

A Robust Neural Fingerprint of Cinematic Shot-Scale

Citation for published version (APA):

Raz, G., Valente, G., Svanera, M., Benini, S., & Kovács, A. B. (2019). A Robust Neural Fingerprint of Cinematic Shot-Scale. *Projections-The journal for movies and mind*, 13(3), 23-52.
<https://doi.org/10.3167/proj.2019.130303>

Document status and date:

Published: 01/12/2019

DOI:

[10.3167/proj.2019.130303](https://doi.org/10.3167/proj.2019.130303)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:


www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



A Robust Neural Fingerprint of Cinematic Shot-Scale

Gal Raz, Giancarlo Valente, Michele Svanera,
Sergio Benini, and András Bálint Kovács

Abstract: This article provides evidence for the existence of a robust “brainprint” of cinematic shot-scales that generalizes across movies, genres, and viewers. We applied a machine-learning method on a dataset of 234 fMRI scans taken during the viewing of a movie excerpt. Based on a manual annotation of shot-scales in five movies, we generated a computational model that predicts time series of this feature. The model was then applied on fMRI data obtained from new participants who either watched excerpts from the movies or clips from new movies. The predicted shot-scale time series that were based on our model significantly correlated with the original annotation in all nine cases. The spatial structure of the model indicates that the empirical experience of cinematic close-ups correlates with the activation of the ventral visual stream, the centromedial amygdala, and components of the mentalization network, while the experience of long shots correlates with the activation of the dorsal visual pathway and the parahippocampus. The shot-scale brainprint is also in line with the notion that this feature is informed among other factors by perceived apparent distance. Based on related theoretical and empirical findings we suggest that the empirical experience of close and far shots implicates different mental models: concrete and contextualized perception dominated by recognition and visual and semantic memory on the one hand, and action-related processing supporting orientation and movement monitoring on the other

Keywords: apparent distance, fMRI, machine learning, motion pictures, neural decoding, shot-scale

Art historians of the early twentieth century already attributed a great importance to the fact that painters represent their subject matters from different apparent distances. The Austrian art historian Alois Riegl (1858–1905) was the first to propose that objects and scenes depicted in a painting from various apparent distances represent different forms of perception. He named these forms *fernsehen* (“distant vision”), *normale sehen* (“normal vision”), and

nahsehen (“close vision”) (Riegel 1901). He supposed that the mere perception of distance or proximity has direct psychological effects, and he attached different sensorial values to “distant vision” and to “close vision.” He proposed that distant vision is “optical,” meaning that the relationship between the beholder and the scene that he is watching is purely optical and therefore evokes a sense of distance. By contrast, he called *nahsehen* “haptic,” meaning that the proximity of the object represented in the image evokes tactile sensations in the beholder. Heinrich Wölfflin (1864–1945), a Swiss art historian, agreed with Riegl that different “shot-scales” exist in painting and that they represent different perceptual modes, but he attributed slightly different psychological effects to different distances. Wölfflin (1921) suggested that different techniques of painting involve distinct distance feelings. The style that he called “painterly” needed a more distanced way of seeing than scenes depicted in a “linear” style. The former is represented with little emphasis on the details and contours, which renders a general impression of the gist of the scene, while the latter is depicted in such a way as to emphasize contours and minor details, which elicits a feeling of closeness.

If these art historians are right, then shot-scales experienced in cinema must have similar psychological effects to those that are experienced with

If these art historians are right, then shot-scales experienced in cinema must have similar psychological effects to those that are experienced with paintings. And so studying the perception of cinematic shot-scales can also help to verify the value of these art historical insights.

paintings. And so studying the perception of cinematic shot-scales can also help to verify the value of these art historical insights. Furthermore, we think that the reason why the use of close-ups and other shot-scales became the universal standard for controlling viewers’ attention after the 1910s—and why it entirely supplanted similar methods such as masking after the 1930s—is due to the perceptual properties of distance perception and the resulting psychological effects thereof.

Hence, the question leading our research is this: is there any evidence that different shot-scales are consistently associated with distinct reaction patterns in the perceiver’s brain? If there is, what could the psychological meaning of these patterns be?

Distance-Perception Mechanisms

Before reviewing recent work in experimental psychology that meaningfully integrates with the aforementioned art historians’ insights, let us briefly describe several key principles of distance perception. This perceptual faculty relies on a variety of neural processes that provide redundant size and depth cues. Major sources for relevant information are the triad of extraretinal mechanisms controlling eye vergence (simultaneous and matched movement

of the eyes toward or away from each other when fixating on proximate or remote objects, respectively), lens curvature accommodation (eye muscles increase the curvature of the elastic lens when perceiving nearby objects and flatten it for distant vision), and pupil diameter (the pupil constricts and dilates when fixating on proximate or remote objects, respectively; see Loewenfeld and Lowenstein 1993; McDougal and Gamlin 2015; and Sperandio and Chouinard 2015). These synchronized processes, which are controlled by reflex mechanisms, are necessary for clear single binocular vision. Since they are fundamentally affected by factors related to viewing distance, they probe this parameter and make it available to the brain.

While this triad of reflexes plays a key role in real-world distance perception in movies that are projected onto a flat screen at invariant an viewing distance, accommodation distance is constant but vergence varies (vergence is a reliable marker in stereoscopic three-dimensional and two-dimensional displays; see Knight et al. 2012). In addition, distance estimation relies on a set of bottom-up and top-down cues including aerial perspective, familiar size, shading, texture gradient, occlusion, and relative height (Sperandio and Chouinard 2015). Naturally, under standard cinematic conditions the distance between the spectator and the two-dimensional display is kept constant so that changes in the viewing distance are necessarily illusionary. Nevertheless, similarly to natural three-dimensional vision (but probably to a smaller extent; see Sperandio et al. 2012), apparent distance still affects size perception in two-dimensional display.

This effect is evident in the well-studied geometrical-optical Ponzo illusion. In this illusion, which has been validated in numerous studies (for review, see Changizi et al. 2008), the apparent size of the same two-dimensional object is changed as a function of its location relative to a pair of converging lines (see Figure 1). Thus, even in the absence of varying accommodation cues, the visual system is “tricked” by two-dimensional perspective distance cues.

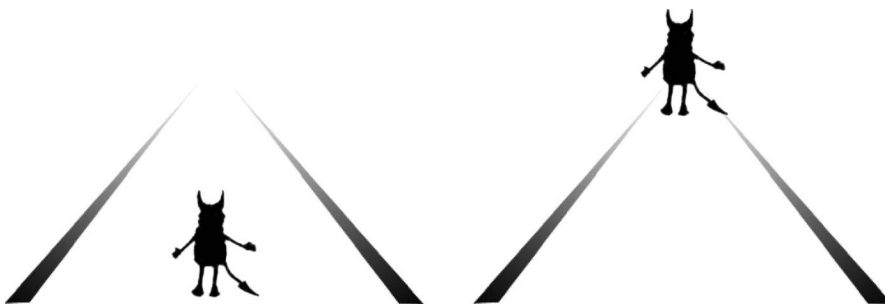


Figure 1. A version of the Ponzo illusion. Although the size of the two creature icons is identical, the left icon will usually be perceived as smaller than the right icon due to its different location relative to the converging lines.

The effects of distance cues on visual perception in two dimensions have been documented in neuroimaging studies as well. For example, in the context of the Ponzo illusion, an image of the same object (a ring) induced more or less distributed activity patterns in the primary visual cortex V1 depending on depth cues (He et al. 2015). Elinor Amit and colleagues (2012) further explored the effects of the Ponzo illusion in brain regions that show selectivity to images of either objects or open scenes. In particular, the researchers tested the effect of two-dimensional distance cues (as in Figure 1) on the activity levels in the following regions of interest: the lateral occipital (LO) object area and the posterior fusiform gyrus (pFg), which assumingly support high-acuity visual processing and which are dominated by stimuli located at the center of the field of view (Larsson and Heeger 2006; Roberts et al. 2013; Woodhead et al. 2011); and the parahippocampal place area (PPA) and transverse occipital sulcus (TOS), which are biased toward noncentral (peripheral) information with low spatial frequency (i.e., coarse details) (Arcaro et al. 2009; Press et al. 2001; Sowards 2011). They found that the two scene-selective regions were more active when the two-dimensional cues created the illusion that the stimuli were distal, whereas the LO area was activated by proximity cues (the pFg showed a similar proximity bias for objects, but not for buildings). In other words, the *same image*, depending on distance construal that is based on two-dimensional background cues, induced distinct (and in some sense, opposite) brain activity patterns.

Psychological Aspects of Perceived Distance

This evidence for a link between specific apparent distances and patterns of brain activity suggests that in line with Riegl's and Wölfflin's theories, distinct perceived distances may be associated with different mental states. Such an association is endorsed by *construal level theory* (Liberman and Trope 2008). According to this theory, construal made at the *distal level* (in terms of temporal, spatial, and social distance between an individual and an object) tends to be more abstract, schematic, and decontextualized. Such construal extracts the gist of the situation and plays down the details. On the other hand, construal made at the *proximal level* is "concrete, relatively unstructured, and contextualized": it is a representations that includes "subordinate and incidental features of the events" (Liberman and Trope 2008, 1201).

This theory relies on empirical research, which links different domains of psychological distance. Thus, for instance, it was reported that individuals process congruent combinations of spatial and temporal distance (e.g., "here" and "today" or "there" and "last year") more quickly than incongruent information (e.g., "there" and "today"). Similarly, spatial distance was found to be implicitly associated with social distance (e.g., "friend" and "enemy"; see Figure 2) and sensitivity to concrete details (distant construal is linked with increased at-



Figure 2. Examples of the type of stimuli presented in Bar-Anan and colleagues' (2007) research into the link between different aspects of psychological distance. [Reproduction]. In the congruent condition (left), the label implies psychological distance that matches the apparently proximal or distal location to which the arrow is pointing. In the incongruent condition (right), the location of the arrow and the verbal element do not match. Congruent conditions were identified more quickly.

tention to general trends, whereas in proximal construal deviations from the trend are outweighed; see Henderson et al. 2006). Even small spatial details may significantly bias one's view of the world. In a study, which compared the interpretation of story versions that varied only in the reported location of described events (distant or proximal), the participants tended to interpret these events either as representing an enduring and general disposition or as limited to the concrete context, respectively (Henderson et al. 2006).

Brain Correlates of Cinematic Shot-Scale

The aforementioned empirical findings and construal level theory bear significant implications for film theory in general and for the theorization of shot-scale in particular. If different perceived spatial distances implicate qualitatively distinct mental dispositions, the properties of these dispositions should be taken into account when considering apparent cinematic distance. The very cinematic decision to frame given content in close-up or in long-shot may significantly affect the cinematic experience. A recent line of research in film studies empirically examined related topics including shifts in shot-scale proportions across film history (J. E. Cutting and Candan 2015; J. E. Cutting et al.

2010; J. Cutting and Iricinschi 2015; Benini et al. 2016) and genres (Kovács and Zentay 2017), the link between shot-scale and the number of characters (J. E. Cutting 2015) or the duration of the shot (J. E. Cutting and Armstrong 2016), shot-scale patterns as a formal characteristic of the works of specific filmmakers and their dynamics (Kovács 2014), and the effect between shot-scales on the viewer's theory of mind (Bálint et al. 2018; Rooney and Bálint 2018).

A relatively major and open issue in this context is the discussion about the inherent emotionality of close-ups. According to one approach, the close-up gains its emotional charge from the emotional facial expression that it represents (Palmer 1920; Schrader 2014). In this case, the close-up's function is to focus the viewer's attention to the facial expression, which has less to do with the proximity of the face than with its enlargement and dominance in the representation (Carroll 1996; Palmer 1920). According to this approach, the viewer sees an enlarged emotional expression and the emotional effect results either from the viewer's mirroring this enlarged expression or from the enlargement only. A recent work (J. E. Cutting and Armstrong 2016) provides empirical support for the notion that viewers are indeed quicker and more accurate in recognizing the emotional valence of facial expressions in close shots than they are in distant shots. The alternative approach regarding the close-up emphasizes the expressive value of the close-up as a result of the perception of proximity regardless of the object represented in the image: "But the magnifying glass of the cinematograph brings us closer to the individual cells of life" (Balázs 1924, 38). According to Béla Balázs, any object represented in a close-up becomes expressive by the mere effect of its closeness. This is how Balázs describes the emotional effect of the close-up: "We often gain the impression that these shots are the product not so much of a good eye as of a good heart. They radiate warmth, a lyricism . . . a tender feeling towards things is aroused without being made explicit" (1924, 39). It is noteworthy that Balázs talks about "things" not just faces. And, according to Jean Epstein, close-up "modifies the drama by the impact of proximity" (1921, 104). In this account, proximity is the primary effect of the close-up independently of the high-level identification of the object represented in the image.

Our present concern is not specifically with the explanation of the emotional effect of close-ups or other shot-scales, but with the broader question about the link between shot-scale and distance perception as evidenced by brain activity patterns. It is through this link that emotional and other psycho-

Do close and long shots implicate distinct patterns of neural activity and mental dispositions?

logical effects of shot-scales can be interpreted. The departure point is the aforementioned empirical evidence, which indicates that different apparent distances are associated with (a) distinct construal levels; and (b) distinct brain activation profiles. What are the implications of these insights for the notion of cinematic shot-scale? Do close and long shots implicate distinct patterns of neural activity and mental dispositions?

“Brainprint”: Image Content versus Apparent Distance

The aforementioned study by Amit and colleagues (2012) suggests that brain activity is modulated by two-dimensional depth cues that are also predominant in the standard context of the moving image. However, in actual cinematic shots the framing scale and the image content may significantly interact with one another. For example, a specific affective mental state may be elicited by close-ups, but not by long shots, only at the presence of social information (e.g., during sequences that depict interpersonal interaction as suggested by Palmer 1920). On the other hand, long shots, but not close-ups, may support a mental mode of seeking, depending on specific narrative cues for spatial expectations. In other words, Amit and colleagues’ (2012) study convincingly shows the coactivation of specific brain areas for apparent distance shifts, but it does not exclude that other effects (e.g., image content) may also have a considerable impact on the very same areas.

Another related neuroimaging finding is the result of Uri Hasson and colleagues’ pioneering study of the cross-viewer similarity of local neural patterns of reactions to movies (Hasson et al. 2004). To assess the functional significance of their finding, these authors adopted a “reverse correlation” approach: they selected a portion of the highest peaks of the intersubjectively correlated signal in predefined brain regions and examined the visual content during cinematic moments that corresponded with these peaks. They found images of faces in sixteen out of sixteen peaks of the fusiform face area (FFA) signal and images of open scenes in twelve out of sixteen peaks of the PPA signal. Allegedly, based on this result, the FFA and PPA can be seen as brain structures that preferably process close and far shots with close and distant shots, respectively. However, Hasson and colleagues did not establish a direct link between these brain regions and shot-scales, as they did not systematically address the question of shot-scale brain correlates. In fact, the face images that were shown by the authors to correspond with FFA signal peaks varied in shot-scale from medium shot to extreme close-up (Hasson et al. 2004, 1637). Moreover, while Hasson and colleagues quantified the ratio of true-to-false positives (i.e., the ratio between shots containing and not containing face/open-scene images during the signal peaks, which refers to a type I error), they did not account for true-to-false negatives (i.e., the ratio between shots containing and not containing these elements but without any correspondence with PPA and FFA signal peaks, which refers to a type II error). Such examination depends on an extended annotation of shot-scales across the entire movie, which is the empirical basis of the study we are presenting in this article.

Our study directly addresses the issue of cinematic shot-scale. It should be noted that shot-scale is a complex construct, which is not limited to apparent distance. Other processes may covary with shot-scale due to either a natural association or a contingency based on cinematic conventions. For example,

optical flow may naturally be more pronounced in close-ups due to the higher impact of small movements on the flow at this scale. Therefore, optical flow may be a component of the “brainprint” of these close-ups, while it is not necessarily linked with apparent distance. Thus, any empirical study that aims to boil shot-scale down to apparent distance should control for a large (and yet unspecified) array of parameters or test the generalizability of shot-scale-specific patterns across these parameters. Since we are aware that in the brain’s response to shot-scales the effect of image content may largely interact with apparent distance perception, we investigate shot-scale as an empirical phenomenon. Rather than decontextualizing shot-scale, we examine whether this feature as implemented in mainstream cinema consistently correlates with specific brain activity patterns across different movies. It is through the variety of genres, styles, and visual content that we think the common perceptual features of different shot-scales indicated by their eventual neural fingerprints can be isolated. We posit that, if a consistent link exists between a certain shot-scale and a specific brain activity pattern to the extent that it allows for a reliable decoding of this cinematic feature across a sample of commercial movies, genres, and viewers (even with the use of different MRI scanners), then the pattern should not be considered as representing a mere *confound* but rather as a neural *correlate* of the shot-scale with its significant associated processes.

Our main hypothesis is that shot-scale reliably covaries with distinctive patterns of brain activity across movies, genres, and viewers.

To recap, our main hypothesis is that shot-scale reliably covaries with distinctive patterns of brain activity across movies, genres, and viewers. We examine whether, despite its heterogeneous nature in terms of content, shot-scale has a reliable “brainprint” on which we can base a “mind reading” of this feature. Because of the variability of the image content and cinematic context in the sample, we further hypothesize that the LO object area and the pFg, on the one hand, and the PPA and TOS, on the other, which were selectively implicated with the processing of short and long apparent distance, respectively (Amit et al. 2012), will be associated with close and far shots, respectively, indicating the distinctive impact of apparent distance perception. With respect to the aforementioned debate on the relations between close-up and face presence (see also Deleuze 1986 for a more abstract analogy between these phenomena), we conducted a secondary empirical analysis. We first tested the validity of the intuitive link between shot-scale categories and the prevalence of face images in our data. Next, we tested whether a shot-scale model would predict shot-scale annotation better than a face-ratio model would.

Methods

Our movie sample data included clips from a set of commercial movies representing genre diversity: drama, comedy, war film, thriller, horror, documentary, and nature film. We adopted a stringent validation approach to the problem of neural correlates of shot-scales (Figure 3): first, fMRI data was collected from

healthy volunteers while they were watching at least one of these clips. Second, we applied a machine-learning approach on this data and developed a computational model that estimates shot-scale given any fMRI film viewing data. And third, the validity of the resulting model was tested by comparing the time course of manually annotated shot-scales for new movies (unseen by the algorithm) with the model's predictions, which were derived from data collected from different participants (except for one case in which one of the movies was shown to different participants in the training and testing groups). We note here that it was rather difficult to obtain reliable decoding that generalized across the training and the testing data. Our previous attempts to validate similar models for the mean brightness level (Raz et al. 2017) and for cut frequency were not successful.

Our decoding procedure relied on a method that was implemented in a recent study (Raz et al. 2017) and that is described in the Appendix (see Figure 3). In short, in this procedure we used a method called “generalized cross-validation with kernel ridge regression” (GCV-KRR) to produce a three-dimensional model, which can be applied to given fMRI data to decode shot-scale. The generation of this model is a statistical procedure in which a certain *weight* is assigned to each voxel (three-dimensional pixel) in the brain. At any given time-point in the movie, the fMRI signal values in each of the voxels is multiplied by the weight assigned to these voxels in accordance with the model. The predicted shot-scale (on a 1–7 scale) is the sum of all of the voxel-wise multiplications between the value of the (preprocessed) fMRI signal at this time-point and the weight. Generally speaking, the weights are computed with the aim that the difference between the predicted and the actual shot-scale will be minimal (for more details, see the Appendix). We used a large dataset of 234 fMRI scans that took place during the viewing of these excerpts (see Table 1). This dataset was used in previous studies (Raz et al. 2014; Raz et al. 2012; Raz, Shpigelman, et al. 2016; Raz, Touroutoglou, et al. 2016). Once the brain model for shot-scales was generated, we tested it using the fMRI data of an additional fifty-five scans, which were obtained from other participants while viewing other movies, and ten scans in which one of the training movies was presented to other participants. For this test, we reconstructed a time series of the predicted shot-scales for these movies based on the fMRI data and compared them with the actual shot-scale series as derived from the manual annotation.

We posit that if a standard intersubjective model of shot-scale indeed produces significant predictions of this cinematic feature under a variety of different cinematic conditions, then the distinct brain activity patterns on which this prediction is based are reliably associated with shot-scale. In other words, such a “mind reading” of shot-scale can succeed only if this feature has a reliable “brainprint.”

Table 1. Details on materials used in the study and the samples. The sample size indicates the net number of datasets after dropout due to technical reasons and head motions.

Movie data								
Movie details					Sample details			
Training set								
Film title	Duration (min)	Theme	Genre	Relevant reference	Sample size	Dropout	Average \pm std age (years)	Females/males
<i>Avenge But One of My Two Eyes</i> (Mograbi, 2005)	5:27	A political activist confronts Israeli soldiers.	Documentary	(Raz, Touroutoglou, et al. 2016)	74	27	19.51 \pm 1.45	0/74
<i>Sophie's Choice</i> (Pakula, 1982)	10:00	A mother is forced to choose which of her two children will be taken from her.	Drama	(Raz et al. 2012)	44	20	26.73 \pm 4.69	25/19
<i>Stepmom</i> (Columbus, 1998)	8:21	A mother talks with her children about her future death.	Drama	(Raz et al. 2012)	53	21	26.75 \pm 4.86	21/32
<i>The Ring 2</i> (Nakata, 2005)	8:15	A child is lost in a bazaar; the child and his mother are attacked by deer.	Horror, thriller	(Raz, Touroutoglou, et al. 2016)	27	3	26.41 \pm 4.12	11/16
<i>The X-Files</i> , the episode "Home" (Manners, 1996)	5:00	Zombies attack a couple in their home.	Horror, thriller	(Raz, Touroutoglou, et al. 2016)	36	6	23.70 \pm 1.23	14/22
Testing set								
Film title	Duration (min)	Theme	Genre	Relevant reference	Sample size	Dropout	Average \pm std age (years)	Females/males
<i>Stepmom</i> (Columbus, 1998)	8:21	A mother talks with her children about her future death.	Drama	(Raz et al. 2012)	10	1	26.4 \pm 3.17	0/10
<i>Alaska's Wild Denali</i> (Thomas, 1997)	5:00	Nature documentary with narration.	Nature film	(Rottenberg et al. 2007)	5	0	26.6 \pm 4.33	4/1
<i>Dead Poets Society</i> (Weir, 1989)	5:18	Parents find out that their son committed suicide.	Drama	(Schaefer et al. 2010)	5	0	26.6 \pm 4.33	4/1
<i>Forrest Gump</i> (Zemeckis, 1994)	5:21	The protagonist is introduced to his unknown son for the first time.	Drama	(Schaefer et al. 2010)	5	0	26.6 \pm 4.33	4/1

<i>Saving Private Ryan</i> (Spielberg, 1998)	6:18	American troops landing on Omaha Beach during World War II.	War film	(Schaefer et al. 2010)	5	0	26.6±4.33	4/1
<i>Se7en</i> (Fincher, 1995)	6:18	A murdered man tells a detective that he beheaded his pregnant wife.	Thriller	(Schaefer et al. 2010)	5	0	26.6±4.33	4/1
<i>The Fly</i> (Cronenberg, 1986)	8:15	A man is transformed into a giant fly after a conversation with his former lover and after attacking her friend.	Horror, science fiction	N/A	20	5	42.55±7.47	8/12
<i>The Shining</i> (Kubrick, 1980)	5:21	A man pursues his wife with an axe.	Horror	(Schaefer et al. 2010)	5	0	26.6±4.33	4/1
<i>There's Something about Mary</i> (Farrelly Brothers, 1998)	5:00	A man fights with his girlfriend's dog.	Comedy	(Schaefer et al. 2010)	5	0	26.6±4.33	4/1

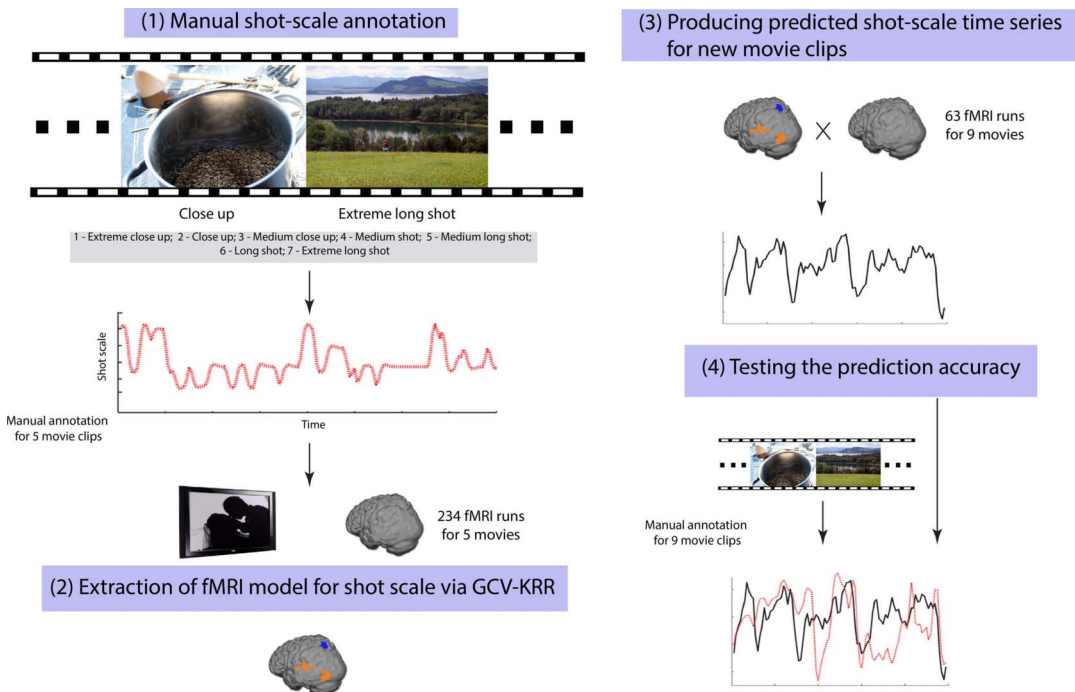


Figure 3: The study outline.

The following procedures are depicted in Figure 3. First, (1) the clips were manually annotated to produce a second-by-second time series of the shot-scale levels (red curve); then (2) a shot-scale model was generated based on a dataset, which included 234 fMRI runs. The orange and blue spots represent the value of the weights given to the colored voxels in the model. While the model assigns weights to any apparent voxel in the illustrated brain, the map is thresholded so only voxels with high positive (orange) or negative (blue) values are colored. Finally, (3) the resulting model was multiplied with fMRI data acquired during the viewing of (mostly) different movies by other individuals; and (4) the predicted shot-scale descriptors (black curve) was compared with a manually annotated descriptor (in red) for each movie at both the individual level and the group average level.

Participants

Volunteers without a known history of neurological or psychiatric disorders were recruited to the study. They had at least twelve years of education with Hebrew as their spoken language. The participants signed a consent form approved by the ethical committees of the Tel Aviv Sourasky Medical Center. In total, the data of 246 volunteers, including 295 fMRI runs, was included in the study. Table 1 presents sample sizes and demographic details.

Stimuli

Heterogeneous cinematic content was selected for this study to represent diversity in terms of genre. We selected sequences from the movies *Dead Poets Society* (Peter Weir, 1989; hereafter named *Poets*), *Forrest Gump* (Robert Zemeckis, 1994; *Forrest*), *Saving Private Ryan* (Steven Spielberg, 1998; *Ryan*), *Seven* (David Fincher, 1995), *The Shining* (Stanley Kubrick, 1980; *Shining*), and *There's Something about Mary* (Farrelly Brothers, 1998; *Mary*). In addition, we presented a sequence from the documentary *Alaska's Wild Denali* (Peter Thomas, 1997; *Denali*). Data from other participants was obtained during the viewing of a sequence from the science fiction film *The Fly* (David Cronenberg, 1986).

We also included data recorded during the viewing of five additional movie excerpts that were collected in previous studies: the documentary *Avenge But One of My Two Eyes* (Avi Mograbi, 2005; *Avenge*), the drama *Sophie's Choice* (Alan Pakula, 1982; *Sophie*), the thriller *The Ring 2* (Hideo Nakata, 2005; *Ring*), the drama *Stepmom* (Chris Columbus, 1998), and a clip from the episode "Home" of the television series *The X-Files* (Fox, 1993–2002, 2016–). More details on the duration, genre, and content of these cinematic excerpts are presented in Table 1.

We presented the movies to the participants, who were lying in the scanner. They were instructed to passively view the movies that were displayed by

an LCD projector to a mirror located above their heads. Active noise-canceling headphones (made by Optoacoustics) were used during the scans.

Shot-Scale, Face Presence, and Face-Ratio Annotation

The shot-scale sequence of each clip was annotated by hand on a seven-grade shot-size scale. Shot-scales were identified as a time sequence on a second-by-second basis. Each second of every clip was assigned with a shot-scale, which provided for each clip a chart showing how long a given shot-scale lasts, when it changes, and what the following shot-scale was. The exact size and name of a given shot-scale varied somewhat by convention, as shot-scales are by nature arbitrary segmentations of a continuous measure. However, the sizes designated by the categories at the close end of the scale are fairly universal. Thus “extreme close-up” (ECU) means an image where a small part of an object or a face fills the entire frame; “close-up” (CU) universally means an image where a full face fills the frame; and medium close-up (MCU) designates a framing where a person is represented from the chest up. The middle range consists of two categories: “medium shot” (MS) from the waist or knee up and “medium long shot” (MLS), where the entire human figure is visible. At the far end of the scale, we distinguished two categories again: “long shot” (LS), an image representing a group of people and a considerable amount of space around them, and “extreme long shot” (ELS), which were pictures with a very big depth of field and where human characters were too small to be identifiable. The exact naming and the corresponding shot-scale categories were not relevant because we did not expect viewers to process visual information using these categories, not the least because most viewers are not even familiar with them. The important thing was to have a categorization detailed enough to use but not so detailed to model how viewers process apparent distance on the picture, which corresponds by and large to the terminology of film criticism and to the way these terms are used in the film industry. Two independent coders worked on each clip, and a third person controlled their coding and made decisions in cases of disagreement.

The interrater reliability of our procedure was computed by comparing the resulting shot-scale annotation with annotations obtained by two additional independent raters who coded the first 121 seconds of each of the 13 clips examined in our study. This sample encompassed over 1,573 seconds, which comprised about 31% of the total duration of the clips. Fleiss’s Kappa index of agreement between the three annotations was computed using a Matlab function (Cardillo 2007). We found a Fleiss’s Kappa coefficient of 0.64, which indicates substantial interrater agreement (95% confidence interval: 0.63587–0.64349). Following a standard procedure in fMRI data analysis, we convolved the annotation with a canonical hemodynamic response function (HRF). The HRF models the expected blood flow changes following a specific neural event

(the implementation of a specific shot-scale in our case). Since fMRI is sensitive to blood flow, it is a common practice to compare its data with an estimated reaction pattern, which applies the HRF on the experimental design (via mathematical convolution).

Data Analysis

We first tested the relations between shot-scale, on the one hand, and face ratio and face presence on the other. Face ratio was automatically annotated for each of the 13 movie clips used in the study. The automatic annotation (also applied to this data in Raz et al. 2017) relied on a method proposed by Xiangxin Zhu and Ramanan in 2012. In each frame, we annotated the ratio of the largest face in the image to the total area of the frame. The resulting values were averaged in one-second bins for the analysis of the relations between face presence and shot-scale and in three-second bins in two further neuroimaging analyses to fit with the temporal resolution of the fMRI signal. No face annotation was made for the clip from *Denali*, which did not include a human face.

To test the hypothesis that face presence is more common in closer shots, we compared the prevalence of face presence across three shot-scale categories: close shots (ECU and CU), medium shots (MCU, MS), and long shots (MLS, LS, and ELS). Face presence was defined as positive for frames in which the value of the automatic annotation of face ratio was higher than zero. One-way ANOVA was used to test whether the shot-scale categories differed in the proportion of bins that included face images. Before testing, we converted the proportion to linearized units using the rationalized arcsine transform (Studebaker 1985).

In the main part of our analysis, the shot-scale model, which was generated via GCV-KRR, was applied to our testing fMRI data obtained during the viewing of nine movies. This step resulted in the generation of predicted shot-scale descriptors for each individual scan. For each of the movies, these descriptors were compared with the manual shot-scale annotation (after they were convolved with the HRF) using Pearson's correlation coefficients either at the individual level or at the level of the group's averaged predicted shot-scale descriptor. A permutation test was used to assess the statistical significance of the results.

We used face-ratio annotation in two additional analyses. To examine the extent to which our shot-scale model captured parameters other than the relative size of the face, we tested for differences between the shot-scale model and a face-ratio model that was generated through the same procedure. Hence, we repeated our decoding procedure, but this time with the automatic face-ratio annotation (after it was convolved with the HRF) for all movies but *Denali*, in which no human face was presented. The resulting face-ratio fMRI model was then used to predict shot-scale annotation for the nine testing movies. To estimate the added value of the shot-scale model relative to the face-ratio model, we compared (a) the average predicted and observed shot-scale descriptors

based on each of the models for each of the movies; and (b) the predicted and observed shot-scale descriptors based on each of the models for every individual fMRI session. For these analyses, we used Student's t -test, comparing the Fisher-Z transformed coefficients (Fisher Z transformation allows the correction of deviations from normal distribution) for the correlation between the original shot-scale annotations and both individual and average predicted annotations.

Results

Shot-Scale Decoding

We generated a standard shot-scale model by applying GCV-KRR to the training fMRI data. This model includes voxelwise weights. To decode shot-scale in a new movie, the model is simply multiplied by a matrix containing the fMRI data that was recorded during the viewing of this movie. The robustness of the model was assessed using a challenging design: fMRI data obtained from new participants watching movies, which was not included in the dataset on which the model was trained (except for *Stepmom*), was in most cases (except for *Stepmom* and *Fly*) also acquired by a different MRI scanner. In this experimental design, significant correlations between the shot-scale annotations that were predicted based on our model and the original annotations (after they were convolved with the HRF) would indicate that this feature can be robustly decoded in a way that is generalizable across viewers.

As illustrated in Figures 4 and 5, the shot-scale time series decoded from the viewers' fMRI data (averaged for each movie) significantly correlated with the original annotations for all nine movie clips tested in our study. A partial conjunction analysis, which is used for the assessment of the reproducibility of the results, confirmed significant correlations in all nine comparisons at $Q_{\text{FDR}} < 0.05$. The correlation ranged between 0.26 and 0.78 with an average of 0.52 across movies. At the individual level, the correlation between the predicted and the original annotation was significant in 31 out of 65 runs (47.79%) as confirmed by an FDR corrected partial conjunction analysis. The average correlation between the predicted and the manual annotation at the individual level was 0.37 with a standard deviation of 0.16.

The "inverse model" (Table A1), which predicts shot-scale based on fMRI signal, was transformed into a univariate "forward model," which predicts the fMRI based on shot-scale annotation, to allow for its interpretation in functional terms (see Appendix and Haufe et al. 2014). The three-dimensional forward model is visualized in Figure 6 (for details, see Table 2) after controlling for the effect of face presence (details on the inverse face presence model are presented in Table A2). Its major components (in terms of highest absolute

The shot-scale time series decoded from the viewers' fMRI data (averaged for each movie) significantly correlated with the original annotations for all nine movie clips tested in our study.

Table 2. Components of the forward MVPA shot-scale model: clusters' anatomical label, location (in Talairach coordinates), and size. The localization relied on the probability atlas published in Wang et al. 2015.

Region label	Mean weight	Cluster size (mm ³)	X	Y	Z
Shot-scale					
Positive weights (close shots)					
R superior temporal cortex	4.59	310	45	-31	1
L superior temporal cortex	3.86	442	-51	-43	7
R temporal pole	3.51	181	48	5	-17
L temporal pole	2.99	84	-51	2	-11
R precentral gyrus	3.02	92	48	-4	49
L precentral gyrus	2.46	43	-45	-4	52
R inferior frontal gyrus	2.24	43	48	23	10
L inferior frontal gyrus	1.72	5	-45	23	7
R fusiform gyrus	3.02	42	39	-46	-17
L fusiform gyrus	3.37	94	-39	-46	-17
R centromedial amygdala	1.94	10	21	-7	-8
L centromedial amygdala	1.84	3	-18	-7	-8
R dorsomedial prefrontal cortex	2.3	75	6	50	34
R supplementary motor area	2.22	64	6	2	64
L inferior / middle frontal gyrus	2.08	51	-51	11	28
Negative weights (long shots)					
R frontal eye field	-2.44	46	24	-1	55
R areas V3a, V3b, V3d, middle temporal gyrus (MT area), visual, precuneus, parietal region V7, transverse occipital sulcus (TOS)	-5.76	778	33	-79	16
L areas V3a, V3b, V3d, middle temporal gyrus (MT area), visual, precuneus, parietal region V7, transverse occipital sulcus (TOS)	-5.03	696	-33	-82	16
R parahippocampal gyrus (PHC1, PHC2), visual area V4	-6.89	392	27	-46	-5
L parahippocampal gyrus (PHC1, PHC2), visual area V4	-6.25	380	-27	-46	5
R parietooccipital sulcus	-3.52	127	21	-58	19
L parietooccipital sulcus	-3.07	36	-21	-61	16
L visual area V1	-2.86	97	-12	-88	-8

weights) are large bilateral clusters whose high weights indicate close shots in the superior temporal gyrus, temporal poles, precentral gyrus, and the pFg. On the other hand, far shots were associated with bilateral large clusters in the associative visual areas, including the TOS and the parahippocampal gyrus.

Shot-Scale, Face Presence, and Face Ratio

Are face images more frequent in close relative to distant shots? We examined the question of whether shot-scale categories differed with regard to the prevalence of face images, and we found that the proportion of face images was $63.04 \pm 8.5\%$ (mean \pm standard deviation), $77.21 \pm 12.5\%$, and $37.88 \pm 28.8\%$ for close, medium, and long shots, respectively (Figure 7). The interaction between shot-scale and face presence was significant ($F(36) = 6.9, p < 0.005$), and a post hoc t -test confirmed that close shots and medium shots had higher proportions of face images relative to long shots ($t = 2.92, p < 0.05$; $t = 5.96, p < 0.0005$, respectively, after Bonferroni correction). However, somewhat surprisingly no significant difference in face image proportion was found between close and medium shots ($t = 2.06$).

As a measure of specificity for our model, we tested whether the original shot-scale annotation is better predicted by the shot-scale model relative to a face-ratio model generated via a similar procedure (Figure 8). The manual shot-scale annotations were more strongly correlated with the predicted time series that were based on the shot-scale model than with those that were based on the face-ratio model both when Pearson's correlation coefficients were compared for every individual scan ($t(59) = 4.3, p = 6.5 \times 10^{-05}$) and when the average predicted time series for each movie ($t(7) = 3, p = 0.02$) were compared.

As can be seen in Figure 4, there was a correlation between the average predicted and the observed shot-scale descriptor for each of the testing mov-

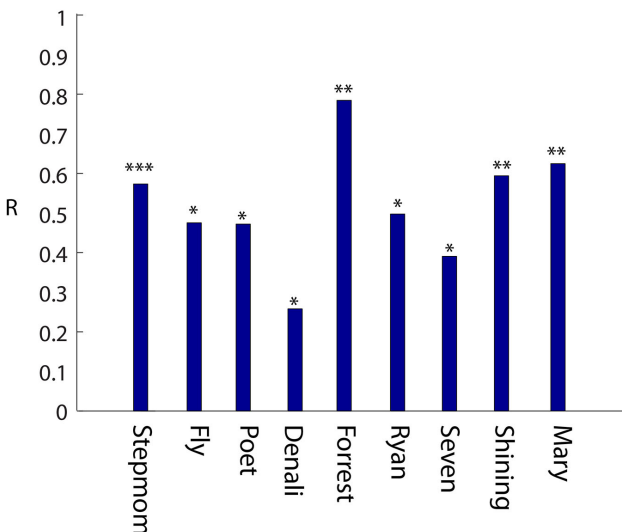


Figure 4. Shot-scale prediction accuracy.

ies. These results indicate the robustness of the shot-scale decoding, as this “mind-reading” procedure yielded successful decoding in movies that were not part of the data that was used to generate the model ($* p < 0.05$, $** p < 0.005$, $*** p < 0.0001$).

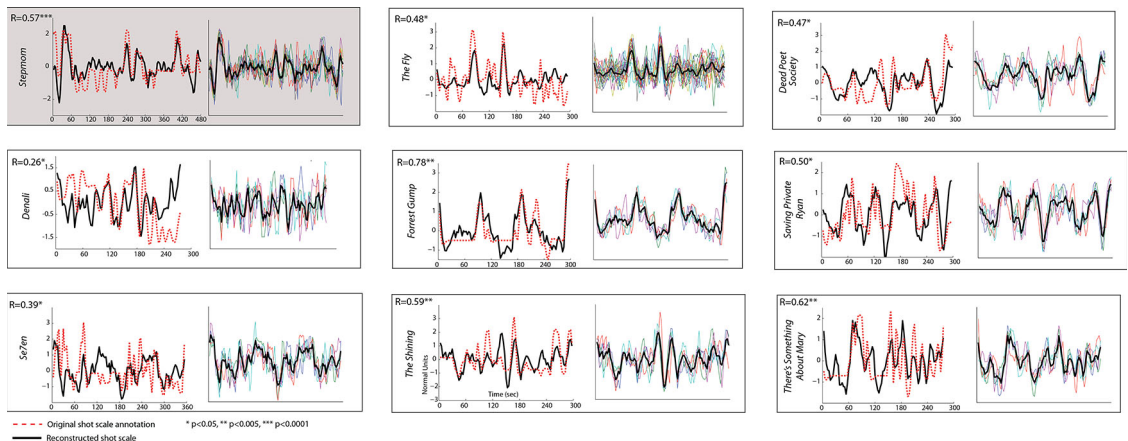


Figure 5. Time courses of predicted and observed shot-scale descriptors

In Figure 5, the left panes present the average predicted descriptor (black) and the observed descriptor (dashed red curve). The right pane presents individual predicted descriptors (colored) and their average (black). A notable similarity between the individual predicted descriptors is evident, pointing to the cross-viewer robustness of this measure.

As indicated in Figure 6, the map was thresholded at $Q_{FDR} < 0.05$. The weights were transformed so that they would range between -8.9 and 8 . As described above, each voxel in the model was assigned with a weight, which reflected its contribution to the prediction of the shot-scale. The orange and blue patches indicate positive and negative correlations with shot-scale, respectively, so that voxels that are colored in orange were activated during close shots and voxels in blue were activated when perceiving distant shots. The abbreviations used were as follows: mPFC—medial prefrontal cortex; MT—middle temporal gyrus; OFA—occipital face area; PHC—parahippocampal cortex; and TOS—transverse occipital sulcus (as above).

As shown in Figure 7, face images were more frequent in close-ups and medium shots relative to more distant shots, but no statistically significant difference was found between extreme close-up and close-up shots, on the one hand, and medium close-ups and medium shots, on the other. As already mentioned above, the abbreviations used here were as follows: ECU—extreme close-up, CU—close-up, MCU—medium close-up, MS—medium shot, MLS—medium long shot, LS—long shot, ELS—extreme long shot ($* p < 0.005$ (Bonferroni corrected)).

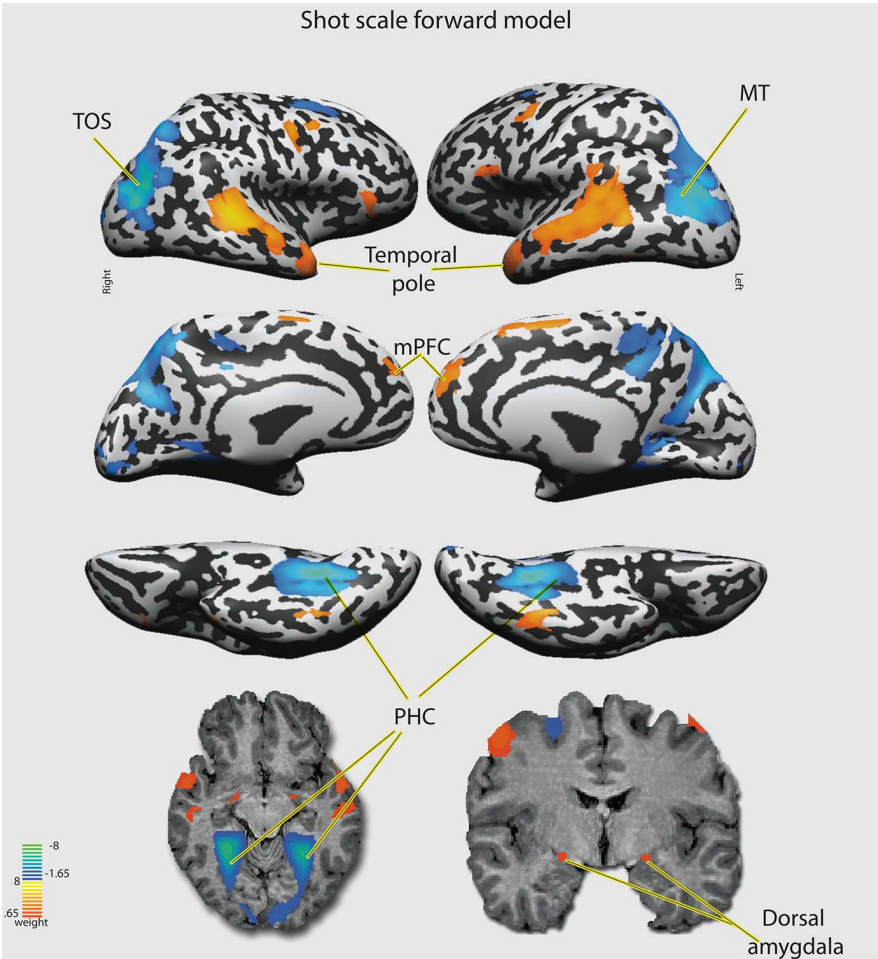


Figure 6. Three-dimensional map of the brain "forward" model for shot-scale.

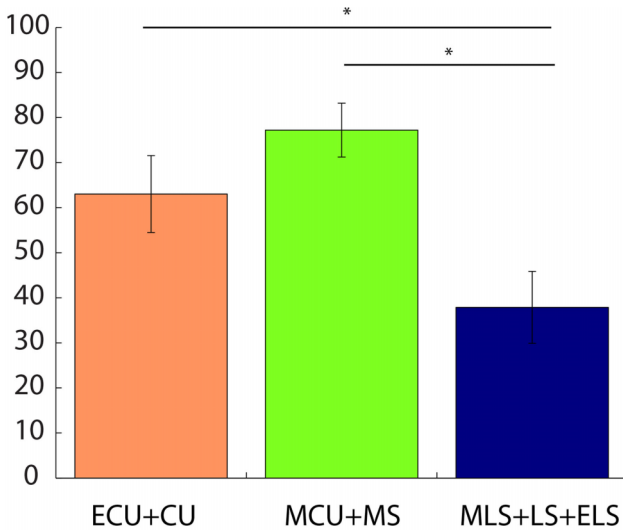


Figure 7. The proportion of face images in three shot-scale categories.

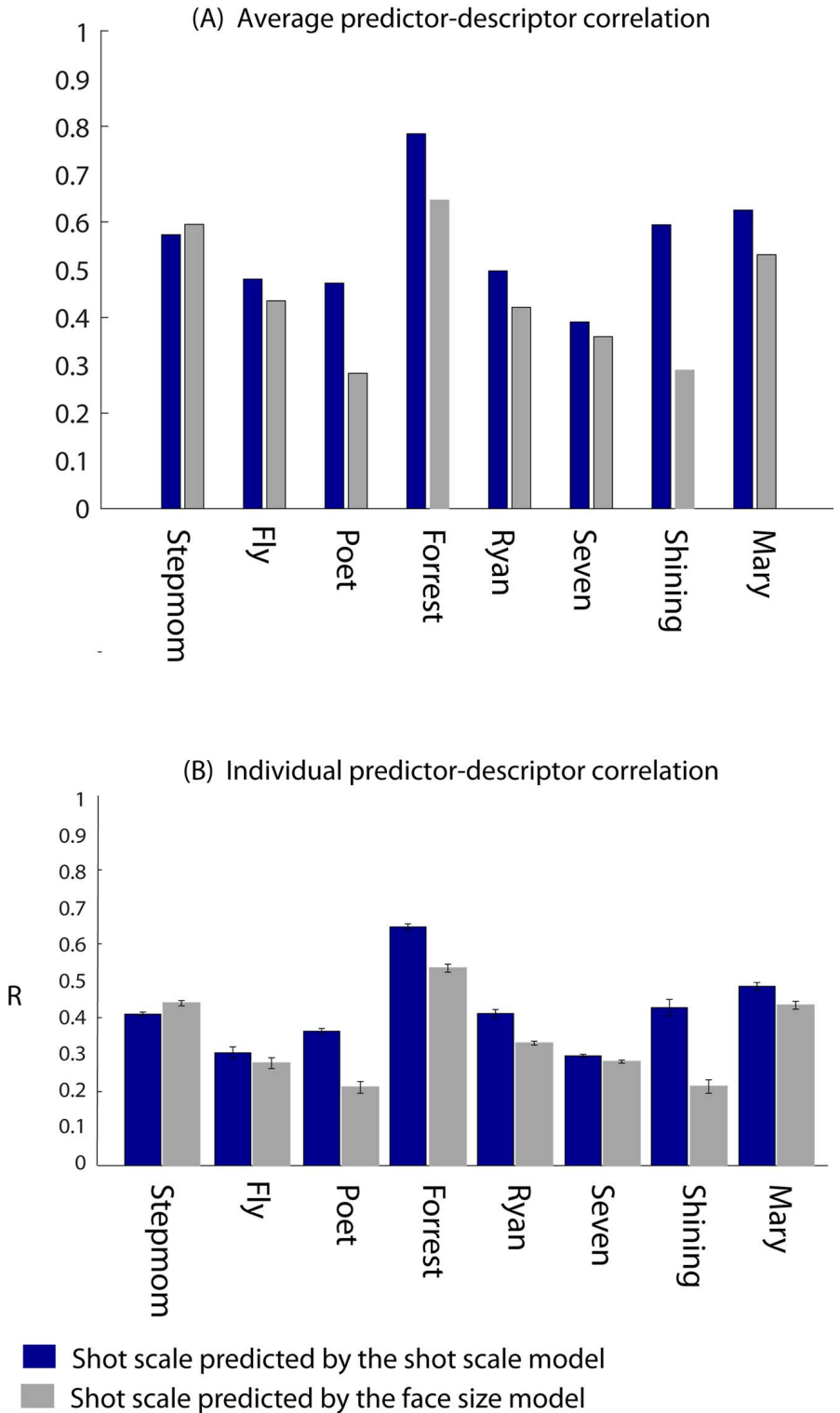


Figure 8. Specificity of the shot-scale model relative to the face-ratio model.

In Figure 8, (A) is the prediction accuracy when using either the averaged observed shot-scale descriptor (blue) or the average face ratio (gray); and (B) is average prediction accuracy and standard error measured for individual shot-scale (blue) and face-ratio (gray) predictors. It is evident that in all movies, except for *Stepmom*, the ground truth manual annotation of shot-scale was better predicted when using the shot-scale rather than the face-ratio model. In other words, the shot-scale brain model captures information that is relevant to shot-scale and that is not captured by the face-ratio model alone.

Discussion

Our study provides robust evidence for the existence of consistent neural patterns that correlate with empirical cinematic shot-scales across genres and movies. To the best of our knowledge, this evidence on the robustness of shot-scale decoding is the first of its kind in the literature, and as argued above, it is not trivial that an elementary cinematic feature would be decoded with such consistency and generalizability. The statistically significant decoding of shot-scale in all nine testing movies indicates that, despite the diversity in terms of content, the cinematic shot-scale is empirically correlated with specific patterns of brain activity. This correlation is even less trivial when considering the fact that the shot-scale annotation was somewhat ambiguous (as indicated by the obtained Fleiss's Kappa coefficient of 0.64 for interrater agreement). Despite the noise in the annotation of this feature as made by human coders, our algorithm detected a correlation that showed sufficient consistency to meet our stringent statistical criteria, which indicates its robustness.

The interpretation of these patterns as they appear in the shot-scale model (Figure 6) is probably the most interesting part of our work for film scholars interested in the psychological significance of shot-scale. However, before we turn to the interpretation of the neuroimaging findings, a disclaimer should be made. The procedure of inferring mental states based on neuroimaging maps is called *reverse inference* (Poldrack 2011). Unlike forward inference procedures, in which a specific psychological parameter is experimentally manipulated (e.g., valence of emotional face) and the resulting pattern is associated with this manipulated parameter, reverse inference interprets the findings in terms of psychological functions that were not manipulated directly. For example, in the interpretation outlined below shot-scales are linked with specific psychological processes (e.g., theory of mind) that were not directly manipulated or proved by us. In that case, reverse inference associates the activation of a certain brain region (or a set of regions) with a specific psychological function without any formal consideration of other functions that are related to this region. Thus, while the following interpretation is motivated by the notion of theoretical coherence, we emphasize that, as long as the likelihood of alternative explanations is not considered, our theoretical proposition is speculative

and hypothetical. It is valuable to the extent that it allows the formulation of specific hypotheses that can be tested using forward or more formal reverse inference (Poldrack 2011). We will suggest such specific directions for future work after outlining our interpretation.

So what may the functional meaning of the shot-scale model visualized in Figure 6 be? First, in this model close shots correlate with a bilateral focal activation of the centromedial section of the amygdala, which has a key role in the processing of emotional information. The centromedial amygdala, in particular, constitutes a major output station of the amygdala (Jalbrzikowski

In line with Balázs’s hypothesis, the findings that the activity of this region correlates with close shots after controlling for face presence and that the shot-scale model yielded significant decoding success in the absence of human face images (in the case of Denali) indicate that some emotional aspects of apparent proximity are independent of face presence in close-ups.

et al. 2017), and it is implicated in the allocation of attention to salient events and in the initiation of situation-appropriate responses of the autonomic nervous system (Davis and Whalen 2001; Gallagher et al. 1990). In line with Balázs’s hypothesis, the findings that the activity of this region correlates with close shots after controlling for face presence and that the shot-scale model yielded significant decoding success in the absence of human face images (in the case of *Denali*) indicate that some emotional aspects of apparent proximity are independent of face presence in close-ups. While probably most close-ups in film history are close-ups of faces, and while the mirroring of facial expressions

is certainly an important factor in their effect, not all close-ups are about faces. It is possible that even when facial expressions are depicted in close-up the very perception of apparent spatial proximity implicates enhanced emotional effects that are not entirely explained by the impact of the increased salience of the larger face on the viewer’s attention. This notion has to be further confirmed by independent studies comparing perception of faces and objects for their emotional effects from different apparent distances as suggested below. This way, it would be possible to discern the effects of individual features in a complex visual stimulus, where some effects are intended and others are not.

The centromedial amygdala was coactivated in close shots with the bilateral pFg, which is part of the ventral visual stream (Rosenke et al. 2018). According to the original “two-streams hypothesis” (Rosenke et al. 2018), the ventral stream (located at the lower-rear part of the brain) specializes in object identification (the “what pathway”), whereas the dorsal stream (which encompasses the upper-rear part of the brain) supports spatial localization (the “where pathway”). In updated accounts (e.g., Goodale 2014; Milner 2017), the ventral pathway underpins the processing of the object’s visual qualities in the context of recognition, memory, and conscious perception, while the

dorsal stream processes spatial information and movement to provide visual control over manual actions:

[T]he dorsal stream plays a critical role in the real-time control of action, transforming moment-to-moment information about the location and disposition of objects into the coordinate frames of the effectors being used to perform the action. By contrast, the ventral stream (together with associated cognitive networks) constructs the rich and detailed visual representations of the world that allow us to identify objects and events, attach meaning and significance to them and establish their causal relations. (Goodale 2014, 2)

Thus, a possible interpretation for the coactivation of the pFg and the centromedial amygdala in close shots is that this type of cinematic framing is empirically associated with a mode of enhanced processing of transient stimuli and recognition in high acuity. It is therefore dominated by visual and semantic memory. Taking into consideration the empirical evidence on psychological distance, which was reviewed above (Liberman and Trope 2008), this mental mode may involve a bias toward concrete and contextualized assessments made at the proximal construal level. On the other hand, the coactivation of bilateral parietal and occipital areas encompassing the dorsal stream in distant shots implies the dominance of an action-related framework, implicating orientation control and movement monitoring. This interpretation is congruent with the functional profile of another major bilateral cluster, which is activated in far shots: the parahippocampal gyrus, which is implicated in goal-directed navigation, spatial contextualization, encoding of topographical scene, and landmark processing (Chan et al. 2012; Hasselmo et al. 2017).

Finally, close shots were correlated with the activation of a large bilateral cluster in the superior temporal cortex (STC). This large area includes regions that have been associated with various functions including language comprehension (Friederici 2012), processing of auditory information in general and vocal expressions in particular (Friederici 2012), joint attention (Friederici 2012), and analysis of various social cues (e.g., Castelli et al. 2000; Isik et al. 2017; Schultz et al. 2005). Large-scale activations across the STC were also observed specifically in the context of narrative processing and comprehension (Szaflarski et al. 2012; Wilson et al. 2008) and narrative event boundaries (Zacks et al. 2010).

Thus, the activation of the STC in close shots may result from the link between these shot-scales and various parameters including increased prevalence of dialogue, abundance and salience of social cues, or, more interestingly, richer syntactic elements such as shot transitions. Further research could ex-

A possible interpretation for the coactivation of the pFg and the centromedial amygdala in close shots is that this type of cinematic framing is empirically associated with a mode of enhanced processing of transient stimuli and recognition in high acuity.

On the other hand, the coactivation of bilateral parietal and occipital areas encompassing the dorsal stream in distant shots implies the dominance of an action-related framework, implicating orientation control and movement monitoring.

amine which of these parameters correlates with close shots. Interestingly, close shots correlate with the bilateral superior temporal sulcus, which is part of the STC cluster, the temporal poles, and part of the medial prefrontal cortex. These regions have been considerably associated with theory of mind—that is, the attribution of mental states (e.g., beliefs, goals) to another individual (Dodell-Feder et al. 2011). This finding is congruent with recent findings that close-ups

enhance theory-of-mind manifestations when viewers recall movies' narratives (Bálint et al. 2018; Rooney and Bálint 2018).

While it is important to note that these patterns of brain activity are empirically, but not necessarily inherently, linked with shot-scales, it is worth noting the agreement between our shot-scale model and Amit and colleagues' (2012) tightly controlled study of apparent distance. In this study, the pFg was associated with apparent proximity, while the TOS and PPA increased their activity level in apparent remoteness. Accordingly, in our study the pFg was coactivated with close shots, whereas the TOS and PPA were linked with far shots. This overlap is congruent with the notion that cinematic shot-scales covary with apparent distance, but as emphasized above it cannot be regarded as conclusive evidence, since apparent distance was not controlled in our study.

The hypotheses outlined above could be tested in a future empirical study of apparent distance in motion pictures. By employing the Ponzo illusion to generate alternative versions of specific shots, it is possible to induce varying degrees of perceived distance while keeping intervening variables (e.g., dialogue and optical flow) constant. Such an experimental setup will allow for the testing of our hypotheses regarding the link between perceived distance and emotion-related functions, perceived granularity, abstraction, and goal-orientation. The link between these notions could be established based on explicit and implicit behavioral measures, as well as on physiological and neuroimaging probes. In the latter case, we expect at least partial replication of the results of the current study.

Finally, in light of our findings on the possible link between shot-scales and mental modes, the statistical distribution of this cinematic feature could reveal a balanced or skewed overall psychological distance effect in specific films or film corpora. The shot-scale distribution is a measure representing the overall presence of a given shot-scale in the entire film expressed as a percentage of the entire playing time of the film (see Kovács 2014). The dominance of closer or wider shots in a film may correlate with other stylistic or thematic features that will evoke similar psychological distance ranges. We have some evidence in this regard. A recent study showed that medium close-ups are consistently and significantly more dominant in Hollywood films than

in European art films. In fact, European art films' shot-scale distributions are more balanced between the different categories, and some even show a dominance of long shots, while Hollywood films' shot-scale distributions consistently show a dominance of medium close-ups (Kovács and Zentay 2017). One plausible explanation of this phenomenon is that Hollywood films are much more concerned with facilitating the viewer's connecting with the protagonist (to bring them "close" to the viewer)—as evidenced by virtually every scriptwriting manual published by Hollywood scriptwriters—than are European art films, where alienating or at least distancing the viewer from the protagonist's world is a more common procedure. An interesting type of shot composition also confirms this hypothesis. In some cases, films use compositions containing two or three shot-scales. Images where a face of an object is placed in the foreground while a scene is shown in the background, and where both planes are in focus, cannot be categorized either as a close-up or as a medium or medium long shot, because it is both. It can be named "foreground shot" (Kovács 2014). These compositions usually create a tension between foreground and background, which could be formulated also in terms of psychological distance as evidenced by the fact that these kinds of compositions abound in horror films and thrillers, where the foreground object always evokes the closeness of a concrete danger threatening the safety of characters or objects in the distance. We would predict that other stylistic features strengthen the effects of closing upon or distancing the viewer from the story world of the films in other dimensions of psychological distance.

Conclusion

Art history discovered "shot-scales" in the beginning of the twentieth century, and later on film theory also attributed a great importance to this visual feature. Up till now, different theories have existed regarding the possible interpretation of the effects of shot-scales depending on the perceptual level at which the theorist imagined shot-scales to exert their effects. The results of our study are congruent with the notion that shot-scales as an empirical phenomenon in cinema are linked with specific mental modes. We speculate that these modes involve bias toward contextual and concrete perceptual analysis and recognition in close shots and more general action-related processes in distant shots. Furthermore, as the effect that was revealed in our study was not a by-product of the increased presence of human faces in close shots, it seems that theorists emphasizing the mirroring effect of facial close-ups are not wrong but may be disregarding an important ingredient in the effect, which is closeness and distance. Shot-scales are to a large extent responsible for the interpretation of films. The more we know about the immediate mechanism of their perception, the more we will be able to understand their high-level impacts.

Acknowledgments

We would like to thank Gadi Gilam, Maya Bleich Cohen, Tamar Lin, Roei Admon, Tal Gonen, and Avner Thaler for collecting fMRI data as well as Nóra Zentay for collecting and managing manual annotations. We would also like to thank Rainer Goebel and Talma Hendler for their support in the work for this article. This article was supported by the Human Enhancement and Learning (HEaL) research program, by the BRAINTRAIN consortium under the EU FP7 Health Cooperation Work Program (602186), and by BIAL Grants for Scientific Research 299/14. We are thankful to the reviewers of this article for their detailed and thoughtful remarks, which have helped to make our article clearer and more accurate.

András Bálint Kovács is Professor and Founding Chair of the Film Department at Eötvös Loránd (ELTE) University, Budapest, Hungary. He teaches history of modern cinema and film analysis. He is a Recurrent Visiting Professor at the University of California, San Diego, and was formerly a Visiting Professor at the École Normale Supérieure, Paris; the Université de la Nouvelle Sorbonne, Paris, and the University of Stockholm. His current research projects include quantitative style analysis and the psychological research of emotion regulation and causal thinking in film viewing. E-mail: kab@btk.elte.hu

Gal Raz is a Senior Lecturer at the Steve Tisch School of Film and Television and the Sagol School of Neuroscience, Tel Aviv University. He is currently establishing an interdisciplinary lab for immersive technologies at the Sagol Brain Institute at the Tel Aviv Sourasky Medical Center. Gal has been using neuroimaging tools to study various aspects of audiovisual experiences, including aesthetics, emotion regulation, empathy, and the potential of multimodal interfaces in neurofeedback learning. E-mail: galraz@post.tau.ac.il

Giancarlo Valente is an Assistant Professor in the Departments of Audition and Cognitive Neuroscience in the Faculty of Psychology and Neuroscience at Maastricht University in the Netherlands. E-mail: giancarlovalente@maastrichtuniversity.nl

Sergio Benini received his MSc in Electronic Engineering (*cum laude*) at the University of Brescia with a thesis granted by the Italian Academy of Science in 2000. Between 2001 and 2003, he worked in the Research and Development Department at Siemens Mobile Communication. He received his PhD in Information Engineering from the University of Brescia in 2006, working on video content analysis. During his PhD, he completed a one-year placement at British Telecom Research, UK, working in the Content Coding Lab. He is currently Assistant Professor at the University of Brescia. E-mail: sergio.benini@unibs.it

Michele Svanera received his MSc in Telecommunications Engineering at the University of Brescia (2013) with a thesis on methods and models for the synthesis and representation of three-dimensional surfaces. He is currently a PhD student at the same institution. E-mail: m.svanera005@unibs.it

References

- Amit, Elinor, Eyal Mehoudar, Yaacov Trope, and Galit Yovel. 2012. "Do Object-Category Selective Regions in the Ventral Visual Stream Represent Perceived Distance Information?" *Brain and Cognition* 80 (2): 201–213. <https://doi.org/10.1016/j.bandc.2012.06.006>.
- Arcaro, Michael J., Stephanie A. McMains, Benjamin D. Singer, and Sabine Kastner. 2009. "Retinotopic Organization of Human Ventral Visual Cortex." *Journal of Neuroscience* 29 (34): 10638–10652. <https://doi.org/10.1523/JNEUROSCI.2807-09.2009>.
- Balázs, Béla. 1924. "Der Sichtbare Mensch Oder Die Kultur Des Films" In *Béla Balázs: Early Film Theory*, ed. Erica Carter, 1–87. New York: Berghahn Books.
- Bálint, Katalin, Thomas Klausch, and Tibor Pólya. 2018. "Watching Closely." *Journal of Media Psychology* 30 (3): 150–159. <https://doi.org/10.1027/1864-1105/a000189>.
- Bar-Anan, Yoav, Nira Liberman, Yaacov Trope, and Daniel Algom. 2007. "Automatic Processing of Psychological Distance: Evidence from a Stroop Task." *Journal of Experimental Psychology: General* 136 (4): 610–622. <https://doi.org/10.1037/0096-3445.136.4.610>.
- Benini, Sergio, Michele Svanera, Nicola Adami, Riccardo Leonardi, and András Bálint Kovács. 2016. "Shot Scale Distribution in Art Films." *Multimedia Tools and Applications* 75 (23): 16499–16527. <https://doi.org/10.1007/s11042-016-3339-9>.
- Cardillo, Giuseppe. 2007. "Fleiss'es Kappa: Compute the Fleiss'es Kappa for Multiple Raters." <http://www.mathworks.com/matlabcentral/fileexchange/15426>.
- Carroll, Noël. 1996. *Theorizing the Moving Image*. Cambridge: Cambridge University Press.
- Castelli, Fulvia, Francesca Happé, Uta Frith, and Chris Frith. 2000. "Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns." *NeuroImage* 12 (3): 314–325. <https://doi.org/10.1006/nimg.2000.0612>.
- Chan, Edgar, Oliver Baumann, Mark A. Bellgrove, and Jason B. Mattingley. 2012. "From Objects to Landmarks: The Function of Visual Location Information in Spatial Navigation." *Frontiers in Psychology* 3: 304. <https://doi.org/10.3389/fpsyg.2012.00304>.
- Changizi, Mark A., Andrew Hsieh, Romi Nijhawan, Ryota Kanai, and Shinsuke Shimojo. 2008. "Perceiving the Present and a Systematization of Illusions." *Cognitive Science* 32 (3): 459–503. <https://doi.org/10.1080/03640210802035191>.
- Cutting, James E. 2015. "The Framing of Characters in Popular Movies." *Art & Perception* 3 (2): 191–212. <https://doi.org/10.1163/22134913-00002031>.
- Cutting, James E., and Kacie L. Armstrong. 2016. "Facial Expression, Size, and Clutter: Inferences from Movie Structure to Emotion Judgments and Back." *Attention, Perception, & Psychophysics* 78 (3): 891–901. <https://doi.org/10.3758/s13414-015-1003-5>.
- Cutting, James E., and Ayse Candan. 2015. "Shot Durations, Shot Classes, and the Increased Pace of Popular Movies." *Projections* 9 (2): 40–62. <https://doi.org/10.3167/proj.2015.090204>.
- Cutting, James E., Jordan E. DeLong, and Christine E. Nothelfer. 2010. "Attention and the Evolution of Hollywood Film." *Psychological Science* 21 (3): 432–439. <https://doi.org/10.1177/0956797610361679>.
- Cutting, James, and Catalina Iricinschi. 2015. "Re-Presentations of Space in Hollywood Movies: An Event-Indexing Analysis." *Cognitive Science* 39 (2): 434–456. <https://doi.org/10.1111/cogs.12151>.

- Davis, Marilyn, and Paul J. Whalen. 2001. "The Amygdala: Vigilance and Emotion." *Molecular Psychiatry* 6 (1): 13–34. <https://doi.org/10.1038/sj.mp.4000812>.
- Deleuze, Gilles. 1986. "The Affection-Image: Face and Close-Up." In *Cinema 1: The Movement-Image*. Trans. Hugh Tomlinson and Barbara Habberjam, 87–101. Minneapolis: University of Minnesota Press.
- Dodell-Feder, David, Jorie Koster-Hale, Marina Bedny, and Rebecca Saxe. 2011. "fMRI Item Analysis in a Theory of Mind Task." *NeuroImage* 55 (2): 705–712. <https://doi.org/10.1016/J.NEUROIMAGE.2010.12.040>.
- Epstein, Jean. 1921. "Grossissement." In *Bonjour Cinéma*, 93–108. Paris: Éditions de la Sirène.
- Friederici, Angela D. 2012. "The Cortical Language Circuit: From Auditory Perception to Sentence Comprehension." *Trends in Cognitive Sciences* 16 (5): 262–268. <https://doi.org/10.1016/J.TICS.2012.04.001>.
- Gallagher, Michela, Phillip W. Graham, and Peter C. Holland. 1990. "The Amygdala Central Nucleus and Appetitive Pavlovian Conditioning: Lesions Impair One Class of Conditioned Behavior." *Journal of Neuroscience* 10 (6): 1906–1911. <https://doi.org/10.1523/jneurosci.3225-10.2010>.
- Goodale, Melvyn A. 2014. "How (and Why) the Visual Control of Action Differs from Visual Perception." *Proceedings of the Royal Society B: Biological Sciences* 281 (1785): 1–9. <https://doi.org/10.1098/rspb.2014.0337>.
- Hasselmo, Michael E., James R. Hinman, Holger Dannenberg, and Chantal E. Stern. 2017. "Models of Spatial and Temporal Dimensions of Memory." *Current Opinion in Behavioral Sciences* 17: 27–33. <https://doi.org/10.1016/J.COBEHA.2017.05.024>.
- Hasson, Uri, Yuval Nir, Ifat Levy, Galit Fuhrmann, and Rafael Malach. 2004. "Intersubject Synchronization of Cortical Activity during Natural Vision." *Science* 303 (5664): 1634–1640. <https://doi.org/10.1126/science.1089506>.
- Haufe, Stefan, Frank Meinecke, Kai Görden, Sven Dähne, John-Dylan Haynes, Benjamin Blankertz, and Felix Bießmann. 2014. "On the Interpretation of Weight Vectors of Linear Models in Multivariate Neuroimaging." *NeuroImage* 87: 96–110. <https://doi.org/10.1016/j.neuroimage.2013.10.067>.
- He, Dongjun, Ce Mo, Yizhou Wang, and Fang Fang. 2015. "Position Shifts of fMRI-Based Population Receptive Fields in Human Visual Cortex Induced by Ponzo Illusion." *Experimental Brain Research* 233 (12): 3535–3541. <https://doi.org/10.1007/s00221-015-4425-3>.
- Henderson, Marlone D., Kentaro Fujita, Yaacov Trope, and Nira Liberman. 2006. "Transcending the 'Here': The Effect of Spatial Distance on Social Judgment." *Journal of Personality and Social Psychology* 91 (5): 845–856. <https://doi.org/10.1037/0022-3514.91.5.845>.
- Isik, Leyla, Kami Koldewyn, David Beeler, and Nancy Kanwisher. 2017. "Perceiving Social Interactions in the Posterior Superior Temporal Sulcus." *Proceedings of the National Academy of Sciences of the United States of America* 114 (43): E9145–E9152. <https://doi.org/10.1073/pnas.1714471114>.
- Jalbrzikowski, Maria, Bart Larsen, Michael N. Hallquist, William Foran, Finnegan Calabro, and Beatriz Luna. 2017. "Development of White Matter Microstructure and Intrinsic Functional Connectivity between the Amygdala and Ventromedial Prefrontal Cortex: Associations with Anxiety and Depression." *Biological Psychiatry* 82 (7): 511–521. <https://doi.org/10.1016/J.BIOPSYCH.2017.01.008>.
- Knight, Tristan, Thomas Steeves, Lundy Day, Mark Lowerison, Nathalie Jette, and Tamara Pringsheim. 2012. "Prevalence of Tic Disorders: A Systematic Review and Meta-Analysis." *Pediatric Neurology* 47 (2): 77–90. <https://doi.org/10.1016/J.PEDIATRNEUROL.2012.05.002>.
- Kovács, András Bálint. 2014. "Shot Scale Distribution: An Authorial Fingerprint or a Cognitive Pattern?" *Projections* 8 (2): 50–70. <https://doi.org/10.3167/proj.2014.080204>.

- Kovács, András Bálint, and Nóra Fanny Zentay. 2017. "The Use of Quantitative Methods in the Humanities." In *A History of Cinema without Names, Vol. 2: Contexts and Practical Applications*, ed. Diego Cavallotti, Simone Dotto, and Leonardo Quaresima, 45–48. Milan: Mimesis.
- Larsson, Jonas, and David J. Heeger. 2006. "Two Retinotopic Visual Areas in Human Lateral Occipital Cortex." *Journal of Neuroscience* 26 (51): 13128–13142. <https://doi.org/10.1523/JNEUROSCI.1657-06.2006>.
- Lieberman, Nira, and Yaacov Trope. 2008. "The Psychology of Transcending the Here and Now." *Science* 322 (5905): 1201–1205. <https://doi.org/10.1126/science.1161958>.
- Loewenfeld, Irene E., and Otto Lowenstein. 1993. *The Pupil: Anatomy, Physiology, and Clinical Applications*, Vol. 2. Hoboken, NJ: Wiley-Blackwell.
- McDougal, David H., and Paul D. Gamin. 2015. "Autonomic Control of the Eye." *Comprehensive Physiology* 5 (1): 439–473. <https://doi.org/10.1002/cphy.c140014>.
- Milner, A. David. 2017. "How Do the Two Visual Streams Interact with Each Other?" *Experimental Brain Research* 235 (5): 1297–1308. <https://doi.org/10.1007/s00221-017-4917-4>.
- Palmer, Frederick. 1920. *Palmer Plan Handbook: Photoplay Writing Simplified and Explained*. Los Angeles: Palmer Photoplay Corporation.
- Poldrack, Russell A. 2011. "Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding." *Neuron* 72 (5): 692–697. <https://doi.org/10.1016/J.NEURON.2011.11.001>.
- Press, William A., Alyssa A. Brewer, Robert F. Dougherty, Alex R. Wade, and Brian A. Wandell. 2001. "Visual Areas and Spatial Summation in Human Visual Cortex." *Vision Research* 41 (10–11): 1321–1332. [https://doi.org/10.1016/S0042-6989\(01\)00074-8](https://doi.org/10.1016/S0042-6989(01)00074-8).
- Raz, Gal, Yael Jacob, Tal Gonen, Yonatan Winetraub, Tamar Flash, Eyal Soreq, and Talma Hendler. 2014. "Cry for Her or Cry with Her: Context-Dependent Dissociation of Two Modes of Cinematic Empathy Reflected in Network Cohesion Dynamics." *Social Cognitive and Affective Neuroscience* 9 (1): 30–38. <https://doi.org/10.1093/scan/nst052>.
- Raz, Gal, Lavi Shpigelman, Yael Jacob, Tal Gonen, Yoav Benjamini, and Talma Hendler. 2016. "Psychophysiological Whole-Brain Network Clustering Based on Connectivity Dynamics Analysis in Naturalistic Conditions." *Human Brain Mapping* 37 (12): 4654–4672. <https://doi.org/10.1002/hbm.23335>.
- Raz, Gal, Michele Svanera, Neomi Singer, Gadi Gilam, Maya Bleich Cohen, Tamar Lin, . . . and Roe Admon. 2017. "Robust Inter-Subject Audiovisual Decoding in Functional Magnetic Resonance Imaging Using High-Dimensional Regression." *NeuroImage* 163: 244–263. <https://doi.org/10.1016/J.NEUROIMAGE.2017.09.032>.
- Raz, Gal, Alexandra Touroutoglou, Christine Wilson-Mendenhall, Gadi Gilam, Tamar Lin, Tal Gonen, . . . and Yael Jacob. 2016. "Functional Connectivity Dynamics during Film Viewing Reveal Common Networks for Different Emotional Experiences." *Cognitive, Affective, & Behavioral Neuroscience*, 16 (4): 709–723. <https://doi.org/10.3758/s13415-016-0425-4>.
- Raz, Gal, Yonatan Winetraub, Yael Jacob, Sivan Kinreich, Adi Maron-Katz, Galit Shaham, . . . and Ilana Podlipsky. 2012. "Portraying Emotions at their Unfolding: A Multilayered Approach for Probing Dynamics of Neural Networks." *NeuroImage* 60 (2): 1448–1461. <https://doi.org/10.1016/j.neuroimage.2011.12.084>.
- Riegel, Alois. 1901. *Die Spätromische Kunstindustrie Nach Den Funden in Österreich-Ungarn*. Vienna: Verlag der Kaiserlich-Königlichen Hof- und Staatsdruckerei.
- Roberts, Daniel J., Anna M. Woollams, Esther Kim, Pelagie M. Beeson, Steven Z. Rapcsak, and Matthew A. Lambon Ralph. 2013. "Efficient Visual Object and Word Recognition Relies on High Spatial Frequency Coding in the Left Posterior Fusiform Gyrus: Evidence from a Case-Series of Patients with Ventral Occipito-Temporal Cortex Damage." *Cerebral Cortex* 23 (11): 2568–2580. <https://doi.org/10.1093/cercor/bhs224>.

- Rooney, Brendan, and Katalin E. Bálint. 2018. "Corrigendum: Watching More Closely: Shot Scale Affects Film Viewers' Theory of Mind Tendency but not Ability." *Frontiers in Psychology* 9: 261. <https://doi.org/10.3389/fpsyg.2018.00261>.
- Rosenke, Mona, Kevin S. Weiner, Michael A. Barnett, Karl Zilles, Katrin Amunts, Rainer Goebel, and Kalanit Grill-Spector. 2018. "A Cross-Validated Cytoarchitectonic Atlas of the Human Ventral Visual Stream." *NeuroImage* 170: 257–270. <https://doi.org/10.1016/j.NEUROIMAGE.2017.02.040>.
- Schrader, Paul. 2014. "The Close-Up." *Film Comment* 50 (5): 58–61. <https://www.filmcomment.com/article/the-close-up-films-that-changed-filmmaking/>.
- Schultz, Johannes, Karl J. Friston, John O'Doherty, Daniel M. Wolpert, and Chris D. Frith. 2005. "Activation in Posterior Superior Temporal Sulcus Parallels Parameter Inducing the Percept of Animacy." *Neuron* 45 (4): 625–635. <https://doi.org/10.1016/j.neuron.2004.12.052>.
- Sewards, Terence V. 2011. "Neural Structures and Mechanisms Involved in Scene Recognition: A Review and Interpretation." *Neuropsychologia* 49 (3): 277–298. <https://doi.org/10.1016/j.neuropsychologia.2010.11.018>.
- Sperandio, Irene, and Philippe A. Chouinard. 2015. "The Mechanisms of Size Constancy." *Multisensory Research* 28 (3–4): 253–283. <https://doi.org/10.1163/22134808-00002483>.
- Sperandio, Irene, Philippe A. Chouinard, and Melvyn A. Goodale. 2012. "Retinotopic Activity in V1 Reflects the Perceived and Not the Retinal Size of an Afterimage." *Nature Neuroscience* 15 (4): 540–542. <https://doi.org/10.1038/nn.3069>.
- Studebaker, Gerald A. 1985. "A 'Rationalized' Arcsine Transform." *Journal of Speech Language and Hearing Research* 28 (3): 455–462. <https://doi.org/10.1044/jshr.2803.455>.
- Szaflarski, Jerzy P., Mekibib Altaye, Akila Rajagopal, Kenneth Eaton, Xiangxiang Meng, Elena Plante, and Scott K. Holland. 2012. "A 10-Year Longitudinal fMRI Study of Narrative Comprehension in Children and Adolescents." *NeuroImage* 63 (3): 1188–1195. <https://doi.org/10.1016/j.NEUROIMAGE.2012.08.049>.
- Wang, Liang, Ryan E. Mruzek, Michael J. Arcaro, and Sabine Kastner. 2015. "Probabilistic Maps of Visual Topography in Human Cortex." *Cerebral Cortex* 25 (10): 3911–3931. <https://doi.org/10.1093/cercor/bhu277>.
- Wilson, Stephen M., Istvan Molnar-Szakacs, and Marco Iacoboni. 2008. "Beyond Superior Temporal Cortex: Intersubject Correlations in Narrative Speech Comprehension." *Cerebral Cortex* 18 (1): 230–242. <https://doi.org/10.1093/cercor/bhm049>.
- Wölfflin, Heinrich. 1921. *Kunstgeschichtliche Grundbegriffe: Das Problem Der Stilentwicklung in Der Neueren Kunst*. Munich: H. Bruckmann.
- Woodhead, Zoe Victoria Joan, Richard James Surtees Wise, Marty Sereno, and Robert Leech. 2011. "Dissociation of Sensitivity to Spatial Frequency in Word and Face Preferential Areas of the Fusiform Gyrus." *Cerebral Cortex* 21 (10): 2307–2312. <https://doi.org/10.1093/cercor/bhro08>.
- Zacks, Jeffrey M., Nicole K. Speer, Khena M. Swallow, and Corey J. Maley. 2010. "The Brain's Cutting-Room Floor: Segmentation of Narrative Cinema." *Frontiers in Human Neuroscience* 4: 168. <https://doi.org/10.3389/fnhum.2010.00168>.
- Zhu, Xiangxin, and Deva Ramanan. 2012. "Face Detection, Pose Estimation, and Landmark Localization in the Wild." In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2879–2886. <https://doi.org/10.1109/CVPR.2012.6248014>.