tracking Film Elements*. Basingstoke: Palgrave Macmillan.

Tseng, C., and Bateman, J. A. (2012). Multimodal narrative construction in christopher Nolan's Memento: a description of method. *J. Visual Commun.* 11, 91–119. doi: 10.1177/1470357211424691

Tseng, C., and Bateman, J. A. (2018). Cohesion in comics and graphic novels: an empirical comparative approach to transmedia adaptation in city of glass. *Adaptation* 11, 122–143

Tseng, C., Laubrock, J., and Pflaeging, J. (2018). "Character developments in comics and graphic novels: a systematic analytical scheme," in *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods*, eds. J. L. Alexandra Dunst, J. Wildfeuer, and A. Dunst (London: Routledge).

Tseng, C.-I., Laubrock, J., and Bateman, J. A. (2021). The impact of multimodal cohesion on attention and interpretation in film. *Discourse, Context Media* 44, 100544. doi: 10.1016/j.dcm.2021.100544

Tseng, C.-I., and Thiele, L. (2022). Actions and digital empathy in interactive storytelling of serious games: Multimodal discourse approach. *Soc. Semiot.* doi: 10.1080/10350330.2022.2128039

Zacks, J. (2015). *Flicker: Your Brain on Movie*. Oxford: Oxford University Press.

Zacks, J. M., Speer, N., Swallow, K., Braver, T., and Reynolds, J. (2007). Event perception: a mind/brain prespective. *Psychol. Bullet.* 133, 273–293. doi: 10.1037/0033-2909.133.2.273