Re-Presentations of Space in Hollywood Movies:
An Event-Indexing Analysis

James Cutting and Catalina Iricinschi
Cornell University

Popular movies present chunk-like events (scenes and subscenes) that promote episodic, serial updating of viewers’ representations of the ongoing narrative. Event-indexing theory would suggest that the beginnings of new scenes trigger these updates, which in turn require more cognitive processing. Typically, a new movie event is signaled by an establishing shot, one providing more background information and a longer look than the average shot. Our analysis of 24 films reconfirms this. More importantly we show that, when returning to a previously shown location, the re-establishing shot reduces both context and duration while remaining greater than the average shot. In general, location shifts dominate character and time shifts in event segmentation of movies. In addition, over the last 70 years re-establishing shots have become more like the noninitial shots of a scene. Establishing shots have also approached noninitial shot scales, but not their durations. Such results suggest that film form is evolving, perhaps to suit more rapid encoding of narrative events.

1. Presenting and Re-Presenting Space

About 7 min into Home Alone (1990) a pizza-delivery car arrives at the McAllister house. In one shot the car topples a metal driveway ornament, as captured in the still at the top panel of Fig. 1. This shot is 3.75 s in duration. Later, almost 48 min into the movie and after his parents have mistakenly left him and gone to Paris, Kevin McAllister (Macaulay Culkin) orders another pizza and a new scene begins when the car returns. Again, it topples the ornament, shown in a still at the bottom of Fig. 1. Notice that this still shows an enlarged view of the car. Moreover, this shot is only 2.17 s in duration.

About 98 min into The Social Network (2010) the first shot of a new scene reveals the open-plan interior of the new Facebook headquarters, shown in the still at top panel of Fig. 2. The shot is 4.13 s in duration. The narrative soon returns to the hearing in which Eduardo Saverin (Andrew Garfield) explains how he was cut out of decision-making at the new company. Then, about three minutes after the first shot of the headquarters, a new scene flashes back to Mark Zuckerberg (Jesse Eisenberg) sitting at his new desk, shown in a still in the bottom panel. The backgrounds of the two shots are highly similar in color, luminance, and layout but
Zuckerberg fills more of the frame than did any of the Facebook workers in the top panel. Moreover, the shot is also only 1.16 s in duration.

Both pairs of shots show quite standard differences in film structure and, we contend, reflect the filmmakers’ tacit understanding of viewers’ psychological processes as a movie visits and revisits a given location. We will demonstrate that these presentational norms are longstanding and have been followed and modified by generations of filmmakers, yet no description of them appears in any text on filmmaking or film theory that we have encountered.¹ The first shots are longer scaled than the second shots – that is, more of the environment is seen – and they are longer in duration. Why such differences?

Hochberg and Brooks (1996, p. 261) noted that: “With any cut the [movie] viewer must make a very fast early ‘decision’ as to whether it opens a different
Fig. 2: Two stills from *The Social Network* (2010) showing the first presentation of the new Facebook Headquarters (98:05) and the second presentation (100:51) after cutting away to a scene at a hearing. Note that in the 2.35 aspect ratio of this movie unlike the 1.85 aspect ratio of *Home Alone* in Fig. 1, plenty of the background is visible, though out of focus, in the second image. From DVD, Columbia TriStar Home Entertainment.

scene or event.” We endorse this view but also endorse its complement: We believe that the physical structure of the movie aids such a decision. With Magliano, Dijkstra, and Zwaan (1996) we believe that there are strong physical guideposts to movie segmentation and understanding. In this article we explore the two visual measures noted with respect to Figs 1 and 2, shot scale and shot duration. To try to ramify and extend these examples we perform a corpus analysis of 24 films released over 70 years. We also link our results to psychological literature on reading and comprehension, analyzing them with a hybrid model from discourse processing.

2. Event Indexing, Movies, and the Given and the New

Zwaan and his colleagues (Zwaan, 1996; Zwaan, Langston, & Graesser, 1995a; Zwaan, Magliano, & Graesser, 1995b; Zwaan & Radvansky, 1998) introduced an event-indexing model applied to literature and reading. With some modifications it provides a theoretical framework for our study (see also Zacks, Speer, Swallow,
Braver, & Reynolds, 2007, for a similar approach). According both to event-indexing theory and to varied accounts of movie understanding, viewers segment the ongoing audiovisual stream into events (Bordwell, 1985; Hochberg & Brooks, 1996; Magliano, Miller, & Zwaan, 2001; Zacks, Speer, Swallow, & Maley, 2010). Film viewers must also encode those events into a model of the narrative that is continually updated over the course of watching a movie. This latter process is likely analogous to the structure-building framework promoted by Gernsbacher (1990, 1997) for language comprehension. Moreover, the whole mental process – shifting attention with the change of scenes and updating mental models – is undoubtedly related to executive function (see, for example, Miyake, Friedman, Emerson, Witzki, & Howarter, 2000). Much of how this is done, of course, remains a mystery (Graesser, Millis, & Zwaan, 1997, p. 181), although see Cutting, Iricinschi, and Brunick (2013) for a small attempt at clarification. Here, however, we focus only on segmentation and the requirements of updating representations, not on the underlying mental models.

Viewers’ respond to changes in narrative attributes called indices. That is, changes in indices force the relations within mental representations to be updated, a process that devours cognitive resources. Indeed, Zwaan et al (1995a) have shown that while reading text narrative shifts in time or in characters cause readers to slow down. This decrease in reading rate is taken as a reflection of mental model revision (although there can be other metrics; see Radvansky & Copeland, 2010), where these processes periodically place a heavier cognitive load on comprehension.

To apply event indexing to movies, we start with data on the structural units of film as determined both by filmmakers’ techniques (shot composition and structure) and by viewers’ judgments (scene segmentation). We then use the viewer segmentation data to focus on narrative changes in location, character, and time, which serve as our indexed dimensions.3 These are also the variables central to the discussion of narrative scenes (see, for example, Polking, 1995, p. 405) and that can be readily measured from the movie frames. In our analysis of films and segmentation, we make three assumptions.

First we assume that filmmakers have, over the years, contoured film form in alignment with the perceptual and cognitive abilities of viewers. Thus, by fitting a psychological model to the physical form of movies we hope to gain simultaneous insight into both filmmaking and, indirectly, film understanding. This is a brazen assumption, but we provide corroborating evidence from our results here, and from our other research on film form, in the concluding discussion.

Our second assumption is more benign. As with narrative shifts in text, we assume that shifts in a film narrative demand greater cognitive resources in the viewer, who then temporarily has less ability to process the content of what is on the screen. Thus, filmmakers need to step back (with longer scale shots; more on this later) and slow down (with longer duration shots) to accommodate the viewers’ need to update. We suggest that, drawing on a century of diffuse knowledge, craft,
and expertise, filmmakers have tacitly learned to fashion their works in this way to allow viewers to absorb the new information and to update their mental models of the narrative.

Our third assumption also has a firm foundation in the discourse processing literature. We assume that there will be differences in presentational form between old (given) and new material, where the latter should receive some emphasis over the former. Such information is known to organize sentences, paragraphs, and larger elements (Chafe, 1970; Clark & Bangerter, 2004; Clark & Haviland, 1977; Prince, 1981). We believe that the needs for the integration of old vs. new information may call for local processes in the viewer a bit like Piaget's larger notions of accommodation and assimilation (see, for example, Ginsberg & Opper, 1979): New locations, new characters, and new time frames may force more representational adjustment; returns to previously seen locations, characters, or time frames, on the other hand, may allow the viewer simply to add the incoming information to pre-existing mental structures with less reorganization.

3. Background

3.1. Shots, Shot Scale, and Establishing Shots

A shot is an unbroken dynamic display that, as a strictly structural unit, is more accurately defined by its boundaries than by its semantic content. Shots in movies are a bit like sentences in text (Carroll & Bever, 1976; Metz, 1974); they can be longer or shorter regardless of the information content they provide. Shots are separated by transitions sometimes likened to punctuation (Monaco, 1977). These transitions are abrupt discontinuities (cuts; 98.5% of all edits in contemporary movies) or more gradual replacements (dissolves, fades, and wipes; 1.5%, Cutting, Brunick, & DeLong, 2011a).4

One determines shot scale from what is shown in the frames of a shot as it depicts the person or object in focus. Fig. 3 shows a still of Tom Joad (Henry Fonda) early in Grapes of Wrath (1940) with outlines representing defined shot scales. Following convention we distinguish seven scales (Bordwell & Thompson, 2004; Salt, 2006). The first might simply show a mountain vista or a cityscape, but more often it contains a focal person or small group of people. If both the top and bottom of the frame include environmental material beyond the body of the character(s) depicted, it is an extreme long shot (1). A long shot (2) is one that is tighter in on a focal character, barely including the feet and the head. A medium long shot (3) progresses inward, showing the person only from the knees up. A shot showing the character from the waist up is a medium shot (4), and one from the chest up is a medium close up (5). A shot showing only the shoulders and head is a close up (6), and one focused on only the head or part of the head is an extreme close up (7). Obviously, this scaling discretizes a logical and perceptual continuum, but it provides a fairly unambiguous seven-category scheme for consistently coding shots in any film. It is worth stressing that we will speak of longer scale shots as those with scales 1, 2, and 3; and shorter scale shots as those of 5, 6, and 7.
Fig. 3: A still from *Grapes of Wrath* (1940), with representations of shot scale differences for seven shot classes as they relate to the human body. XLS = 1, extreme long shot; LS = 2, long shot; MLS = 3, medium long shot; MS = 4, medium shot; MCU = 5, medium close up; CU = 6, close up; XCU = 7, extreme close up. *Longer shots* are 1, 2, and 3; *shorter shots* are 5, 6, and 7. Manipulations of camera lenses alter not only the scope of the background included, shown here, but also its degree of focus. The tighter the shot on the character the less the background will be in focus. From DVD, Twentieth Century Fox Home Entertainment.

Shot scales can also be generalized to other objects or body parts. For example, it is occasionally important for the filmmaker to show the hands of a character or the details in a small object like an envelope or a smart phone. These would typically be shown in an extreme close up. Shot scales can also vary across a given shot, particularly during a pan or a zoom. For our analyses, however, we ignore these and only consider the shot scale of the first frames.

Another effect of shot scale can be seen in the lower panel of Fig. 2. If a camera moves in on a character for a tighter shot (5, 6, or 7), the camera lens is adjusted for focal length and the character remains in focus while the background becomes increasingly out of focus. This blurring removes high spatial frequencies from the background, and sometimes the foreground. Fortunately, everyday scenes can be identified and remembered on the basis of low spatial frequencies alone (De Cesarei & Loftus, 2011).
Filmmakers typically begin a scene with an *establishing shot*. This is usually a longer shot in which the camera takes in much of the surround of a given spatial location. The purpose of such a shot is to orient the movie viewer to a new environment and to the arrangement of the characters within it. As the scene progresses the camera typically includes less of the surround (and blurs what remains) focusing on the characters, their faces and emotions (Bordwell & Thompson, 2004). The focus of this article is on the relationship between the establishing shot and the *re-establishing shot*, one that revisits a previously seen location, character or set of characters, or a time frame. For purposes of simplicity we consider the first shot of a scene or subscene always to be an establishing or re-establishing shot.

### 3.2. Scenes and Subscenes

Events in life are generally defined as taking place in a single location over contiguous time (Cutting, 1981; Gibson, 1979; Zacks & Swallow, 2007), and they are in movies as well (Zacks & Magliano, 2011; Zacks, Speer, & Reynolds 2009). Scenes are a movie’s events. However, the temporal contiguity constraint for scenes necessitates a finer unit that we call the *subscene* (Cutting, Brunick, & Candan, 2012). Subscenes, particularly in action films, successively present two or more parallel, interleaved threads of the story that are crosscut with one another. That is, a typical sequence may show the protagonist in one location, then the antagonist in another; and then the two types of subscenes will alternate until the characters come together in conflict. Subscenes can also occur, particularly in older movies, with the entrance or exit of characters within a scene, which often marks a change in the tone of the story.

But how does one determine where a scene begins and ends? Such boundary questions are addressed in the film literature precisely because of potential ambiguity. Gendler (2012), for example, concluded that only a mixture of formal properties of the visual narrative and the viewer’s inferential processes could account for defining the beginning and end of a scene. We have followed Gendler’s lead.

### 3.3. Previous Results: Shot Analyses and Segmentation

In explorations of the physical parameters of movies and how they have changed over 75 years, Cutting, DeLong, and Nothelfer (2010) and Cutting, DeLong, Brunick, Iricinschi, and Candan (2011c) measured the durations of all shots in 160 popular, English language movies. These were among the most popular films of their release years as determined from the Internet Movie Database (IMDb, [http://www.imdb.com/](http://www.imdb.com/)). From that set of movies Cutting et al (2012) selected 24, three each from release years 1940 to 2010 at ten-year intervals – one action film, one comedy, and one drama, as categorized on the IMDb. Those movies appear in the appendix of Cutting et al (2012), but many are listed in the filmography at the end of this article. Each shot was then digitally analyzed for its average motion, luminance, and color, and manually measured for its shot scale.
Cutting et al (2012) then had eight viewers, three per movie, segment the movies into events with no specific instruction as to what an event might entail. Without consultation viewers agreed on their segmentations about 91% of the time across the 24 films (median [KAPPA] $\kappa = .56$ across the 72 pairs of observers). Among the physical variables of shots – shot scale, shot duration, motion, luminance, and color – they found that by far the two most potent predictors of scene segmentation were shot scale and shot duration, which provided the foci of analyses for this article.

As a result of viewers’ segmentations Cutting et al (2012) discovered that scenes and subscenes have a reasonably stereotypical structure. Although they can be encompassed in a single shot or in a sequence of fifty or more, scenes have a median length of seven shots. Cutting et al normalized the shot-duration profiles in scenes to a single standard, first affine transforming and then averaging them. In the left panel of Fig. 4 we plot a new version of the resulting variation as a function of the proportion of time through the movie event. These findings are grouped by release years – six movies each from 1940 and 1950, from 1960 and 1970, from 1980 and 1990, and from 2000 and 2010. It is easy to see that the more recent the movie, the stronger is its tendency to use closeups. Nonetheless, the overall pattern is strikingly similar across release years; scene and subscenes tend to have a stereotypic shot-scale arc.

Most importantly, across all movies the average scale of the first shot is longer than the subsequent shots. That is, the first shot is close to a medium-long shot (MLS = 3 in Fig. 3) and those throughout the rest of the scene or subscene cluster closer to a medium shot (MS = 4 in Fig. 3). The occasional use of a long shot to end a scene, suggested in the downturn at the end of each function, is also part of standard film practice, often signifying a change in tone (Mercado, 2010).

Cutting et al (2012) also normalized shot-duration profiles within scenes and subscenes in the same way as for shot scale. They then averaged them within a movie, then across movies, and we show a new version of those results in the right panel of Fig 4, for the same four periods of film release. Notice again the variation by release years, with considerably shorter-duration shots used in more recent movies. Notice also that the pattern is consistent across decades – another arc-like pattern – and that the first shot tends to be about 30% longer in duration than the subsequent shots, which have a fairly uniform length. The exception is the final shot, which is also longer in duration.7 By convention, and so as not to be confused with the longer shots of shot scaling, longer-duration shots are called longer takes.

4. Methods

For this article, we started with the segmentation results for the 24 films of Cutting et al (2012). We sought to determine whether there was variation in two measures – shot scale and shot duration – of the establishing and re-establishing shots of the segmented scenes. Thus, we needed to perform a number of new coding
analyses. First, we went back through each movie, firming up the choices of scene boundaries. We chose all those boundaries agreed upon by at least two of the three viewers and added a few that one viewer had determined. This latter group never totaled more than about 5% of the original number of scene or subscene boundaries. Second, we logged and categorized all spatial locations, characters, and time shifts. The visual presentation and re-presentation of these over the course of an entire movie allows an analysis of the differences and variability across establishing and re-establishing shots.

The revised viewer segmentations of these movies yield a mean of 133 scenes and subscenes. Since these movies average 118 minutes in length, the mean duration of a scene or subscene is just under a minute. Unsurprisingly, because of ongoing parallel plot lines, our action films have more scenes and subscenes (mean = 189) than our comedies or dramas (109 and 128, respectively). Also, our movies released from 1940 to 1970 have fewer scenes (92) than those from between 1980 and 2010 (166). Since our analyses are based on spatial locations, characters, and temporal shifts, we needed next to count these in each movie.
4.1. Coding Location Shifts

Spatial locations, although intuitively non-problematic, are hard to define. Magliano et al (2001) suggested that a location is the visibly and memorably segregable environment shown in the shots of a scene. We agree. In addition, we require character- and action-based distinctions among locations. Consider some examples.

The different rooms of a house are different locations if different actions take place within them, particularly if there is no smooth transition between rooms; but if a character simply walks through the rooms, the house is a single location. By extension, walking, running, or driving through any region is counted as a single location. However, longer journeys might be broken into sections of movement and rest, and each movement and rest section counts as a separate location. As with rooms, the outside and the inside of any enclosure are different locations if different actions take place in each.

Doors and doorways are particularly important as boundaries (Radvansky, Tamplin, & Krawietz, 2010). Burch (1973) identified six offscreen spaces: the left, right, top, and bottom of the setting as presented on the screen; the areas beyond the movie set, and the area behind the camera. The region beyond a closed door most often coincides with the area beyond the set (Burch, 1973; Saxton, 2007). Indeed, as Bordwell and Thompson (2004, p. 259) noted: “The use of the fifth zone of off-screen space, that behind the set, is, of course, common; characters go out a door and are now concealed by a wall or a staircase.” Thus, a scene that follows another with a closed door between them is a separate location. Rides shown in elevators are separate locations from the floors above and below. The inside of a single subway car is a separate location, but not if the doors of separate cars are opened between them.

A telephone call typically reveals two locations in one scene, one location for each speaker. Locations from a character’s dream or from memory count in the same way as those in diegetic (cinematic) “reality.” Montage sequences of several shots separated by dissolves are a single, if diffuse, location. Different places in a large room, such as a restaurant or an airport entrance hall are the same location. However, peculiar “enclosures” (e.g. Transportation Security Administration screening stations) are separate locations and also divide the locations on either side. Similarly, events taking place on a sports field versus the sidelines or in the stands are different locations. On the other hand, scenes in undifferentiable natural environments — different parts of a forest, a mountain terrain, or a watery environment — are coded as the same location (Magliano et al, 2001).

Nonetheless, we did not count every novel point of view of the camera as a separate spatial location. In particular, we did not include certain autonomous shots (Metz, 1974), a special kind of establishing shot that can precede the core of a scene. Examples include showing the outside of an airplane in flight or the aerial shot of a cruise ship before cutting to a conversation among staff or passengers; or showing
the outside of a tall building before cutting to office workers inside on a high floor. Because these shots are signals of location shift and present no relevant characters or plot-related action, we excluded them from our location analyses, but retained them as they denote the beginning of, and are part of, a new scene.

We developed these general guidelines by identifying locations in our 24-movie sample and then crosschecking for consistency. We recorded which scene-opening shots showed locations new to the movie, and which revisited previously shown locations. We counted all the different locations used in the scenes of these movies following the guidelines listed above. The average number of locations was 31. By far the most, however, appeared in Goodfellas (1990) and Inception (2010) with 74 and 66 locations, respectively. The fewest used were in older movies – 14 each in The Flame and the Arrow (1950) and Harvey (1950). There are no reliable differences across genres but, as suggested from the examples above, movies released from 1980 to 2010 have almost twice as many locations (39) as those from 1970 and before (21).

The most frequently used spatial location in a movie appears in about 18% of all scenes. The range, however, is large. The courtroom in Inherit the Wind (1960) appears in just over 50% of all of its scenes. This value is hardly surprising for a movie that re-enacts the Scopes trial and the early 20th-century controversy of teaching evolution in a public school classroom. At the lower end of the distribution is the outer warehouse at the first dream level in Inception, which appears in only 5% of scenes. This too is unsurprising since that movie dodges back and forth among many locations across four levels of dream states. Across all of our movies, the average frequencies of the second- through fifth-ranked scenes are 12, 9, 7, and 6%, respectively.

These data force significant constraints on film structure. In particular, movies return to previously shown locations for an average of 65% of all scenes and subscenes. The range here is also considerable. The lowest values are 41% and 43% for Goodfellas and Grapes of Wrath, respectively. Basically, these two movies are about journeys: the latter is geographical and the former metaphorical through a young life. We suggest that locations would always repeat less in journey narratives. The highest values for repeated locations occur for action films – 82, 79, and 74%, respectively, for Die Hard 2 (1990), Inception, and Mission: Impossible II (2000) – and for the older comedies Harvey (78%) and Philadelphia Story (1940, 75%). The values for action films are understandable, since they often cut back and forth between narrative threads. The values for older comedies are also unsurprising. As adaptations from live theater, they generally follow theater-production constraints and present the same physical spaces many times with subscenes of characters moving in and out of those few locations.

4.2. Coding Character Shifts

Characters are easy to determine; they typically have a constant identity across a movie. Across the set of 24 films we found that only ten characters appear
in more than 8% of all scenes and subscenes of a movie. More importantly, the most frequently appearing character (typically the protagonist) is present in 60% of them. Again the range is high. Robert Dupea (Jack Nicholson) appears in 95% of the scenes in *Five Easy Pieces* (1970) and Nick Marshall (Mel Gibson) appears in 92% in *What Women Want* (2000). However, at the other extreme, Luke Skywalker (Mark Hamill) appears in only 31% of scenes in *The Empire Strikes Back* (1980) and Danny Ocean (Frank Sinatra) in only 34% of those in the original *Ocean’s Eleven* (1960).

The second most frequent character in a movie appears in an average of just over 35% of all scenes; the third, fourth, and fifth characters appear in about 30, 25, and 20%, respectively, and so forth. There are no reliable differences in this pattern across release years or genres. Most importantly, these results show that there are not very many characters to remember when watching a typical movie.8

Shifts among main characters across scenes or subscenes, whether by addition or subtraction of characters, or by complete change, were all coded as character shifts. In sum, across these movies and by our counts, the characters across a scene or subscene shifted for 80% of the all scene changes, more so than locations. Again, the range is considerable, from 24% in *Nine to Five* (1980) to 97% *Die Hard 2.*

4.3. Coding Time Shifts

To record shifts in time, we first counted temporal ellipses (skips forward in time), and flashbacks and flashes forward. To these we added what Messaris (1994) called reality shifts – going back and forth between depicted memories and diegetic reality, as well as dream states and diegetic reality. Together, time shifts also co-occur with location shifts in nearly 40% of the cases, and the most of the remainder occurred with character shifts. The range across these movies is also considerable, from 0% in *Die Hard 2* to 70% in *Inception* and 77% in *Ordinary People* (1980).9

But frankly, shifts in time are often difficult to discern. As Gibson (1975) and James (1890, chap 15) claimed, time itself escapes perception. Instead, perceivers record events that register as evident changes in the environment. Therefore, the only sure ways to note these in movies are through marked shifts in time of day; through marked shifts in the weather; through flashbacks and flashes forward in time or through shifts in mental state from diegetic reality to dreams or memories and back, both of which typically are subsets of the cases in which a character appears in successive scenes in different and noncontiguous locales and often dressed differently; through the movie’s use of fades or dissolves, which traditionally denote a time shift (Cutting et al, 2011b); or through written overlays (such as “Four Years Later”). Most problematic are cases where different characters appear in successive scenes. It is often ambiguous whether time has elapsed between them, and we almost always coded these as parallel action, with no shift in time.10
5. Index Change and the Establishment and Re-Establishment of Scenes and Subscenes

5.1. Model Preliminaries
Because of the large differences in mean shot scales and mean shot durations across the years of our sample (seen in Fig. 4), we first normalized both of these within each movie (mean = 0, standard deviation = 1), and then combined data from all 24 movies into one large file with 31,037 shots and 2994 scenes and subscenes. Three independent, ordinal variables were constructed to reflect index changes for every shot in every film – one each for spatial location, character, and time shifts. The lack of a location shift across shots was coded as 0 (n = 28,501), the reappearance of an old location was coded as 1 (n = 1791), and the appearance of a new location coded as 2 (n = 745). Similarly, no shift in characters (0, n = 28,480), the re-appearance of a character (1, n = 2376), and the appearance of a new character (2, n = 181) comprised the second variable. And for temporal changes, codes were applied to nonshifts in time (0, n = 29,906), any apparent shift to a time frame previously seen (1; the recurrence of a dream sequence, a replay of an event, or the return to a particular mental state; n = 336), and any shift to a time frame new to the narrative (2, n = 795). There were only three movies in our sample – Ordinary People, Inception, and The Social Network – that contained many flashbacks and flashes forward to previously seen time frames. Indeed, these three films accounted for 98% of all temporal codes of 1.

The three codes were highly intercorrelated. Location and character shift codes were most closely aligned (r = .75, t(31,306) = 202.4, p < .0001, d = 2.28); location and time were next (r = .60, d =1.50); and characters and time followed (r = .48, d = 1.08). These values have statistical consequences; the three variables compete with one another for significance and often reduce the potency of one another.

Using normalized shot scales and shot durations as dependent variables we then performed least-squares multiple regressions with release year, genre, and the three indices of location, character, and time shifts as independent variables. Because of the very large number of degrees of freedom we chose a very stringent criterion (α = .0001) for assessment.

5.2. Results
Mean results and standard errors are shown in Fig. 5 for the raw (non-normalized) data based on the means taken within each film and then across films. Assessments of statistical significance in the figure, however, are based on the least-square regressions on the normalized data, and the means given below are derived from them as well.

After normalization there are no reliable effects of release year or of genre for either shot scale or shot duration. On the other hand, the upper left panel of Fig. 5 shows substantial effects on shot scale of location shifts. Establishing shots are
Fig. 5: The main results of this study concerning the shot scales (left panels) and shot durations (right panels) for the establishing shot (the first shot of a scene in a new location, character, or time frame in a movie), the re-establishing shot (the first shot of a scene or subscene returning to a location, character, or time frame previously seen in a movie), and the remaining shots of each scene. Error bars correspond to standard errors for the means of each movie. Numbers for the ordinate of shot scales come from Fig. 3. Asterisks indicate reliable results ($\alpha = .0001$) from least-square multiple regressions on the entire dataset, not from the movie means.

Reliably longer in scale (smaller numbers in Fig. 3) than re-establishing shots (2.77 vs. 3.24, $t(836.9) = -5.55$, $p < .0001$, $d = 0.38$) and re-establishing shots are longer in scale than those of the rest of the scene or subscene (3.87, $t(1900.8) = -11.06$, $p < .0001$, $d = 0.51$). The effect of character shifts on shot scale, as shown in the left middle panel of Fig. 5, was similar but statistically much weaker due to their correlation with location shifts. They are not reliable by our stringent criterion. Establishing shots are not different than re-establishing shots (2.69 vs 3.22; $t(178.9) = -2.39$, $p < .02$, $d = 0.36$), and the latter are marginally different than those of the rest of the scene (3.22 vs 3.86; $t(2516.3) = -3.53$, $p < .0004$, $d = .14$). Finally, consider the temporal-shift results. Only seven films had shifts to previously seen time frames, and again three of these (Ordinary People, The Social Network, and Inception) account for almost all of them. Thus, no meaningful across-corpus comparisons can be made for shifts involving them. However, shifts to new time frames are greatly
different than nonshifts (2.86 vs 3.64, respectively; \(t(808.1) = -4.36, p < .0001, d = .31\)) as shown in the lower left panel.

Results for index changes and their effects on shot duration are similar. As shown in the upper right panel of Fig. 5, there is a reliable duration difference for spatial location shifts between establishing and re-establishing shots (15.3 vs 12.2 s; \(t(801.9) = 8.32, p < .0001, d = 0.59\)), and a reliable difference between the latter and remaining shots within a scene (12.2 vs 6.79 s; \(t(1870.9) = 7.93, p < .0001, d = 0.37\)). There is no reliable difference in shot duration as a function of new vs. old characters (14.8 vs 12.6 s; \(t(178.6) = 1.08, p > .50, d = .16\)), but there is a reliable difference between the shifts to previously seen characters and the residual non-initial shots in scenes and subscenes (12.6 vs 6.8 s; \(t(2454.5) = 9.91, p < .0001, d = 0.40\)). Finally, and again, there are too few films with shifts to previously seen time frames, but the difference between shots that shift to new time frames and those without shifts is highly reliable (codes 0 vs 2; \(t(808.1) = 8.44, p < .0001, d = .59\)).

Regressions were run on individual films as well. Taken alone, codes for spatial location shifts (0, 1, 2) reliably predict differences in shot scale in 22 movies, codes for character shifts (0, 1, 2) also reliably predict scale differences in 22, and codes for time shifts (0, 1, 2) predict reliable differences in 18 \((\alpha = .0001)\). However, when shifts in location, characters, and time are regressed together against shot scale, seven movies show reliable effects of location shift, only three show reliable effects of character shift, and none a reliable effect of time shift. These results are generally consistent with those seen in Fig. 5.

Codes for shifts in spatial location, characters, and time reliably predict shot duration differences in 19, 20, and 13 of the movies, respectively. However, when all three variables are regressed against shot duration, location shifts are reliable predictors in seven movies, character shifts are reliable in four, and time shifts in one. These results are also generally consistent with those in Fig. 5.

5.3. An Evolution in Hollywood Movie Making

In a final analysis on our sample movies, we looked for modifications over the 70 years in how scene shifts are implemented. Since shifts in spatial location, characters, and time are highly correlated, we combined them. That is, we recoded as 2 \((n = 1226)\) shifts to a new location and/or character and/or time frame; recoded as 1 \((n = 1768)\) whether the shift was to previously seen locations, characters, or times; and recoded the remainder of all shots as 0 \((n = 28,043)\). We call the variable for shifts to previously seen locations, characters, and times Familiar Change (1); and the shifts to new locations, characters, or times frames Novel Change (2). More particularly, we were interested in Release Year interactions of Familiar and Novel Change for shot scale and shot duration. We report these analyses for both the raw data and the normalized data, shown in the left and right panels in Fig. 6, respectively.
Fig. 6: Regression lines and 95% confidence intervals for raw (left panels) and normalized (right panels) shot scales (upper panels) and shot durations (lower panels) in the combined dataset of all 24 films. Shot scales are exemplified in Fig 3. Normalization entailed standardizing (mean = 0, standard deviation = 1) values within each film, and then averaging those values across films. Nonshifts (0) are shots that continue a scene or subscene; Familiar shifts (1) are those to previously seen locations, characters, or time frames; Novel shifts (2) are those to new locations, characters or time frames not previously seen in the movie.

The left panels show the clear main effects of Release Year over 70 years towards shorter shot scales (t(31,026) = 25.66, p < .0001, d = .29) and durations (t = -41.10, p < .0001, d = .48), as was also shown in Fig 4. The normalized data in the right panels allow one to pull apart effects not otherwise easily discerned. In the upper right panel the shot scales for nonchange shots are quite different from those for Familiar Change (0 vs 1; t(1844.9) = -18.99, p < .0001, d = .88) as are those for Familiar Change vs. Novel Change (1 vs 2; t(1375.7) = -8.84, p < .0001, d = .45). More
importantly, however, there is a reliable interaction of release year with the various shifts (codes 0, 1, and 2) for shot scale \( (F(2;31,031) = 9.31, p < .0001) \), which plays itself out mostly in the contrast between nonchange and Familiar Change (0 vs 1, \( t(1844.9) = 3.30, p < .001, d = .15 \)). In particular, although not readily discernable in the upper-left panel, the upper-right panel shows that, beyond the general shortening of shot scale over the same period, the shot-scale differences between shifts and nonshifts have diminished from 1940 to 2010.

The lower-right panel of Fig. 6 shows reliable differences for the durations for nonchange shots compared to those for Familiar Change (0 vs 1; \( t(1844.9) = 20.62, p < .0001, d = .92 \)), and in turn those for Familiar Change vs. Novel Change (1 vs 2; \( t(1375.7) = 10.48, p < .0001, d = .57 \)). But there are two interactions in the normalized shot duration data that are more interesting and barely visible in the raw data of the lower-left panel. That is, there is a reliable interaction of Release Year with nonchange vs. Familiar Change (0 vs 1; \( t(1844.9) = -6.15, p < .0001, d = .29 \)) and another interaction in the reverse direction for Familiar Change vs. Novel Change (1 vs 2; \( t(1375.4) = 4.39, p < .0001, d = .24 \)). This means that, beyond the overall decreases in shot durations over 70 years seen in the lower-left panel, Familiar Change shot durations are approaching the nonchange durations. However, Novel Change durations do not follow this pattern, and thus remain longer in proportion to the nonchange durations.

Thus, over the last 70 years we find that movie form has been modified in two previously unknown ways. First, not only have mean shot scales decreased (Cutting et al., 2012; Salt, 2006), but the scales of establishing and re-establishing shots have decreased relative to noninitial shots of scenes and subscenes. This suggests that filmmakers have found that viewers can generally get by with relatively less environmental information than they previously thought. Second, not only have the mean shot durations decreased (Bordwell, 2006; Cutting et al., 2012, Salt, 2006), but the durations of re-establishing shots have also decreased in their proportion to noninitial shots. This suggests that filmmakers have found that returns to previously seen locations, characters, and time frames do not require the proportional durations that establishing shots still require.

6. Shot Scale and Boundary Extension?

Finally, it is tempting to link our shot scale data to boundary extension (Intraub & Bodamer, 1993; Intraub & Richardson, 1989), a robust phenomenon that occurs across many methodologies (Hubbard, Hutchison, & Courtney, 2010) for the memory of pictures. After a set of test images has been presented to viewers, and when they are given a choice among images to recognize, they systematically choose one including more background than the original. Viewers falsely remember having seen information that is “likely to have existed just outside the camera’s field of view” (Intraub & Berkowitz, 1996, p. 581). This additional information is never visually available to them during the experiment. When viewers are shown the complete images actually presented during the experimental training phases, they
recognize the content of the image as familiar, but report that “it shows a closer view than did the original” (Intraub, 1997, p. 219).

In a movie context, this broader information is visually available for a new location. It is therefore tempting to infer that the establishing shots provide ready-made schematic encoding of locations that in turn facilitate recognition of subsequent narrower views. It seems intuitive that re-establishing shots are encoded within the same schematic blueprint as the establishing shot. However, one might challenge the application of the boundary extension to movies because almost every return to a location entails a shift in camera angle. The situation shown for Home Alone in Fig 1 is an exception; a more typical return to a location is shown for The Social Network in Fig 2 where one sees the environment from a different angle.

Nonetheless, behavioral evidence may contravene this problem. After being trained on the same scenic content presented from different angles, picture viewers can readily identify the content as seen before, but are unable to identify which of images from different camera angles they had seen and which they had not (Hock & Schmelzkopf, 1980). This finding suggests that boundary extension may apply to the different points of view in the re-establishing shots of movies.

7. Summary and Conclusions

In popular movies, the establishing shot is usually the first shot of a scene in a new location, and a re-establishing shot is the first of a scene returning to a previously shown location (Bordwell, Staiger, & Thompson, 1985; Bordwell, 2006). Nonetheless, nothing in the film literature suggests these should be different in scale or duration. In this article we adapted and modified an event-indexing model from discourse processes (e.g. Zwaan & Radvansky, 1998) to include the consideration of old (given) and new information (Chafe, 1970; Clark & Haviland, 1977) and presented four new results centered on location shifts across the narrative: (1) re-establishing shots include less background than establishing shots, but (2) more background than the average of remaining shots of a scene or subscene, and (3) re-establishing shots are shorter in duration than establishing shots, but (4) longer than the remaining shots. Analogous shifts for characters or time frames are similar, but generally not reliable by our stringent criterion.

It might seem surprising that shifts in location outstrip shifts in characters in this aspect of the physical form of movies. After all, movies are about characters and characters are the most important focus of any story (Özyürek & Trabasso, 1997; Scott Rich & Taylor, 2000; Zunshine, 2012). We have no doubt that characters are more vital to narratives than either locations or time, but the issue here is about information pickup, not importance. A character’s actions always take place in spatial locations and these locations are, by necessity, physically “larger” than the characters themselves. That is, for a changed location to be registered and remembered while the character acts within it, the character must temporarily assume a relatively “smaller” stature (placed in a longer shot). When a new location
is shown, it must be made salient by further lengthening shot scale. Moreover, we assume that time is needed to register these locations. After all, the film viewers will be looking at the character, not typically at the surrounding location (Smith, 2012), whether new or old.

Finally, we found that, beyond the general shortening of all shot scales and durations over the last 70 years, re-establishing shots have also gotten relatively shorter in both dimensions, suggesting an evolution in filmmaking that capitalizes on viewers’ abilities to shift gears rapidly, assimilating information about previously seen locations, characters, and time frames. Establishing shots have also gotten relatively shorter in scale, but not in duration, implicating a continued temporal need for viewers to mentally shift gears when new information is presented.

Obviously, our results come from corpus analyses; they stem solely from stimulus correlations after obtaining behavioral data from scene and subscene segmentation. We have no direct measures of shot scales and durations needed for viewers to process narrative shifts in a movie. Nonetheless, we believe that over the last 70 years and more the structural patterns in popular movies have been transformed as filmmakers have discovered better and more economical ways to engage viewers’ perceptual and cognitive systems. Indeed, we have found decreases in shot durations (Fig 6 and Cutting et al, 2011c; see also Bordwell, 2006; Salt, 2006), which force more frequent reallocations of attention (Mital, Smith, Hill, & Henderson, 2011; Smith & Henderson, 2008); increases in the match between exogenous attention fluctuations demanded by cuts (Cutting et al, 2010) and those driven endogenously in psychological tasks (Gilden, 2001); increases in the variation and the amount of motion (Cutting et al, 2011b; Cutting, 1914), which captures attention (Mital et al, 2011) and heightens physiological responses (Ando et al, 2002;); increases in the coupling of shot duration and motion (Cutting et al, 2011b), which yoke attention to those physiological responses (Soleymani, Chanel, Kierkels, & Pun, 2008); changes to shorter shot scales (Cutting et al, 2012; Salt, 2006), which magnify the expressions and perceived tensions in characters; and decreases in overall luminance with attendant increases in within-frame contrast (Cutting et al, 2011c; Cutting, 2014) to control attention and eye movements (Smith et al, 2012). To these functionally supported modifications in movie structure, we would add the modulations of shot scales and durations across new and old locations in service of viewers’ ability to shift attention and update narrative schemata.

Acknowledgments

The authors are members of the Psychology Department at Cornell University. Requests for information should be sent to J. Cutting, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853-7601 USA, or to james.cutting@cornell.edu. We thank Kate Brunick, Ayse Candan, Jordan DeLong, David Field, and Khena Swallow for discussion; and Michael Masucci who first undertook coding scenes for us.
References


Erlbaum.

Gibson, J. J. (1975). Events are perceivable but time is not. In J. T. Fraser & N. Lawrence (Eds.) *The study of time, II* (pp. 295-301). New York: Springer Verlag.


**Filmography**


*Films in our sample
Endnotes

1 The closest description we have found concerns 1920s Hollywood cinema. Bordwell, Staiger, & Thompson (1985, p. 58) suggest no difference between the two: “The reestablishing shot” as the first shot in a previously seen location “will usually be from the same angle and distance as the establishing shot.” Later, they discuss changes in position (perhaps implying changes in shot scale) that occurred later in cinema but there is no mention of differences in shot duration.

2 Movies are not alone in providing visual guideposts to understanding. Ample evidence for reader-targeted discourse structure comes from the development of punctuation in text. Written text in antiquity was meant to be heard, listened to; words addressed the ear, not the eye (Parkes, 1993). Early accounts of silent reading – with its first documented instance in the 4th century – describe the act as either peculiar (for example, St. Augustine’s account of St. Ambrose’s voiceless reading) or, if self-reported, explanatory and almost apologetic (for example, 6th century Isidore of Seville’s account of the effortless and reflection-inducing silent reading). This individual, quiet act of reading led to the development of a punctuation system that facilitated information extraction by the eye (for example, word and paragraph separation as we know it today). Although punctuation existed since the 3rd century BC (contrived by Aristophanes of Byzantium), the first signs of punctuation that addressed meaning (and not sound and breathing patterns) appeared about five hundred years later. Our current punctuation system, which grants meaning and structural information, took another two thousand years to come into being with the advent of the printing press in Western Europe (Fischer, 2003). Text editing thus experienced a rather long history of changes increasingly reflecting the readers’ visual – and silent – engagement with the written word (Cobley, 2001; Parkes, 1993). Analogously, changes in film editing reflect advances in technology and viewers’ watching strategies.

3 Semantic and narrative content indices such a causation and goals cannot be analyzed as a formal feature of the movie from the visual stimulus.

4 Indeed, Monaco (1977, p. 192) suggested that most of the transition types were like periods, but that “If there is a comma in films amongst this catalogue of periods, it is the dissolve.”

5 In his forthright account of filmmaking, Lumet (1995, p. 149) refers to the strict rules that governed Hollywood film production in the 1930s and 1940s. For example, “every scene had to be ‘covered.’” This meant there was progression of shooting from an extreme long shot to close-ups. Scenes were then compiled in the cutting room in roughly the same scale order.

6 Bellour (1976) also identified these units, calling them segments and subsegments, respectively.

7 This last shot may present what is called a dangling cause (Bordwell & Thompson, 2004), an intentional pun on part of an ill-formed sentence. The dangling cause presents information critical for the next or some future scene. Nonetheless, we will not consider this
last shot further; we will simply average it with the other shots that follow the (re)establishing shot.

To be sure exceptions do occur, for example in what are called network narratives (Thompson & Bordwell, 2010). Grand Hotel (1932), Nashville (1975), and Valentine’s Day (2010), among others, are all films in which many more than the usual number of characters and plot lines are followed in a network of interactions. In Valentine’s Day, for example, one is asked to keep track of twenty main characters and their inter-relations. But even in this case there are only twelve that appear in more than 8% of all scenes, and the most prevalent character, Reed Bennett (Ashton Kutcher), appears in only 22% of all scenes, well below that for a typical popular film. Thus, network narratives take away scenes from central characters more than they add significantly to those of less central ones.

No movie in our sample had a jump cut, typically defined as a short ellipsis within what would otherwise be a longer shot within a scene.

There is an argument from philosophy and the philosophy of physics that time can change without a change in space, but that there can be no change of space without a change in time (e.g. Sherover, 1975). For example, Croll (1897, p. 249) suggested that the "Lapse of time is necessarily implied in the very conception of change... There can be no change without change of position in time." However, our views on mental shifts ally with Tulving (2004, p. 7), who suggested that “Mental ‘space travel’ (imagining different spatial locations) does not require mental time travel (imagining oneself at different times).”