Revisiting the Kuleshov Effect with First-Time Viewers

Sermin Ildirar and Louise Ewing

Abstract: Researchers have recently suggested that historically mixed findings in studies of the Kuleshov effect (a classic film editing–related phenomenon whereby meaning is extracted from the interaction of sequential camera shots) might reflect differences in the relative sophistication of early versus modern cinema audiences. Relative to experienced audiences, first-time film viewers might be less predisposed and/or able to forge the required conceptual and perceptual links between the edited shots in order to demonstrate the effect. This article recreates the conditions that traditionally elicit this effect (whereby a neutral face comes to be perceived as expressive after being juxtaposed with independent images: a bowl of soup, a gravestone, a child playing) to directly compare "continuity" *perception* in first-time and more experienced film viewers. Results confirm the presence of the Kuleshov effect for experienced viewers (explicitly only in the sadness condition) but not the first-time viewers, who failed to perceive continuity between the shots.

The Kuleshov effect is a film-editing effect that was demonstrated during the late 1910s and early 1920s by the pioneering Russian filmmaker and theorist Lev Kuleshov (1899–1970). Famously, Kuleshov is reported to have intercut a close-up of the Russian actor Ivan Mozhukhin's neutral, expressionless face with various other camera shots, including a bowl of soup, a woman in a coffin, and a child playing with a toy bear. He observed that these additional

shots interacted with the original, leading viewers to perceive the (objectively neutral) face as expressing happiness, sadness, and hunger/thoughtfulness, respectively (Pudovkin 2013). As the years have passed, the reliability and validity of this effect have come into question. The original footage used by Kuleshov has long since been lost, and superficial issues with the design of the experiment[1] have prompted some to reclassify it as part of the "mythology of film" (Holland 1989) or the "folklore of the cinema" (Pearson and Simpson 2005). Yet, this disapproval may be unwarranted. [Callout 1 about here]

Despite the somewhat anecdotal nature of Kuleshov's original observations, other (more rigorous) studies provide converging evidence that a single film scene can generate a profoundly different perceptual meaning for viewers when placed in different contexts. Herman Goldberg (1951), for example, found that the emotional quality and intensity of a fearful face accompanied by a scream can differ depending on the order of camera shots (e.g., it can come to be perceived as rage or even joy). Similarly, studies by J. B. Kuiper (1958) and J. M. Foley (1966) (as cited in Isenhour 1975) demonstrate that neutral faces can be perceived as happy or sad, depending on their contexts in films. Support has also come from psychological studies utilizing brain-imaging (Mobbs et al. 2006) and eye-tracking (Aviezer et al. 2008; Barratt et al. 2016) techniques during the viewing of edited film clips. Dean Mobbs and colleagues (2006) observed differential neural responses (e.g., in the bilateral temporal pole, superior temporal sulcus, and anterior cingulate cortex) when identical faces were paired with different emotionally salient contextual movies. At the end of the scanning session, they also asked their subjects to judge the faces. Despite the fact that the faces were identical, attributions of facial expression and mental state were altered when the faces were juxtaposed with contextual movies of different valences. Hillel Aviezer and colleagues (2008) reported that the pattern of participants' eye movements to facial regions

changed systematically as a function of the affective context in which these images appeared.

The most recent replication (and extension) of the Kuleshov experiment was conducted by Daniel Barratt and colleagues (2016: 865); they concluded that "some sort of Kuleshov effect does in fact exist." These authors considered the original film sequences to be an instance of point-of-view editing, so they carefully constructed their set of test stimuli to encourage participants to infer that the glance shot and the object shot were spatially related (i.e., the gazer did not look directly into the camera). Their results confirmed that the emotional context influenced participants' judgments of the target face stimulus in each of the five emotional conditions (happiness, sadness, hunger, fear, and desire), with the most pronounced effects observed for sadness.

Importantly, however, previous replication attempts have been less successful. Stephen Prince and Wayne Hensley (1992) found that the majority of their subjects reported seeing an actor with a neutral expression (i.e., no editing-induced appearance of emotion), regardless of the sequence into which his face was edited. These authors suggested that the "naïveté of early cinema audiences," compared with their more experienced, modern participants (university undergraduates), might explain their original findings. [Callout 2 about here]

To our knowledge, there has been no empirical study of the Kuleshov effect with naïve participants. However, there have been anecdotal reports (Forsdale and Forsdale 1966) and direct investigations of their perception of other aspects of editing (e.g., Hobbs et al. 1988; Schwan and Ildirar 2010). Renee Hobbs and colleagues (1988) compared single-shot recordings with edited versions of the same content, and reported no effect of editing on comprehension in first-time viewers. Crucially, however, more recent studies with first-time viewers (Schwan and Ildirar 2010; Ildirar and Schwan 2015; Ildirar et al. 2017) have found that participants' familiarity with

the depicted content can powerfully modulate this effect. In these studies, first-time viewers struggled to construct a spatiotemporal relationship between adjacent shots (e.g., shot reverse shot, outdoor to indoor shot). Instead, they perceived adjacent camera shots as independent images. For example, a shot-reverse-shot sequence of a man looking right, followed by a shot of a man looking left, with both actors shown against the same scenic background, was not interpreted as 'Two men looking at each other' but, instead, as two completely independent scenes: 'First, there was a man, then he was gone, and then, another man appeared (see Figure 1). However, an ongoing line of actions that they were familiar with, a salient gaze cue, or clear dialogue helped naïve viewers to perceive continuity between adjacent camera shots. Given that the film clips that have been historically used in Kuleshov experiments do not include any such cues (relying instead on participants' connecting of the shots together through emotion), it remains an open question whether this editing effect will help naïve viewers to perceive a spatiotemporal relationship between the adjacent shots. In order to answer this question, we conducted a field experiment that attempted to elicit the Kuleshov effect with a unique sample of first-time film viewers and a comparison group, both from regional Turkey.

<Figures 1a & 1b around here, side by side>

Figure 1. Example of a sequence from Stephan Schwan and Sermin Ildirar (2010). First-time viewers could not construct a spatiotemporal relationship between adjacent shots (video stills by Sermin Ildirar).

The Kuleshov-Type Sequence as an Instance of Artificial Landscape

There are two components to the Kuleshov effect: perception of spatiotemporal continuity between the juxtaposed camera shots and the perception of a change in emotion of the target

(neutral) face. Although the first component is a critical prerequisite for the latter, it is rarely directly considered or discussed in any detail. An exception is a consideration raised by David Bordwell, Kristin Thompson, and Jeremy Ashton (2004): they argue that the Kuleshov effect can arise when any series of shots that, in the absence of an establishing shot, prompts the spectator to infer a spatial whole on the basis of seeing some of the spatial parts. Here, the authors are describing the concept of artificial landscape without actually naming it. [Callout 3 about here]

While shooting his film *The Project of Engineer Prite* (1918), Kuleshov discovered that it was possible to create a cinematic terrain that does not exist anywhere in reality. This was the first of several properties of the montage that he described in his later articles and books. His film required shots of actors looking at electrical cables strung up on poles that had not been filmed. Kuleshov supposed that the same effect could be achieved by splicing shots of actors looking off camera with separately recorded shots of the row of poles. Since the poles and the actors were in different parts of Moscow, Kuleshov (1974) termed the effect the "artificial landscape" (also known as "creative geography"). After this discovery, Kuleshov, created other artificial landscapes in his movies. For example, he presented scenes in which actors walked up the steps of a well-known Moscow building to then arrive at the White House in Washington, DC. In one film, he even combined close-up shots of different women's body parts to create a "new" woman. In this way, he created cities, buildings, and bodies that existed only on screen.

The artificial landscape is a ubiquitous feature of modern film and television. For example, when two characters are shown in single shots looking right and left (usually in dialogue scenes), respectively, viewers readily assume that they are filmed in the same place at the same time, though this may not have been the case. A well-known example is the dialogue between David Bowie and Marlene Dietrich in *Just a Gigolo* (David Hemmings, 1978), which

was filmed with these actors individually, in separate rooms, months apart. It is interesting to note that, although the viewers of *Just a Gigolo* did not realize this production trick and perceived the shots as being in spatiotemporal continuity, first-time film viewers were not similarly fooled (Ildirar and Schwan 2015; Schwan and Ildirar 2010). These naïve viewers saw people in the same place but not at the same time.

Kuleshov-Type Sequence as an Instance of a Point-of-View (POV) Shot

Another master of editing, Alfred Hitchcock, noted that the primary editing structure of his film *Rear Window* (1954) was based on the Kuleshov effect. In the film, James Stewart's character (Jeff) is a voyeur, who peeks through his window into people's private lives. In the framing of the shots, Hitchcock consistently kept his POV shot aligned with Stewart's eyeline. Since Stewart often has an emotionally ambiguous face during the film, the views out of his apartment window powerfully drive the emotional context (Truffaut 1984: 213–223). In an interview, Stewart later claimed not to remember playing the role the way he had seen it on screen. Thus, it appears that Hitchcock's manipulation of the Kuleshov effect was so successful that he was able to alter the montage to create completely different meanings (Sharff 1997).

From this perspective, a Kuleshov-type sequence can be considered an instance of a POV shot, which is a short film scene that shows what a character (the subject) is looking at (represented through the camera). Viewers link these two images together in their minds and perceive them to be depicting a continuous moment—concluding that the subject is looking at the object.

The POV shot is one of the techniques that filmmakers discovered in the early years of cinema that helps viewers to integrate diverse views separated by cuts—in other words, it helps

them to perceive continuity through film cuts. One proposed explanation of how viewers

perceive cinematic continuity despite the spatiotemporally discontinuous nature of the visual

information presented to them is that films produce a stream of audiovisual information that is

similar to our veridical perception of real scenes and events (e.g., Anderson 1998; Bordwell et al.

1985; Cutting 2005; Gibson 2014 Lindgren 1948; Münsterberg 2013).[2] In line with this

ecological view of film cognition, explaining how a POV shot is easily comprehended by

viewers, Noël Carroll (1993) and Tim J. Smith (2012) argue that it mirrors natural attentional

shifts between a gazer and an object.

Gaze following (looking where someone else is looking) emerges in infancy as early as

six months of age to targets within a baby's own visual field (D'Entremont et al. 1997) and

within the first year to targets more broadly (Butterworth and Jarrett 1991; Corkum and Moore

1998). By twelve months, infants will turn to see what another is looking at (Tomasello et

al.1993). Adults, however, spontaneously monitor a person's eyes and use gaze direction to

support inferences about their intentions, emotions, attention, knowledge states, and likely future

actions. Indeed, although other cues such as head orientation, body posture, or even pointing

gestures may also provide important information in the determination of where gazers are

directing their attention, the information from gaze cues has been shown to be exceptionally

powerful in this regard (Perrett et al. 1992). Past research has shown that a salient gaze cue

and congruent head orientation can help even first-time viewers to construct a spatiotemporal

relationship between adjacent shots (Ildirar & Schwan, 2015). Despite not otherwise perceiving

spatiotemporal continuity – when naïve viewers were shown edited footage of a woman lowering

her head (shot one) and a pair of shoes lying on the ground (shot two), they reported that the

woman was looking down to her shoes (see Figure 2).The location of objects in the proximity of

the viewer can also influence the interpretation of gaze direction (Lobmaier et al. 2006);

however, in a Kuleshov-type sequence, these are unlikely to influence responses unless

participants perceive spatiotemporal continuity between the adjacent shots.

.

<Figure 2a & 2b around here, side by side>

Figure 2. Example of sequence from Ildirar and Schwan (2014). A salient gaze cue

helped first-time viewers to construct a spatiotemporal relationship between adjacent shots

(video stills by Sermin Ildirar).

According to Carroll (1993: 128), the fact that a head movement is replaced with an edit

does not matter, because "it is the endpoints of the activity, and not the space between, that

command our attention." Per Persson (2003) developed this theory by describing the POV

structure as an instance of deictic gaze or joint visual attention. According to Persson, the

presentation of the object in a POV scenario involves an unnatural "jump" from one optical

perspective/camera position to another. He suggested some conditions that could increase the

likelihood that the viewer will make a "POV inference," and the first of these conditions is that

the gazer should not look directly into the camera (the so-called "fourth wall" rule).[3] Perhaps

crucially, the original Kuleshov sequences did not meet this condition. Moreover, since we aim

in this article to replicate the original sequences as closely as possible, in our core stimuli the

gazer will look directly into the camera.

The technique of direct address—when a character looks to the audience—is rare in

fictional cinematic discourse, except in instances of comedy (Renov 2004: 30). However, this

technique has become increasingly popular with documentary filmmakers since the 1990s. It is

believed to stand in for what would be eye contact in daily life and increase the sense of intimacy

and/or confrontation felt by viewers (Rosenheim 1996: 221). Interestingly, a study investigating

perceptions of credibility during testimony reported that witnesses who averted their gaze were

perceived to be less credible and were more likely to be associated with a guilty verdict

(Hemsley and Doob 1978). Other studies have since found that maintaining eye contact with an

interviewer facilitates deception detection (Vrij et al. 2010). It follows, then, that looking directly

into the camera might have an effect (positive or negative) on the perception of continuity and

emotion, which are both components of the Kuleshov effect.

Kuleshov-Type Sequence as a Place for Emotion Seeds to Sprout

In everyday life, face stimuli are rarely perceived in isolation, and the context in which they

appear can be very informative. Researchers have explored three types of contexts vis-à-vis their

effects on facial emotion perception: (a) the stimulus-based context, in which a face is physically

presented with other sensory input that has informational value; (b) the perceiver-based context,

in which processes within the brain or body of a perceiver can shape emotion perception; and (c)

the cultural context, which affects either the encoding or the understanding of facial actions

(Barrett et al. 2011). The Kuleshov experiment deals with the stimulus-based context.

Emotion perception studies investigating the influence of the stimulus-based context have

shown that facial expression judgments are influenced by any number of cues, including

descriptions of social situations (e.g., Carroll and Russell 1996), voices, body postures, visual

scenes (e.g., Aviezer et al. 2008; Righart and de Gelder 2008; for reviews, see Barrett et al. 2011

and de Gelder et al. 2006), and even other faces (e.g., Masuda et al. 2008). For example,

scowling faces (posed, exaggerated facial expressions of anger) are more likely to be perceived as fearful when paired with a description of danger (Carroll and Russell 1996, Study 1) or disgusted when paired with a body posture involving a soiled object (Aviezer et al. 2008, Study 1). Aviezer and colleagues (2008) propose a model of context effects using the metaphor of "emotion seeds." They suggest that the same perceptual information might be shared by different facial expressions (i.e., emotion seeds) and lie dormant in isolated faces. This information can, however, be activated by the appropriate context.

If a given context activates a facial expression that shares enough emotion seeds with the expression displayed by a target face, these seeds will "sprout" and override the original expression of the target face. By contrast, an equally powerful context will have little impact if its associated facial expression shares few emotion seeds with the expression of the target face (Aviezer et al. 2008). In the case of naïve viewers watching a Kuleshov sequence, we hypothesize that the sprouting of seeds might function to not only help them perceive an expression on an otherwise expressionless face, but also to help them make a link between the discontinuous shots.

Method

Participants

Forty participants (half female, 56–72 years old, M = 64.1 years) took part in the study. All subjects gave informed consent, and the study was approved by the Research Ethics Committee of the Istanbul University Faculty of Medicine. The experimental group (twenty participants, half female, 58–72 years, M = 66.4 years) knew of the existence of television and had some abstract ideas about it, but had no prior direct experience with the medium. This group lived in small

isolated houses in the mountains south of Isparta, Turkey, that had only recently been connected

to the electrical grid. All of these of participants had some photos (mostly head shots of their

children or grandchildren), and four had radios with a very limited broadcast range. Many

assumed that television was a "visual radio" with programs that showed pictures of the people

who speak or sing on the radio. Seven members of the group were illiterate, and the average

education level was 1.95 years.

The control group (half female, 56–72 years, M = 61.9 years) was from a similar

geographic and cultural background as the experimental group. Critically, these participants all

had some experience with television. They spoke the same dialect and had a similar lifestyle as

the experimental group (socially and geographically isolated, working in agricultural industries),

but with a little more access to luxuries such as refrigerators, ovens and most importantly

televisions. Three members of the group were illiterate, and the average education level was 3.1

years. This control group was significantly younger than the experimental group ($F(57,2) = 3.7$,

$p = .03$), but there was no significant difference in educational level ($x^2(4) = 4.48$, $p = .3$).

Stimuli

Two sets of video clips were produced, with each set containing six two-shot sequences that

were eight seconds in length (see Table 1). In Set A, each sequence started with an

expressionless man's face; this image was followed by images of a plate of soup, a gravestone,

and a little girl. In Set B, the structure of the sequences matched the structure of those in Set A,

but here the facial expression of each man matched the intercut images: he licked his lips and

gulped to express hunger when he preceded the soup image, looked sad when he preceded the

gravestone, and smiled when he preceded the little girl. Two versions of each set were created,

and they featured different actors. We independently validated the perception of these expressions (i.e., as being neutral, hungry, sad, and happy) with a large separate group of undergraduate students (n = 80). To replicate the conditions in Kuleshov's original experiment, in both clips the actors looked directly into the camera, the sequences were in grayscale, and there was no sound.

An additional sequence was produced during testing in the field following responses from the first three experimental (naïve) participants, who strongly signaled that they were not making any connections between the intercut images. In light of these responses, we made an alternate version of the hunger sequence in Set A, where the actor was replaced with a shot of an old woman looking down and a plate of soup on a floor table, which is where these participants tend to eat their own meals.

Table 1: Overview of the Two-Shot Sequences

<Table 1 around here>

Note: Each eight-second sequence in Set A featured a face with a neutral expression, and those in Set B featured a face with an overt expression that matched the following image. Alternate versions of each sequence were created with a second actor. Set B included an additional sequence that was intended to more closely mirror the conditions in participants' lives (old woman looking down) (image stills by Sermin Ildirar).

Procedure

Participants were tested individually in their homes in sessions lasting thirty to sixty minutes. In order for us to check for possible auditory, visual, or cognitive deficits, participants were asked

to describe their present situation (i.e., what they could see outside the window). They were also interviewed about their experience with and their knowledge about television and film. No participants were excluded on the basis of these discussions.

After the participants answered the questions, a laptop with a 17.3-inch display was presented to them (viewing distance about 60 centimeters). Participants were told that they would see something on the display and be asked to describe it much as they had previously described their present (real-life) situation. The video sequences were shown in a fixed order (as in Table 1) with a short break after each presentation, in which to answer questions from the experimenter. The first question was always "Could you please tell me what you have seen?" If their answer clearly indicated an understanding of spatiotemporal continuity and/or the Kuleshov effect (e.g., "I saw a man smiling at the baby across from him"), no further questions were asked regarding spatiotemporal continuity perception. When the participants mentioned just one of the shots (e.g., "I saw a man looking at me"), they were always asked what else they saw, which usually led them to talk about the content of the other shot (e.g., "There was a man first. Then he disappeared, and there appeared a stewpan"). If the answer did not mention any connection between the shots (e.g., "I saw a gravestone too"), follow-up questions were asked (e.g., "Where was the gravestone?") until their perception of the edited sequence was clear. All the participants were also asked how the person on the screen was feeling.

Coding and Analysis

All sessions were video recorded, transcribed, and then double coded (reliability, Cohen's kappa coefficient > .92) using the qualitative analysis program Atlas-ti. Each participant's qualitative responses to each clip were numerically classified. When there was no spatiotemporal linkage

between the camera shots (i.e., no sense that the person in the first shot was in the same place or time as the objects in the second shot), the participants received a score of 0. When they *did* make a clear spatiotemporal link between the shots, they received a score of 1. When participants demonstrated a clear Kuleshov effect (i.e., perceived variation in the (neutral) facial expression of the first shot when it interacted with the content in the second shot), they received a score of 2. After the coding process, the data was transferred from Atlas-ti to SPSS, and the differences in the frequencies between the first-time viewers and the experienced viewers were tested for significance by Fisher's exact test.

Results

The percentage values reported below reflect the participants' responses averaged across the two identities that appeared in Set A and Set B (see Table 2 and Table 3).

Table 2: Perception of Spatiotemporal Continuity and the Kuleshov Effect across Groups for the First Set of Film Sequences (Intercut Faces with Neutral Expressions)

<Table 2 around here>

Set A (Neutral Faces)

First-time viewers. The first-time viewers interpreted all the sequences in Set A as independent images. Responses in this group did not suggest any spatiotemporal linkage between the shots or the existence of a Kuleshov effect on the perceived expression of the neutral face. A typical response was that there was an image of a man sitting in silence and looking toward the viewer that came and went. When asked what else they saw, participants commented that the man

15

disappeared and that something else subsequently appeared: a plate (often described as something bigger, e.g., a cooking pot or saucepan), a gravestone, or a little girl. When asked additional questions to probe their perception of these sequences, their responses revealed very limited consideration of the context of or the interaction between the shots. With regard to the perceived spatial location of these objects (i.e., when asked, "Where was the plate/man?"), they responded that the plate "should be in the kitchen" or "on the stove" or, pointing to the screen, "How can I know it? It appeared there." [Callout 4 about here]

When asked what the man was feeling or thinking, first-time viewers said that they "cannot know" that, or that "he was looking with empty eyes." When asked whether the little girl was alone, all participants answered "yes," adding that they did not see her parents next to her. The customized additional video clip added during testing, which featured a face with directed gaze (looking in the direction of the soup), helped the first-time viewers link the shots spatiotemporally. All of them reported that she was sitting at a floor table and waiting. The reasons provided for her waiting were diverse, and they related mostly to the individual backgrounds of the first-time viewers. For example, one female participant said that the woman in the video clip was afraid of her husband's anger because she did not know whether he would like the meal. Given that these attributions regarding the woman's emotion were elicited in a perceiver-based rather than a stimulus-based context, this was not considered evidence of the Kuleshov effect. [Callout 5 about here]

Experienced viewers. In contrast to the first-time viewers, 100 percent of the experienced viewers constructed spatiotemporal links between the shots in the Set A sequences. A Kuleshov effect was also observed for 55 percent of participants in the gravestone sequence.

For the soup sequence, 100 percent of participants reported that they saw a man with a

meal in front of him, with many (65 percent) also making a forward inference and saying that the man will eat the meal. When asked about how he was looking and feeling, 30 percent of participants said that he looked indecisive and was thinking about whether he should eat the meal, and 45 percent of them said that he was waiting for someone else to arrive before she started eating. The remaining 25 percent said: "Nothing special . . . he will just eat the meal." Here, the absence of motion through the cuts led the viewers to seek an explanation for the two shots (i.e., the meal would be eaten by the actor). This expectation may be explained by the dramatic principle called "Chekhov's gun." Here, every element in a narrative is required to be irreplaceable (Bill 1987). Thus, just as whenever you introduce a rifle in the first act it must go off in the second act, to give Chekhov's example, so too it seems in our case that if you show a meal in the first shot of an edited sequence it must be eaten in the second shot.

For the gravestone sequence, 100 percent of the experienced viewers made spatiotemporal links between the shots, and 55 percent demonstrated a Kuleshov effect. That is, they all said that the man was standing in front of a gravestone, and when they were asked how he was feeling, 55 percent of them said that he looked sad or sorry. Other responses were that he was praying (15 percent) or keeping a minute of silence (20 percent), which might also be considered as an interpretation of sadness, since these are what people do in memory of people who have died. Only 10 percent of the experienced viewers said that the person was feeling nothing.

For the child sequence, once again 100 percent of the participants made spatiotemporal links between the face and the second image. All of the experienced viewers reported that they saw a man and a girl. When asked where they were, participants said that they must be at home or at school. No participants showed a clear Kuleshov effect. Forty-five percent of participants

said that he felt "nothing," and 20 percent said that he was miles away and thinking of something else. Interestingly, 25 percent of the participants linked this sequence with the gravestone sequence (that preceded it) by saying that the man was trying to forget someone who had been lost by thinking that their life goes on.

For the old woman sequence, all participants reported that she was waiting before eating her meal. The reasons for her waiting were several: she was allowing the meal to get cold (10 percent); she was expecting someone to come (20 percent); and she just did not have her appetite (45 percent). The rest did not offer an explanation. When asked what she felt or thought, the participants most frequently answered, "Who knows what problem she has?" Just as for the other "soup" video clip (showing the male actor), however, no one inferred that she was hungry.

Set B (Expressive Faces)

First-time viewers. Even with these emotionally congruent stimuli, first-time viewers rarely constructed any links between the camera shots. Only for the graveyard sequence was there any evidence of any interaction. Critically, however, this did not constitute a full spatiotemporal association. Rather, participants said that they thought that the man was sorry for his loss, but did not seem to perceive him to be spatially located in the graveyard. When asked where he was, they did not say that he was across from or next to gravestone. They said that he was "here," looking at them.

When they were further probed regarding where the gravestone was, participants responded that "it was gone." In the other sequences, even this limited interaction was not observed. For the soup sequence, for example, participants described the man to be licking his lips/gulping (0 percent said he looked hungry) and then said that the plate (or pot/well/hole/pool)

appeared "*again*." When asked the reason for this man's behavior, they said that they "cannot know" it. For the child sequence, the two shots were also interpreted as two independent pictures. The little girl and the man were said to be looking happy, but no participants commented that they were together.

Experienced viewers. Descriptions of the soup, graveyard, playing child, and old woman (with directed gaze) sequences all indicated that 100 percent of the experienced viewers made clear spatiotemporal associations between these shots. Furthermore, most of these participants perceived the emotions of the persons in the predicted manner, describing the man as hungry in the soup condition (95 percent for Actor A and 100 percent for Actor B), sad in the gravestone condition (100 percent for both actors), and happy in the child condition (100 percent for both actors).

Table 3: Perception of Spatiotemporal Continuity and the Kuleshov Effect across Groups for the Second Set of Film Sequences (with Emotionally Congruent Facial Expressions)

<Table 3 around here>

Discussion

The Kuleshov effect occurs when an observer makes a conscious connection between—and subsequently mentally interacts with—edited camera shots. The camera shots used in the sequences typically associated with this effect are not connected to each other with commonalities on a perceptual level, but are rather linked through intentions, motivations, and emotions. In other words, any continuity between juxtaposed shots is an illusion created in the mind of a viewer, and the landscape in which both shots are located in is an artificial one existing

outside of reality. The present study investigated whether first-time viewers construct spatiotemporal relations between the shots like experienced viewers do (i.e., forging a narrative connection between them and conceiving of the artificial landscape created in the video clips). Here, we coded first-time and experienced viewers' responses to classic Kuleshov experiment sequences in order to establish whether or not there are differences in how first-time film viewers spontaneously connect edited shots and generate the Kuleshov effect.

The current study did not address different theories of emotion, the existence or discreteness of specific emotions, or other related topics.[4] A person's ability to make sense of facial expressions is affected by several factors, which we attempted to control for as much as possible in the experiment. Responses from the experienced viewer participant group validated our chosen stimulus set. These participants all connected the shots on the spatiotemporal level and had no problem identifying the facial expressions used in the Set B sequences. Moreover, even the first-time viewers accurately categorized the emotions in the happiness and sadness conditions. They did not do so for the emotions in the hunger condition, which could reflect the fact that hunger is not one of the "basic" or universal emotions.

Our results reveal that the first-time viewers did not demonstrate either of the two key components of the Kuleshov effect. Despite an intact ability to perceive and understand the content of each shot, they perceived them to be wholly separate from each other and did not relate them spatiotemporally. Even when the coherent facial expressions were juxtaposed with the causes of such expressions, they still considered them as if they were independent photographs: a visual format they are familiar with. [Callout 6 about here]

The first-time viewers did not seem to have the notion of what constitutes a film (i.e., sequences of shots that are linked to one another in coherent ways). In the sadness condition, for

example, they said that the person was sad because of someone he had lost (in relation to the gravestone shot), but crucially there was no indication that they thought that the sad person and the gravestone were in the same place at the same time. The image of the person was not "here" anymore as the image of gravestone. These results are consistent with the results of a study that looked at young children viewing picture books (Berman 1988); it suggested that, as far as these children were concerned, once a page is turned a new story begins. Ruth Berman (1988) concluded that the narrative abilities that function to allow children to link events are constrained by broader cognitive development issues, expressive language abilities, and children's (un)familiarity with the narrative norms of their literate society.

Somewhat surprisingly, the customized additional video clip added during testing revealed that first-time viewers *can* connect edited sequences spatiotemporally under at least some conditions. For example, this can happen when a person's gaze in the first shot is coherent with the location of the depicted object in the second shot. Here, the eyeline matching, which is the filmic equivalent of joint attention (something acquired in early childhood, e.g., Moore and Dunham 2014), may have provided an instance of a conceptual relation that was clear enough for even naïve viewers to interpret. Eyeline matches, in other words, appeared to open the eyes of first-time viewers to the artificial landscape created in the video clip. Unfortunately, there was no scope for interpretation of the facial expression of the lady depicted, because her face was not readable (head and eyes were turned downward), thus preventing evaluation of the second component of the Kuleshov effect.

The "classic" Kuleshov effect was clearly observed for experienced viewers only in the sadness condition. Here, participants reported that the man standing in front of the gravestone was sad for his loss, although the footage showed the same neutral face that was juxtaposed with

the shots of the soup and the little girl. It could be argued that the image of the gravestone is much more intense and salient than the images of a bowl of soup and a cute child playing. This study, however, followed the procedures described in other studies of the Kuleshov effect, so as to be comparable with this previous research. It is possible that these participants' interpretations of the emotional state of the faces shown before the bowl of soup could also be considered evidence of the Kuleshov effect in action. Although there was no clear attribution of a specific emotional or mental state, the experienced viewers tried to find an explanation for what caused the man to not eat the soup in front of him. Thirty percent of them said that the man was unsure as to whether he should eat it, and forty-five percent thought that he was waiting for someone else. [Callout 7 about here]

When considering participants' responses to the video sequences with the little girl, it may be helpful to consider that viewing one facial expression can shift the wider scale of judgment. That is, a strongly salient "anchor" face can skew the emotion perceived in subsequent faces in the opposite affective direction (Russell and Fehr 1987), making a neutral face appear sad when presented after a happy face, or happy when presented after a sad face. Thus, the happy face of the little girl in the test sequences might have biased participants' interpretations of the actor's facial expression.

Prince and Hensley (1992) cited the naïveté of the early audiences as a possible reason for discrepancies in the appearance of the Kuleshov effect with contemporary audiences. Our results challenge this notion. They indicate that first-time film viewers do not even link intercut camera shots edited in sequence, let alone demonstrate the Kuleshov effect. We propose, instead, that experienced viewers are more likely to "collaborate" with the filmmaker. That is, they are more likely to try to understand the filmmaker's intentions and make sense of what they see

because they know that films comprise shots that come together to convey a narrative. Such viewers stand in stark contrast with naïve viewers, who seem to be unaware of the existence of a filmmaker or a camera. It should be noted here that the experienced viewers in the present study (like the first-time viewers) had no prior experience of taking part in research experiments. Both participant groups were first-time participants in a study and had no idea what a study was. Even though the experiment was explained to them, they supposed that they would simply be watching videos, not realizing that they were intentionally made for research purposes. [Callout 8 about here]

It also seems worth mentioning here that the first-time viewers (mis)interpreted the objects shown in the close-up shots as things bigger than they really were (e.g., bowl, hole) and the people as sitting (only upper bodies were shown) in the medium shots. This is evidence that the first-time viewers recruited for this particular study had only a very basic understanding of what film was. It was also interesting that neither the first-time nor the experienced viewers made any comment on the black-and-white quality of the video clips. Further research is needed to determine the role of such prior knowledge by explaining the concept of film to first-time viewers. Further research is also needed to test the Kuleshov effect with other images (e.g., those as perceptually salient as a gravestone), which might elicit stronger emotions and modulate perception more powerfully. Direction of gaze and the order of the shots have also been identified as key variables that should also be taken into account in such future research.

Acknowledgments

Sermin Ildirar is a Marie S. Curie postdoctoral research fellow at the Centre for Brain and Cognitive Development, Birkbeck, University of London. She studied film and media studies at Istanbul University and the University of Vienna, and she has worked at the Knowledge Media Research Center in Tübingen, Germany. She is currently investigating the role of cinematic features on the cognitive processes of both adult and infant viewers. She has articles published in *Psychological Science* and the *British Journal of Psychology*. She is also the director of several short movies and co-scriptwriter of a feature movie and a computer game. E-mail: ildirarsermin@gmail.com

Louise Ewing is a lecturer at the University of East Anglia, Norwich. She studied psychology at the University of Western Australia. Her research focuses on face and person perception. She has articles published in *Perception*, *Vision Research*, and the *Journal of Experimental Psychology*. E-mail: l.ewing@uea.ac.uk

References

Anderson, Joseph D. 1998. *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*. Carbondale: Southern Illinois University Press.

Aviezer, Hillel, Ran R. Hassin, Jennifer Ryan, Cheryl Grady, Josh Susskind, Adam Anderson, Morris Moscovitch, and Shlomo Bentin. 2008. "Angry, Disgusted, or Afraid? Studies on the Malleability of Emotion Perception." *Psychological Science* 19 (7): 724–732.

Barratt, Daniel, Anna Cabak Rédei, Åse Innes-Ker, and Joost Van de Weijer. 2016. "Does the Kuleshov Effect Really Exist? Revisiting a Classic Film Experiment on Facial

Expressions and Emotional Contexts." *Perception* 45 (8): 847–874.

Barrett, Lisa Feldman, Batja Mesquita, and Maria Gendron. 2011. "Context in Emotion Perception." *Current Directions in Psychological Science* 20 (5): 286–290.

Berman, Ruth A. 1988. "On the Ability to Relate Events in Narrative." *Discourse Processes* 11 (4): 469–497.

Bill, Valentine T. 1987. *Chekhov: The Silent Voice of Freedom*. New York: Allied Books.

Bordwell, David, Janet Staiger, and Kristin Thompson. 1985. *The Classical Hollywood Cinema: Flm Style and Mode of Production to 1960*. London: Routledge.

Bordwell, David, Kristin Thompson, and Jeremy Ashton. 2004. *Film Art: An Introduction*. 7th ed. New York: McGraw-Hill.

Butterworth, George, and Nicholas Jarrett. 1991. "What Minds Have in Common Is Space: Spatial Mechanisms Serving Joint Visual Attention in Infancy." *British Journal of Developmental Psychology* 9 (1): 55–72.

Carroll, James M., and James A. Russell. 1996. "Do Facial Expressions Signal Specific Emotions? Judging Emotion from the Face in Context." *Journal of Personality and Social Psychology* 70 (2): 205–218.

Carroll, Noël. 1993. "Toward a Theory of Point-of-View Editing: Communication, Emotion, and the Movies." *Poetics Today* 14 (1): 123–141.

Corkum, Valerie, and Chris Moore. 1998. "The Origins of Joint Visual Attention in Infants." *Developmental Psychology* 34 (1): 28–38.

Cutting, James E. 2005. "Perceiving Scenes in Film and in the World." In *Moving Image Theory: Ecological Considerations*, ed. Joseph D. Anderson and Barbara Fisher Anderson, 9–27. Carbondale: Southern Illinois University Press.

D'Entremont, Barbara, Sylvia M. J. Hains, and Darwin W. Muir. 1997. "A Demonstration of

    Gaze Following in 3- to 6-Month-Olds." *Infant Behavior and Development* 20 (4): 569–

    572.

Forsdale, Joan Rosengren, and Louis Forsdale. 1966. "Film Literacy." *Journal of the University*

    Film Producers Association 18 (3): 9–27.

de Gelder, Beatrice, Hanneke K. M. Meeren, Ruthger Righart, Jan Van den Stock, Wim A. C.

    Van de Riet, and Marco Tamietto. 2006. "Beyond the Face: Exploring Rapid Influences

    of Context on Face Processing." *Progress in Brain Research* 155: 37–48.

Gibson, James J. 2014. *The Ecological Approach to Visual Perception: Classic Edition*. New

    York: Psychology Press.

Goldberg, Herman D. 1951. "The Role of 'Cutting' in the Perception of the Motion

    Picture." *Journal of Applied Psychology* 35 (1): 70–71.

Hemsley, Gordon D., and Anthony N. Doob.1978. "The Effect of Looking Behavior on

    Perceptions of a Communicator's Credibility." *Journal of Applied Social*

    Psychology 8 (2): 136–142.

Hobbs, Renee, Richard Frost, Arthur Davis, and John Stauffer. 1988. "How First-Time Viewers

    Comprehend Editing Conventions." *Journal of Communication* 38 (4): 50–60.

Holland, Norman N. 1989. "Film Response from Eye to I: The Kuleshov Experiment." *South*

    Atlantic Quarterly 88 (2): 415–442.

Ildirar, Sermin, Daniel T. Levin, Stephan Schwan, and Tim J. Smith. "Audio Facilitates the

    Perception of Cinematic Continuity by First-Time Viewers." *Perception* (2017):

    0301006617745782.

Ildirar, Sermin, and Stephan Schwan. 2015. "First-Time Viewers' Comprehension of Films:

Bridging Shot Transitions." *British Journal of Psychology* 106 (1): 133–151.

Isenhour, John Preston. 1975. "The Effects of Context and Order in Film Editing." *AV Communication Review* 23 (1): 69–80.

Kuleshov, Lev Vladimirovich. 1974. *Kuleshov on Film: Writings*. Berkeley: University of California Press.

Levaco, Ronald. 1974. "Eikhenbaum, Inner Speech and Film Stylistics." *Screen* 15 (4): 47–58.

Lindgren, Ernest. 1948. *The Art of the Film*. London: George Allen & Unwin.

Lobmaier, Janek S., Martin H. Fischer, and Adrian Schwaninger. 2006. "Objects Capture Perceived Gaze Direction." *Experimental Psychology* 53 (2): 117–122.

Masuda, Takahiko, Phoebe C. Ellsworth, Batja Mesquita, Janxin Leu, Shigehito Tanida, and Ellen Van de Veerdonk. 2008. "Placing the Face in Context: Cultural Differences in the Perception of Facial Emotion." *Journal of Personality and Social Psychology* 94 (3): 365–381.

Messaris, Paul. 1994. *Visual "Literacy": Image, Mind, and Reality*. Boulder, CO: Westview Press.

Mobbs, Dean, Nikolaus Weiskopf, Hakwan C. Lau, Eric Featherstone, Ray J. Dolan, and Chris D. Frith. 2006. "The Kuleshov Effect: The Influence of Contextual Framing on Emotional Attributions." *Social Cognitive and Affective Neuroscience* 1 (2): 95–106.

Moore, Chris, and Phil Dunham. 2014. *Joint Attention: Its Origins and Role in Development*. New York: Psychology Press.

Moors, Agnes, Phoebe C. Ellsworth, Klaus R. Scherer, and Nico H. Frijda. 2013. "Appraisal Theories of Emotion: State of the Art and Future Development." *Emotion Review* 5 (2): 119–124.

Münsterberg, Hugo. 2013. *Hugo Munsterberg on Film: The Photoplay—A Psychological Study and Other Writings*. London: Routledge.

Nelson, Nicole L., and James A. Russell. 2013. "Universality Revisited." *Emotion Review* 5 (1): 8–15.

Pearson, Roberta, and Philip Simpson, eds. 2005. *Critical Dictionary of Film and Television Theory*. London: Routledge.

Perrett, David Ian, Jari K. Hietanen, Michael W. Oram, Philip J. Benson, and E. T. Rolls. 1992. "Organization and Functions of Cells Responsive to Faces in the Temporal Cortex [and Discussion]." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 335 (1273): 23–30.

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

Prince, Stephen, and Wayne E. Hensley. 1992. "The Kuleshov Effect: Recreating the Classic Experiment." *Cinema Journal* 31 (2): 59–75.

Pudovkin, Vsevolod Illarionovich. 2013. *Film Technique and Film Acting: The Cinema Writings of V. I. Pudovkin*. Vancouver: Read Books.

Renov, Michael. 2004. *The Subject of Documentary*. Minneapolis: University of Minnesota Press.

Righart, Ruthger, and Beatrice de Gelder. 2008. "Recognition of Facial Expressions Is Influenced by Emotional Scene Gist." *Cognitive, Affective, and Behavioral Neuroscience* 8 (3): 264–272.

Rosenheim, Shawn. 1996. "Interrotroning History: Errol Morris and the Documentary of the Future." In *The Persistence of History: Cinema, Television, and the Modern Event*, ed.

Vivian Sobchack, 219–234. London: Routledge.

Russell, James A., and Beverley Fehr. 1987. "Relativity in the Perception of Emotion in Facial Expressions." *Journal of Experimental Psychology* 116 (3): 223–237.

Schwan, Stephan, and Ildirar, Sermin. 2010. "Watching Film for the First Time: How Adult Viewers Interpret Perceptual Discontinuities in Film." *Psychological Science* 21 (7): 970–976.

Sharff, Stefan. 1997. *The Art of Looking in Hitchcock's Rear Window*. Winona, MN: Hal Leonard.

Smith, Tim J. 2012. "The Attentional Theory of Cinematic Continuity." *Projections* 6 (1): 1–27.

Smith, Tim J., Daniel Levin, and James E. Cutting. 2012. "A Window on Reality Perceiving Edited Moving Images." *Current Directions in Psychological Science* 21 (2): 107–113.

Tomasello, Michael, Ann Cale Kruger, and Hilary Horn Ratner. 1993. "Cultural Learning." *Behavioral and Brain Sciences* 16 (3): 495–511.

Truffaut, François. 1984. *Hitchcock*. Rev. ed. with the collaboration of Helen G. Scott. New York: Simon & Schuster.

Vrij, Aldert, Samantha Mann, Sharon Leal, and Ronald Fisher. 2010. "'Look into My Eyes': Can an Instruction to Maintain Eye Contact Facilitate Lie Detection?" *Psychology, Crime and Law* 16 (4): 327–348.

CALLOUTS

Callout 1: The original footage used by Kuleshov is long-since lost and superficial issues with the design of the experiments have prompted some to re-classify it as part of the "mythology of film" (Holland 1989) or "folklore of the cinema" (Pearson and Simpson 2005).

Callout 2: Prince and Hensley (1992) suggested that the "naïveté of early cinema audiences," compared with their more experienced, modern participants, might explain the original findings.

Callout 3: There are two components to the Kuleshov effect: perception of spatiotemporal continuity between the juxtaposed camera shots and perception of a change in emotion of the target (neutral) face.

Callout 4: A typical response was that there was a man looking towards the viewer sitting in silence that came and went. . . . When they were asked where he was, he was not reported to be across or next to gravestone but rather "here," looking at us.

Callout 5: The first-time viewers (mis)interpreted the objects shown in close-up shots as things bigger than they really were (e.g. plate, hole) and the people as sitting (only upper bodies were shown, in medium shots).

Callout 6: The first-time viewers do not demonstrate either of the two key components of the Kuleshov effect. . . . They do not seem to have the notion of what constitutes a film, i.e., sequences of shots that are linked in coherent ways.

Callout 7: The 'classic' Kuleshov effect was clearly observed for experienced viewers only in sadness condition.

Callout 8: We propose, instead, that it is experienced viewers that are more likely to 'collaborate' with the filmmaker.

Notes

---

[1] It has been (conflictingly) reported that Kuleshov found a long strip of film with Mozhukhin's close-up and used it for his experiment (Levaco 1974) and that he purposely filmed Mozhukhin

after having instructed him to appear expressionless (Messaris 1994).

[2] For further discussions about the perception of continuity in films, see Ildirar and Schwan (2015), and Smith (2012), Smith et al. (2012).

[3] For a summary of Persson's theory, see Barratt et al. (2016).

[4] For cutting-edge theories of emotion, see Moors et al. (2013), and for current debates on the universality of emotion recognition, see Nelson and Russell (2013).