

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Pushing the button: Why do learners pause online videos?

Martin Merkt^{a,*}, Anett Hoppe^b, Gerrit Bruns^b, Ralph Ewerth^{b,c}, Markus Huff^{d,e}

^a German Institute for Adult Education, Bonn, Germany

^b TIB - Leibniz Information Centre for Science and Technology, Hannover, Germany

^c L3S Research Center, Leibniz University Hannover, Germany

^d Leibniz-Institut für Wissensmedien, Tübingen, Germany

^e Department of Psychology, Eberhard Karls Universität Tübingen, Germany

ARTICLE INFO

KEYWORDS:

Online videos
Pauses
Logfile analysis
Video difficulty
Video structure

ABSTRACT

With the recent surge in digitalization across all levels of education, online video platforms gained educational relevance. Therefore, optimizing such platforms in line with learners' actual needs should be considered a priority for scientists and educators alike. In this project, we triangulate logfiles of a large German online video platform for educational videos with behavioral data from a laboratory study and the objective characteristics of the selected videos. We aim to understand the potential motives for why participants pause educational videos while watching such videos online. Our analyses revealed that perceived difficulties in comprehension and meaningful structural breakpoints in the videos were associated with increased pausing behavior. In contrast, pausing behavior was not associated with the videos' formal structural features highlighted in the video platform. Implications of these findings and the potentials of our methodological approach for theory and practice are discussed.

1. Introduction

It is hard to dispute the continuing popularity of online videos on platforms such as YouTube (Rat für kulturelle Bildung, 2019). Whereas entertainment may still be the dominant motive to watch videos online, the educational relevance of video-based learning materials has gained additional traction during the COVID-19 pandemic, with schooling, university courses, and courses in adult education centers partially relying on video-based online materials during national lockdowns (Biebeler & Schreiber, 2020; Chick et al., 2020; Wildemann & Hosenfeld, 2020; Winde et al., 2020; www.volkshochschule.de/online-lernen). Next to allowing easy access to information virtually anytime and anywhere, viewers can watch online videos at their own pace (Kay & Kletschin, 2012; Rat für kulturelle Bildung, 2019). Given this increased popularity and relevance of educational online videos, it is essential to optimize online videos to provide learners with a learning experience that best aligns with their individual needs and natural usage patterns.

Whereas there is an increasing number of research focusing on characteristics of educational videos that support learning (e.g., Beege et al., 2017; Boucheix et al., 2018; de Koning et al., 2018; Fiorella et al., 2017; Hoogerheide et al., 2016; Merkt et al., 2011, 2020; Schneider et al., 2021; Tarchi et al., 2021), the current manuscript focusses on one of the major challenges that is associated with video-based learning. In particular, videos confront learners with the task of paying continuous attention to a steady stream of new information that has to be integrated with existing knowledge structures (Ayres & Paas, 2007). It is suggested that the dynamic presentation of transient information in videos induces extraneous cognitive load that may result in cognitive overload and thus in

* Corresponding author. German Institute for Adult Education, Heinemannstr. 12-14, 53175, Bonn, Germany.
E-mail address: merkt@die-bonn.de (M. Merkt).

<https://doi.org/10.1016/j.compedu.2021.104355>

Received 31 May 2021; Received in revised form 15 October 2021; Accepted 16 October 2021

Available online 20 October 2021

0360-1315/© 2021 The Authors.

Published by Elsevier Ltd. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

suboptimal learning outcomes. Addressing this issue, the availability of a stop button (e.g., Hasler et al., 2007; Lee et al., 2020; Schwan & Riempp, 2004) as well as system-determined pauses (Hasler et al., 2007; Spanjers et al., 2012) have been shown to mitigate the harmful effects of the transient flow of information in videos; however, some may argue that system-determined pauses are superior over self-determined pauses because novice learners may not be able to find the appropriate position for stopping the videos themselves (Biard et al., 2018). To provide learners with the best possible learning experience, it could be a promising approach to provide system-determined pauses at positions where the viewers would most likely stop the videos themselves. Thus, it is worth looking at learners' motives to stop videos in a natural viewing setting to identify such positions.

Drawing from one's own experience while watching videos online, it is evident that there is a plethora of potential reasons why viewers may want to pause a video. Whereas some of the possible reasons are unrelated to the video's actual contents (e.g., external disturbances) and may thus be considered background noise in logfile analyses, there are potential reasons for pausing a video that are inherent to the video that the viewers are watching. Because logfiles do not contain information about contextual factors such as external disturbances, we focus our research on the characteristics of the video that may constitute motives for pushing the pause button. In this regard, we identified content difficulty, meaningful structural boundaries, and formal structural cues as potential reasons for pushing the pause button. The investigation of content difficulty and meaningful structure is motivated by theoretical assumptions about the mechanisms underlying the beneficial effects of pauses in videos or animations (Merkt et al., 2018; Rey et al., 2019; Spanjers et al., 2010; Spanjers et al., 2012). Please refer to Sections 1.1.1 and 1.1.2 for a more elaborate discussion of these mechanisms. Further, analyzing the formal structural features is essential to control for more superficial cues that may also affect how viewers structure a video and thus decide on when to push the pause button (see Section 1.1.3). Finally, the current study serves as a proof of concept that the triangulation of online logfiles with data of lab experiments is a suitable approach to explaining real-world behavior.

Whereas the effects of both self-determined and system-determined pauses on learning are well-investigated (e.g., Chen & Yen, 2021; Hasler et al., 2007; Rey et al., 2019; Spanjers et al., 2012), there is a lack of research on why learners actively pause educational videos in an ecologically valid setting outside of the laboratory (e.g., at home or in the library). To address this open research question, we triangulated data from three different data sources. First, we used anonymous logfiles of an online platform for scholarly videos¹ to identify the positions at which the videos were paused by the platform's actual users in their natural viewing environments. Second, we conducted a laboratory study in that we had participants analyze the same videos with regard to the videos' meaningful structure and difficulties in comprehension. Third, a computer vision algorithm was used to identify the objective structural features (i.e., cuts) of the videos highlighted in the online video platform. With these different data sources, we gained some insights into potential reasons why viewers push the pause button while watching educational videos online.

1.1. Learners' potential motives to push the pause button in educational videos

In learning analytics, logfiles of online learning platforms are used to shed light on self-regulated learning processes (Roll & Winne, 2015; Seo et al., 2021). Whereas this approach may be used to identify associations of the use of interactive features in online platforms and the learning outcomes (Li et al., 2020), the current manuscript focuses on viewers' motives to use interactive features by investigating relationships between the features of the learning materials and the viewers' use of the interactive features (i.e., the pause button) provided in an educational online video platform. In this regard, we focus on three potential reasons for viewers to push the pause button when watching educational videos online: (1) difficulty, (2) meaningful content structure, and (3) formal structural cues. Subsequently, we will outline the rationale for selecting these characteristics for the current study, highlighting the characteristics' hypothesized relevance for the learning process.

1.1.1. Difficulty

Both the Cognitive Load Theory (CLT, Sweller et al., 1998, 2019) and the Cognitive Theory of Multimedia Learning (CTML, Mayer, 2021) postulate that limited working memory resources constrain learning processes. Thus, learning materials should be designed to make optimal use of these limited resources. According to CLT (Sweller et al., 1998, 2019), the cognitive demands associated with the learning materials may stem from three different sources, referred to as intrinsic cognitive load, extraneous cognitive load, and germane cognitive load. Intrinsic cognitive load is related to the inherent complexity of the learning materials, extraneous cognitive load is associated with the processing required due to (suboptimal) instructional design, and germane cognitive load is related to learners' deliberate elaborations of the learning materials. It is widely agreed that the sum of intrinsic and extraneous cognitive load should be kept at a minimum to free cognitive capacities for deliberate elaboration processes (i.e., germane cognitive load).

Regarding learning with dynamic audiovisual media, the transience of these kinds of materials has been identified as one characteristic that may induce extraneous cognitive load that may prevent a deeper elaboration of the learning materials (e.g., Ayres & Paas, 2007; Merkt et al., 2018). In line with this reasoning, it is often argued that the opportunity to take pauses in the learning process is beneficial for learning (Lee et al., 2021; Rey et al., 2019). Consequently, increased use of interactive features that allowed for adjusting the pace of a video presentation (i.e., fast forward and rewind) was shown for more complex learning materials (Schwan & Riempp, 2004). Further, regarding pauses, various studies demonstrated that the provision of system-determined pauses or the availability of a pause button can support learning (Hasler et al., 2007; Lee et al., 2020).

¹ <https://av.tib.eu>.

More specifically, a recent study has shown that taking self-determined pauses in a complex simulation-based learning environment temporarily reduced the learners' cognitive load during the pauses (Lee et al., 2020), even though learners' overall cognitive load was increased by the availability of a pause button. Whereas we will come back to the latter observation in Section 1.1.2, the observation that cognitive load is reduced during pauses supports the assumption that pauses help learners avoid cognitive overload. Further support for this assumption may be gathered from findings that pauses are most beneficial for complex learning materials (Hasler et al., 2007), for learners with low prior knowledge (Spanjers et al., 2011), and learners with low working memory capacities (Lusk et al., 2009). To sum up, there is ample evidence that the positive effects of pauses are associated with reducing cognitive task demands and thus the avoidance of cognitive overload. Consequently, it is feasible to assume that learners actively use the pause button when the contents of a video are too difficult to understand in a system-paced presentation.

1.1.2. Meaningful structure

Some studies demonstrated that the availability of a pause button might benefit learning, even though learners did not even make extensive use of the pause button or did not pause the video at all (Hasler et al., 2007; Lee et al., 2020). Whereas these findings cannot be explained with the avoidance of cognitive overload by pausing the video, the more so because the availability of a pause button increased overall cognitive load in a simulation-based learning environment (see Lee et al., 2020), they are well in line with the explanation that the availability of a pause button triggers learners to monitor the video for appropriate positions where to pause the video (see Hasler et al., 2007). In other words, viewers may monitor the video for meaningful structural breakpoints at which the video can be stopped without disrupting the video within meaningful units.

The relevance of highlighting the structure of educational videos by pauses has been recently discussed as one potential mechanism underlying the beneficial effects of system-determined pauses (i.e., structuring explanation; Merkt et al., 2018; Rey et al., 2019; Spanjers et al., 2012). This explanation is based on the Event Segmentation Theory (EST, Zacks & Swallow, 2007) and the Event Horizon Model (EHM, Radvansky, 2012), which address humans' perception of continuous events. Whereas a detailed description of EST and EHM is beyond the scope of this manuscript, it is essential to note that both models share the assumption that humans automatically segment ongoing events into discrete and meaningful units. In particular, there is robust evidence that recipients perceive a new situation when various dimensions of a narrative (e.g., time, space) change concurrently (Magliano et al., 2001; Zwaan et al., 1995). Even though there is evidence that grasping the meaningful structure of events is associated with better memory for events (Sargent et al., 2013), it has not yet been investigated whether learners' self-determined pauses in educational videos coincide with meaningful structural breakpoints. The current study will address this research gap.

1.1.3. Formal structural cues

Next to meaningful semantic boundaries determined by a video's content structure, the perceived structure of a video may also be shaped by its formal features. In particular, previous studies used short pauses or cuts to highlight the meaningful semantic structure of dynamic presentations (Merkt et al., 2018; Spanjers et al., 2012). Whereas instructional designers may use external structuring cues such as cuts to highlight the semantic structure of the learning materials, these formal structural features of videos do not necessarily coincide with viewers' perceptions of the meaningful structure. For example, film cuts did not result in increased segmentation behavior when participants were asked to identify the meaningful structure of a video (Schwan et al., 2000).

Even though the formal structure of a video may not coincide with the meaningful structure of the contents, it may still have implications for cognitive processes. In particular, the structural division of learning materials that were not connected in a meaningful semantic structure (i.e., word lists) affected learners' memory, with more fine-grained divisions of the word lists across multiple windows on a computer screen, resulting in better recall (Pettijohn et al., 2016). Hence, it is feasible to assume that formal structural features of an online video platform may affect learners' cognitive processing independent of the videos' actual meaningful content structure that is perceived by the viewers. Admittedly, the division of word lists across multiple windows on a computer screen (see Pettijohn et al., 2016) may be considered a less subtle means of providing learners with formal structural cues than film cuts (see Schwan et al., 2000).

Next to film cuts, instructional designers may use other structural cues that imply the structure of a video. For example, a division of the timeline that is displayed below the video may serve as an external structural cue that highlights the structure of the learning materials. Even though these divisions of the timeline are often designed to reflect the chapters of a video (see Cojean & Jamet, 2017; Merkt et al., 2011; Merkt & Schwan, 2014), they do not necessarily reflect the majority of the learners' perceptions of a video's structure. More importantly, on the online video platform that was used to gather data for the current study, the timeline below the video was divided based on automatic cut detection (see Section 2.2.3.), so that the division of the timeline did not necessarily reflect the videos' content structure, but rather the editing of the videos.

To the best of our knowledge, there is no study that investigated whether providing learners with an external structure for the video actually triggers them to push the pause button at the externally provided structural breakpoints. However, previous research has demonstrated that structuring learning materials may have cognitive effects on learners beyond simply providing cues for the meaningful structure of the materials (see Pettijohn et al., 2016). Therefore, it is an interesting research question whether external structural cues also serve as an orientation for learners when deciding to push the pause button in online videos.

1.2. The current study: research questions

In the current study, we triangulated logfiles from an online video platform, data from one laboratory study, and the results of automated video analyses to explore viewers' potential motives for pushing the pause button while watching educational videos

online. In this regard, the previous sections identified three potential reasons why viewers may decide to push the pause button while watching educational videos online. First, based on the reasoning of the cognitive load theory (Sweller et al., 1998, 2019), it is feasible to assume that viewers push the pause button whenever difficulties in comprehending the videos occur. Second, based on the reasoning in empirical articles arguing that including a pause button benefits learning because it prompts learners to scan the video for suitable breakpoints (Hasler et al., 2007; Lee et al., 2020), it is feasible to assume that viewers pause videos at meaningful structural breakpoints. Finally, providing viewers with external structural cues (e.g., division of a video's timeline) may prompt them to stop the video at the breakpoints that are provided on the online video platform.

To wrap up, the current study serves to test various potential motives for the use of the pause button in an ecologically valid viewing environment. As such, it may be considered important groundwork for understanding how online video platforms may be tailored to viewers' individual needs. Even though our research questions and assumptions are well-grounded in theory, we refrain from formulating definite hypotheses because the study was initially designed with an explorative character that shall not be misrepresented in this manuscript.

2. Method

2.1. Materials

For the current analysis, we selected four videos that covered different physical topics. These videos were selected from an online video platform that provides more than 30,000 scientific videos. The platform targets academic users such as researchers and university students and mainly addresses the fields of natural sciences, engineering, and architecture, but also provides videos for other subject areas. Metadata are automatically generated for the videos belonging to the main subject areas, using algorithms that extract annotations, transcriptions, and a structural overview based on cut detection. The results of these automatic analyses are displayed as vertical lines in the timeline (at positions of detected cuts) and on the right-hand side of the video in tabs for annotations and transcripts, which are divided according to the identified cuts. The visual division of the timeline is only shortly visible at the beginning of the video or on mouseover, whereas the annotations and transcripts that are grouped according to the identified structure are only visible when viewers select the tab "annotations" or "transcript" (see Fig. 1). The platform is typically used in informal learning scenarios and not part of formal education.

The selection of the videos for the study was based on three main criteria for inclusion. Most importantly, only videos belonging to the platform's key subject areas were eligible for selection because only these videos are automatically analyzed. Second, the videos had to attract a sufficient number of viewers interacting with the videos. Therefore, only the more popular videos on the platform were eligible for selection. In particular, the videos included in our analyses attracted between 353 and 687 visitors in the period in that the logfiles were collected. Third, the videos had to contain an accumulation of pauses in specific sections of the videos. Such accumulations could be observed by visual inspection using graphs produced in the R package SegMag (Papenmeier & Sering, 2016). SegMag allows for a visual inspection of time series data resulting from user interactions. This is done by fitting Gaussian distribution functions around each interaction time point of each participant. Thus, user interactions with high temporal alignment result in significantly higher peaks than user interaction with low temporal alignment (for an application of this procedure for videos, see Meitz et al., 2020). The four videos were selected based on visual inspection from a preselected pool of ten videos that fulfilled the first two criteria mentioned above and additionally showed local accumulations of pausing behavior in the graphs produced using SegMag. These selection criteria for the videos reflect the rationale for using videos including positions with an increased likelihood to push the pause button. The aim of this project was to explain this increased likelihood of pausing behavior by triangulating the logfiles with data produced in a laboratory study. The four selected videos were between 147 and 392 s long. Next to the visual information, all videos included auditory information such as speech and ambient noise. In particular, the videos covered the streaming behavior of water (<https://av.tib.eu/media/10886>), rotational force (<https://av.tib.eu/media/12493>), adsorption (<https://av.tib.eu/media/15686>), and absorption (<https://av.tib.eu/media/15712>). They were produced as explanatory videos by domain experts.

2.2. Data sources

For our analyses, we gathered data from three different sources: logfiles from the online video platform, segmentation data from a laboratory study, and the results of an automatic analysis of the videos. Subsequently, we give an overview of these different data sources. All data and analyses scripts (for the statistical programming language R and SPSS) have been made publicly available via the Open Science Framework.²

2.2.1. Logfiles

The logfiles reflect the natural pausing behavior of visitors watching the videos on the online platform between June 20th, 2018 and December 31st, 2018. Out of an overall of 2114 visitors that watched the videos on the video platform, there was data from 626 unique IDs that paused one of the videos at least once, with 104 to 195 unique user IDs per video. Please note that logfiles were extracted only for users with at least one interaction (i.e., pausing) so that these numbers do not reflect the overall number of viewers of

² <https://osf.io/nd54f/>.

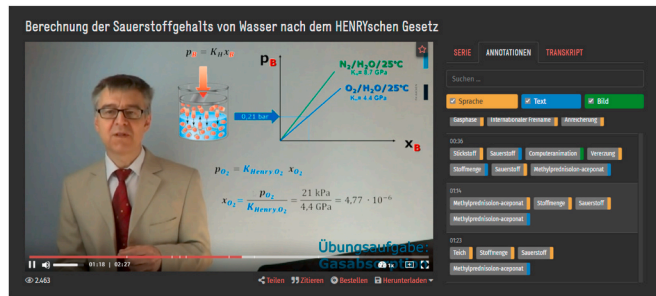


Fig. 1. Screenshot of the online video platform (video about absorption). Note. The screenshot includes the division of the timeline (vertical lines) and the division of the annotations (right hand side of the video) highlighting the structure of the video according to a cut detection algorithm. The timeline is only visible at the beginning of the video and on mouseover, whereas viewers actively have to select the tab annotations in order to see them.

the videos (i.e., 2114 visitors), but rather the number of users that interacted with the video by pushing the pause button (i.e., 626 visitors). In the video with the least unique IDs, 104 viewers pushed the pause button at least once, whereas, in the video with the most unique IDs, 195 viewers pushed the pause button at least once. There was a list of time stamps at which the corresponding visitor had paused the video on the online platform for each anonymous user ID. These time stamps were automatically logged on the online video platform. Across all videos selected for the analyses, the number of pauses ranged between 1 and 51 per ID ($M = 3.37$, $SD = 5.08$). Because the logfiles reflect natural usage behavior in an uncontrolled, real-world setting, it could not be controlled that each user watched each of the four videos and that the users finished watching each video. However, it is essential to highlight that not the users but the videos (divided into 1-s bins) constituted the unit of analysis in this study (see Section 2.3 for a more comprehensive overview of the analyses).

2.2.2. Segmentation data

We collected the segmentation data in our laboratory in a study with 24 participants ($M_{\text{age}} = 22.29$, $SD_{\text{age}} = 4.80$; 19 female, 5 male) studying a wide variety of different subjects at the local university (e.g., economics, nutrition science, agricultural science). We had intentionally excluded students studying physics when recruiting participants using the institute's research management database. After giving informed consent to participate in the experiment, each participant was asked to watch each of the four videos twice. When gathering the segmentation data, we did not provide learners with external structural cues such as the division of the timeline so that the segmentation data was not affected by these cues. In the first pass, participants were asked to pay attention to video difficulty and press the spacebar whenever they had difficulties understanding the contents of the video. In the second pass, participants were instructed to press the spacebar whenever they perceived the end of a meaningful unit and the beginning of a new meaningful unit (see Newton, 1973). The order of the four videos was randomized across participants; however, the two passes for each video were always presented as a unit. Having participants watch each video twice is in line with previous segmentation studies that confronted learners with the same materials twice, either with different task instructions (e.g., Zacks et al., 2001) or with a familiarization phase in that the participants viewed the experimental materials without a segmenting instruction (e.g., Mura et al., 2013). In our study, watching the video with the difficulty instruction may serve as a familiarization phase for watching the video with the instruction to segment the video into meaningful units. After watching the final video, participants were debriefed and received financial compensation for participating in the experiment. The entire experiment was run in PsychoPy (Peirce et al., 2019) and lasted no longer than 60 min per participant. The local ethics committee approved the procedure of the laboratory study.

2.2.3. (Automatic) video analysis

On the video platform, a shot boundary detection algorithm is applied to automatically provide a structural overview of the videos that are uploaded to the online platform. This algorithm analyzes the videos with regard to abrupt camera shot changes (see Hentschel et al., 2013). The detected structure of the video is reflected in two ways on the video platform: (a) The timeline below the video displays vertical lines to show the temporal position of detected cuts; (b) Further automatic annotations and transcripts on the right-hand side of the video are grouped based on the detected shots that they belong to. These kinds of annotations consist of transcribed audio information using automatic speech recognition (Milde & Köhn, 2018; Povey et al., 2011), recognized text in the video frames (optical character recognition³), as well as results of visual concept classification, (e.g., objects, places) (Springstein & Ewerth, 2016).

To evaluate the accuracy of the cut detection algorithm (see Section 3.1.3.), we had two human raters watch the videos in order to identify cuts. In line with previous research, a cut was defined as an abrupt content change in the entire video frame (Cutting et al., 2011; Smith & Henderson, 2008). Both raters provided Excel sheets with time stamps of the positions in that they had identified cuts. To analyze inter-rater-agreement, all of these time stamps were listed in a table in that the four videos were divided into 1-s bins,

³ <https://github.com/tesseract-ocr>.

resulting in 1040 bins. Overall, there were 40 bins in that at least one of the two raters had identified a cut. Out of these bins, the raters agreed on 35 bins that there was a cut (87,5 %). Additionally, there were 1000 1-s bins in the videos in that none of the two raters identified a cut, resulting in an excellent overall inter-rater-agreement (Cohen's Kappa = .93). For the 5 bins that included conflicting ratings, the first author resolved conflict by visual inspection of the videos. Overall, 1 of the 5 discrepancies resulted from one rater missing a cut, whereas 4 out of the 5 discrepancies resulted from one rater falsely identifying camera movement or short blips in the video as a cut. To sum up, based on human inspection, there were 36 cuts across the four different videos.

2.3. Analysis

For our analyses, we triangulated data from the three different sources described above in order to gain some insights into motives that may trigger learners to push the pause button when watching videos online. Importantly, the logfiles from the online video platform, the segmentation data from the laboratory, and the cut detection data from the automatic analysis of the video do not stem from the same people (i.e., participants). Consequently, we used the division of the videos into 1-s bins as the unit of analysis. For each bin, it was counted how many viewers pushed the pause button (source: logfiles), how many participants perceived a difficulty (source: segmentation data), how many participants perceived a meaningful boundary (source: segmentation data), and whether there was a structural breakpoint indicated on the online video platform (source: automatic analysis, 0 = no cut, 1 = cut). Based on this data, we conducted a regression analysis with the number of pauses per bin as the dependent variable and the three other variables (i.e., difficulty, meaningful boundary, and cut) as the independent variables. To increase the power of our analyses, we merged the data for the four videos into a single dataset for analysis because we did not expect any differences between different types of videos. Further, we make predictions about general associations between different variables that should be independent of a video's actual content domain and be applicable independent of the actual contents, as long as these contents are characterized by the same levels of difficulty and the same levels of meaningfulness. In total, there were 1040 bins in the four videos.

3. Results

3.1. Descriptive data

3.1.1. Logfiles

Overall, there was a total of 2083 presses of the pause-button for the four videos, with the numbers for the individual videos ranging between 1 and 51 per user ID. These key presses were recorded in 818 of the 1040 bins (78.65 %) and were generated by 626 unique users, ranging between 104 and 195 unique users for the individual videos.

3.1.2. Segmentation data

For segmentation data analysis, the videos were split into 1-s bins (total: 1040 bins). Overall, difficulties were noted by at least one participant in 349 bins (33.56 %), whereas meaningful structural breakpoints were identified by at least one participant in 286 bins (27.50 %). Finally, both difficulties and meaningful structural breakpoints were identified by at least one participant in 122 bins (11.73 %), whereas it is important to note that the difficulty and the meaningful structural breakpoint could have been identified by different participants.

3.1.3. Automatic video analysis

Overall, the automatic cut detection identified 45 cuts in the four videos, with the number of cuts per individual video ranging from 7 to 16. In contrast, human inspection of the videos revealed 36 actual cuts (see Section 2.2.3). Out of 36 actual cuts in the videos, 35 were correctly detected by the cut detection algorithm (i.e. detection rate is 97.2 %). In addition, there were 10 false positive detections, (i.e. precision is 77.8 %). The F1 score (harmonic mean of recall and precision) was 86.4 %. Nearly all false positive detections were caused by abrupt changes in smaller or larger parts in the frames, for example when a diagram or text was abruptly inserted and displayed, whereas we had defined a cut as an abrupt content change in the entire video frame. Because the provided structure on the video platform reflects the results of the automatic cut detection, we decided to use this data for the regression analysis.

3.2. Regression analysis

The multiple linear regression analysis with pausing behavior in the online video platform as the dependent variable revealed a significant overall regression model, $F(3,1036) = 4.72, p = .003, R^2 = 0.01$. Regarding the individual predictors, there were associations between perceived difficulty and natural pausing behavior, $\beta = 0.09, p = .006$, and between perceived meaningful breakpoints and natural pausing behavior, $\beta = 0.07, p = .042$. In contrast, there was no association between the external structure provided in the online video platform and natural pausing behavior, $\beta = -0.02, p = .631$. There was a significant intercept, $p < .001$.

4. Discussion

Triangulating data from three different sources, the current study is the first to investigate viewers' potential motives for pushing the pause button while watching educational videos online. Most importantly, we found some initial evidence that both the perception of difficulties in comprehension as well as the perception of meaningful breakpoints may constitute motives for viewers to push the

pause button, whereas more formal structural features such as the division of the timeline based on the detection of cuts did not constitute motives to pause the video. Hence, it may be concluded that viewers' decisions to push the pause button are more strongly linked to the contents of the videos rather than to formal structural features. This latter finding nicely extends the findings of [Schwan et al. \(2000\)](#) who also found no effect of formal structural features (i.e., film cuts) on participants' segmentation behavior, even though more explicit ways of structuring the videos (i.e., division of the timeline and overviews next to the video) were implemented on the video platform that was used as a data source in the current experiment.

Notably, the current study provides some evidence that the explanations for the effectiveness of pausing in educational videos (i.e., transience explanation and structuring explanation) do not only apply to a laboratory setting. In particular, based on the current analysis, we can assume that viewers of educational online videos are more likely to push the pause button when videos turn difficult and when there is a meaningful structural breakpoint. These findings provide further evidence for the relevance of both the transience explanation and the structuring explanation that are often used to explain the benefits of pauses (see [Merkt et al., 2018](#); [Spanjers et al., 2012](#)). According to the transience explanation, adding pauses or a pause button to the video supports learning because pausing the video may prevent cognitive overload if the complexity of the video increases. According to the structuring explanation, system-determined pauses support learning because they may structure the learning material (e.g., [Spanjers et al., 2012](#)), whereas a pause button should support learning because viewers monitor the video for appropriate positions where to stop the video (e.g., [Hasler et al., 2007](#); [Lee et al., 2020](#)).

In contrast to the content-related features investigated in this study, there was no association between the external structuring cues (i.e., division of the timeline and annotations) and the participants' pausing behavior. This may, to some extent, reflect the fact that these divisions are only shown briefly at the beginning of each video (i.e., division of the timeline) or on-demand (i.e., division of the annotations; mouseover to re-activate the timeline). On the other hand, this division reflects the automatic analysis of the videos, which is based on the detection of cuts that constitute abrupt changes of the visual information presented in two consecutive frames. However, based on the current data, there is no evidence that such structural cues that are independent of the videos' contents are associated with viewers' pausing behavior in educational online videos.

Because our analyses rely on correlational data, we refrain from making any causal statements at this point. Nevertheless, the current study provides some first insights into viewers' potential motivations for taking pauses online. Building on these insights, future research could take two different approaches to gain causal evidence, both with their strengths and caveats. First, researchers could insert automated questions into an online video platform, asking learners why they took a pause directly after taking a pause. This type of questioning may allow insights into learners' motives for taking pauses while watching videos online. However, these questions interrupt the natural viewing process and may thus affect viewers' subsequent pausing behavior so that such data may be more artificial than the logfiles used for the current analysis. To overcome this caveat, researchers may use cued retrospective reports in that viewers are provided with a screencast of their viewing behavior and asked why they paused videos after watching the entire video online. However, an approach including the recording of screencasts would come with challenges regarding users' privacy and may thus not be compatible with collecting usage data in a scenario where the user does not feel observed. Second, researchers could systematically vary the characteristics of videos (i.e., difficulty, meaningful structure) and investigate whether viewers' use of the pause button increases at the predefined positions in the videos. However, whereas this method may be less intrusive than questioning viewers after pushing the pause button, it also does not come without its caveats. In particular, if implemented in a naturalistic viewing setting, this method requires the manipulated videos to gain attention from a sufficiently large number of viewers. Importantly, these viewers should not just watch the video but also use the pause button. Given that a substantial number of users does not use the pause button even under highly controlled settings in the laboratory ([Biard et al., 2018](#); [Hasler et al., 2007](#); [Lee et al., 2020](#)), collecting sufficient data under natural viewing conditions may be challenging, even if the videos follow design characteristics of popular videos on platforms such as YouTube (see [ten Hove & van der Meij, 2015](#)).

Another limitation of the current study, which may be mainly attributed to the collection of anonymous logfiles in a natural, uncontrolled viewing scenario, is that the observed effects are rather small and only explain a limited amount of variance regarding viewers' potential motivations to push the pause button when watching educational videos online. Nevertheless, the study has its merit in pointing out that viewers may be more inclined to push the pause button by content features rather than by structural elements of the video or the video platform, at the same time leaving room for further research that tries to dive deeper into viewers' motives to pause educational online videos. For example, additional motives may be associated with causes in viewers' environment (e.g., external disturbances), with viewers' individual characteristics (e.g., prior knowledge, working memory capacity, interest) or motivations to watch the videos (e.g., learning vs. entertainment), learning context (formal or informal), and with other characteristics of the videos that were beyond the scope of this manuscript (e.g., changes in presentation format). Building on our current findings using comparable methodology, it is a promising pathway for future research to dive deeper into viewers' potential motives for pushing the pause button while watching educational videos online. Maybe, the previous section is inspiring in identifying additional factors that explain variance in viewers' natural pausing behavior.

Despite its limitations, the current manuscript demonstrated that it is possible to use logfiles of online platforms to test hypotheses that are derived from theoretical frameworks. In this regard, the current study provides evidence for the generalizability of the hypothesized associations from laboratory settings to ecologically valid settings because the viewers' actual pausing behavior was collected when "real" visitors of an online video platform watched educational videos online, without feeling observed by the researchers. Thus, the use of logfiles, as advocated in the context of learning analytics (e.g., [Roll & Winne, 2015](#)), may be considered a fruitful approach to demonstrate the ecological validity of theories that are often tested in the laboratory or in settings in that participants know that they participate in a scientific study. The triangulation of logfiles with other data sources can support research in gathering evidence using unobtrusive measures that reflect learners' actual behavior. We hope that the current work provides a

convincing proof of concept that inspires researchers to use comparable methodology in order to test the generalizability of scientific theories in practice.

Next to these methodological implications, these findings may contribute to the provision of educational videos that are tailored to the learners' needs while watching videos, for example by providing system-determined pauses at positions that would be considered natural stopping points for the majority of viewers. These pauses could either be used to avoid cognitive overload for individual learners depending on video difficulty or to display comprehension questions or reflection prompts that elicit more active processing of the videos (Cheon et al., 2014; Roediger & Karpicke, 2006; Wouters et al., 2007). Further, if pauses at specific positions in videos can be identified as indicators for difficulties in comprehension, such information could also be used in predictive learning analytics dashboards that have already been subject to investigation in other contexts (Herodotou et al., 2021). Therefore, the current study may be considered as laying some important groundwork for the development of online video platforms that are tailored to viewers' needs and that are more engaging in terms of fostering active processing of the video-based learning materials.

Admittedly, the current approach is mostly suitable to identify potential motives to push the pause button for "the average user", whereas tailoring online video platforms to each learners' individual needs comes with challenges both on an assessment and on a privacy level. First, research has to develop efficient methods to assess learners' prior knowledge or working memory span, which may in turn be used to adapt the system-determined pauses to the users. However, such an individual adaptation requires tracking the users by means of cookies or individual user profiles. Whereas such tracking may be easily implemented technologically, tracking individual users comes with issues regarding privacy protection that should be carefully addressed before engaging in such endeavors.

5. Conclusion

In conclusion, the current study may be considered a proof of concept that it is possible to explain at least some variance in users' natural viewing behavior of educational online videos by triangulating logfiles of internet platforms with laboratory data. In this regard, we observed that viewers' natural pausing behavior in online video platforms is related to theoretically derived content characteristics of the videos (i.e., difficulty and meaningful structure), whereas external structural cues such as abrupt visual changes (i.e., cuts) that are additionally highlighted by some features of the online platform (i.e., division of the timeline below the video; division of the annotations) were not associated with pausing behavior.

Credit author statement

Martin Merkt: Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review & Editing. Anett Hoppe: Software, Writing – Review & Editing. Gerrit Bruns: Data Curation, Writing – Review & Editing. Ralph Ewerth: Software, Writing – Review & Editing. Markus Huff: Conceptualization, Methodology, Formal Analysis, Data Curation, Writing – Review & Editing.

Acknowledgments

The position of Anett Hoppe was funded within the project SALIENT (Leibniz Association, K68/2017).

References

- Ayres, P., & Paas, F. (2007). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology*, 21, 695–700. <https://doi.org/10.1002/acp.1343>
- Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2017). Look into my eyes! Exploring the effect of addressing in educational videos. *Learning and Instruction*, 49, 113–120. <https://doi.org/10.1016/j.learninstruc.2017.01.004>
- Biard, N., Cojean, S., & Jamet, E. (2018). Effects of segmentation and pacing on procedural learning by video. *Computers in Human Behavior*, 89, 411–417. <https://doi.org/10.1016/j.chb.2017.12.002>
- Biebeler, H., & Schreiber, D. (2020). *Ausbildung in Zeiten von Corona. Ergebnisse einer empirischen Studie zu Auswirkungen der Corona-Pandemie auf Ausbildungsbetriebe*. Retrieved on 19.02.2021 from <https://www.bibb.de/veroeffentlichungen/de/publication/download/16929>.
- Boucheix, J.-M., Gauthier, P., Fontaine, J.-B., & Jaffeux, S. (2018). Mixed camera viewpoints improve learning medical hand procedure from video in nurse training? *Computers in Human Behavior*, 89, 418–429. <https://doi.org/10.1016/j.chb.2018.01.017>
- Chen, C.-Y., & Yen, P.-R. (2021). Learner control, segmenting, and modality effects in animated demonstrations used as the before-class instructions in the flipped classroom. *Interactive Learning Environments*, 29, 44–58. <https://doi.org/10.1080/10494820.2019.1572627>
- Cheon, J., Chung, S., Crooks, S. M., Song, J., & Kim, J. (2014). An investigation of the Effects of different types of activities during pauses in a segmented instructional animation. *Journal of Educational Technology & Society*, 17, 296–306. <http://www.jstor.org/stable/jeductechsoci.17.2.296>.
- Chick, R. C. C., Clifton, G. T., Peace, K. M., Propper, B. W., Hale, D. F., Alseidi, A. A., & Vreeland, T. J. (2020). Using technology to maintain the education of residents during the COVID-19 pandemic. *Journal of Surgical Education*, 77, 729–732. <https://doi.org/10.1016/j.jsurg.2020.03.018>
- Cojean, S., & Jamet, E. (2017). Facilitating information-seeking activity in instructional videos: The combined effects of micro- and macroscaffolding. *Computers in Human Behavior*, 74, 294–302. <https://doi.org/10.1016/j.chb.2017.04.052>
- Cutting, J. E., Brunick, K. L., & Delong, J. E. (2011). The changing poetics of the dissolve in Hollywood film. *Empirical Studies of the Arts*, 29, 149–169. <https://doi.org/10.2190/EM.29.2.b>
- Fiorella, L., Van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology*, 109, 653–665. <https://doi.org/10.1037/edu0000161>
- Hasler, B., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology*, 21, 713–729. <https://doi.org/10.1002/acp.1345>
- Hentschel, C., Blümel, I., & Sack, H. (2013). Automatic annotation of scientific video material based on visual concept detection, 1-16:8 *Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies (I-KNOW)*, 16. <https://doi.org/10.1145/2494188.2494213>, 2013.

- Herodotou, C., Maguire, C., McDowell, N., Hlosta, M., & Borooa, A. (2021). The engagement of university teachers with predictive learning analytics. *Computers & Education*, 173, Article 104285. <https://doi.org/10.1016/j.compedu.2021.104285>
- Hoogerheide, V., Van Wermeskerken, M., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction*, 44, 22–30. <https://doi.org/10.1016/j.learninstruc.2016.02.004>
- ten Hove, P., & van der Meij, H. (2015). Like it or not. What characterizes YouTube's more popular instructional videos? *Technical Communication*, 62, 48–62. <https://www.ingentaconnect.com/content/stc/tc/2015/00000062/00000001/art00005>.
- Kay, R., & Kletschin, I. (2012). Evaluating the use of problem-based video podcasts to teach mathematics in higher education. *Computers & Education*, 59, 619–627. <https://doi.org/10.1016/j.compedu.2012.03.007>
- de Koning, B., Hoogerheide, V., & Boucheix, J. M. (2018). Developments and trends in learning with instructional video. *Computers in Human Behavior*, 89, 395–398. <https://doi.org/10.1016/j.chb.2018.08.055>
- Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., & van Merriënboer, J. (2020). Different effects of pausing on cognitive load in a medical simulation game. *Computers in Human Behavior*, 110, Article 106385. <https://doi.org/10.1016/j.chb.2020.106385>
- Lee, J. Y., Szulewski, A., Young, J. Q., Donkers, J., Jarodzka, H., & van Merriënboer, J. (2021). *The medical pause: Importance, processes and training*. Medical Education. <https://doi.org/10.1111/medu.14529>
- Li, Q., Baker, R., & Warschauer, M. (2020). Using clickstream data to measure, understand, and support self-regulated learning in online courses. *The Internet and Higher Education*, 45, Article 100727. <https://doi.org/10.1016/j.iheduc.2020.100727>
- Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S., & Doolittle, P. E. (2009). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Educational Technology*, 40, 636–651. <https://doi.org/10.1111/j.1467-8535.2008.00848.x>
- Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, 15, 533–545. <https://doi.org/10.1002/acp.724>
- Mayer, R. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.
- Meitz, T. G. K., Meyerhoff, H. S., & Huff, M. (2020). Event related message processing: Perceiving and remembering changes in films with and without soundtrack. *Media Psychology*, 23, 733–763. <https://doi.org/10.1080/15213269.2019.1636660>
- Merkt, M., Ballmann, A., Felfeli, J., & Schwan, S. (2018). Pauses in educational videos: Testing the transience explanation against the structuring explanation. *Computers in Human Behavior*, 89, 399–410. <https://doi.org/10.1016/j.chb.2018.01.013>
- Merkt, M., Lux, S., Hoogerheide, V., van Gog, T., & Schwan, S. (2020). A change of scenery: Does the setting of an instructional video affect learning? *Journal of Educational Psychology*, 112, 1273–1283. <https://doi.org/10.1037/edu0000414>
- Merkt, M., & Schwan, S. (2014). Training the use of interactive videos: Effects on mastering different tasks. *Instructional Science*, 42, 421–441. <https://doi.org/10.1007/s11251-013-9287-0>
- Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction*, 21, 687–704. <https://doi.org/10.1016/j.learninstruc.2011.03.004>
- Milde, B., & Köhn, A. (2018). Open source automatic speech recognition for German. In *Speech communication; 13th ITG-symposium* (pp. 1–5). VDE.
- Mura, K., Petersen, N., Huff, M., & Ghose, T. (2013). IBES: A tool for creating instructions based on event segmentation. *Frontiers in Psychology*, 4, Article 494. <https://doi.org/10.3389/fpsyg.2013.00994>
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28(1), 28–38. <https://doi.org/10.1037/h0035584>
- Papenmeier, F., & Sering, K. (2016). *segmag: Determine event boundaries in event segmentation experiments*. R package version 1.2.4. Retrieved on 09.02.2021 from <http://CRAN.R-project.org/package=segmag>.
- Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Pettijohn, K. A., Thompson, A. N., Tamplin, A. K., Krawietz, S. A., & Radvansky, G. A. (2016). Event boundaries and memory improvement. *Cognition*, 148, 136–144. <https://doi.org/10.1016/j.cognition.2015.12.013>
- Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., & Vesely, K. (2011). The Kaldi speech recognition toolkit in, 2011. In *IEEE 2011 Workshop on automatic speech recognition and understanding*. IEEE Signal Processing Society.
- Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science*, 21, 269–272. <https://doi.org/10.1177/096372141214251274>
- Rat für kulturelle Bildung. (2019). *Jugend/YouTube/kulturelle Bildung*, 18.02.2021 https://www.rat-kulturelle-bildung.de/fileadmin/user_upload/pdf/Studie_YouTube_Webversion_final_2.pdf.
- Rey, G. D., Beege, M., Nebel, S., Wirzberger, M., Schmitt, T. H., & Schneider, S. (2019). A meta-analysis of the segmenting effect. *Educational Psychology Review*, 31, 389–419. <https://doi.org/10.1007/s10648-018-9456-4>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, 2, 7–12. <https://doi.org/10.18608/jla.2015.2.2>
- Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, 129, 241–255. <https://doi.org/10.1016/j.cognition.2013.07.002>
- Schneider, S., Kriegelstein, F., Beege, M., & Rey, G. D. (2021). The impact of video lecturers' nonverbal communication on learning – an experiment on gestures and facial expressions of pedagogical agents. *Computers & Education*, 176, Article 104350. <https://doi.org/10.1016/j.compedu.2021.104350>
- Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activity sequences? *Memory & Cognition*, 28, 214–223. <https://doi.org/10.3758/BF03213801>
- Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction*, 14, 293–305. <https://doi.org/10.1016/j.learninstruc.2004.06.005>
- Seo, K., Dodson, S., Harandi, N. M., Roberson, N., Fels, S., & Roll, I. (2021). Active learning with online video: The impact of learning context on engagement. *Computers & Education*, 165, Article 104132. <https://doi.org/10.1016/j.compedu.2021.104132>
- Smith, T., & Henderson, J. M. (2008). Edit blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research*, 2, 1–17. <https://doi.org/10.16910/jemr.2.2.6>
- Spanjers, I. E., van Gog, T., & van Merriënboer, J. G. (2010). A theoretical analysis of how segmentation of dynamic visualizations optimizes students' learning. *Educational Psychology Review*, 22, 411–423. <https://doi.org/10.1007/s10648-010-9135-6>
- Spanjers, I. E., van Gog, T., Wouters, P., & van Merriënboer, J. G. (2012). Explaining the segmentation effect in learning from animations: The role of pausing and temporal cueing. *Computers & Education*, 59, 274–280. <https://doi.org/10.1016/j.compedu.2011.12.024>
- Spanjers, I. E., Wouters, P., van Gog, T., & van Merriënboer, J. G. (2011). An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior*, 27, 46–52. <https://doi.org/10.1016/j.chb.2010.05.011>
- Springstein, M., & Ewerth, R. (2016). On the effects of spam filtering and incremental learning for web-supervised visual concept classification. In *Proceedings of ACM International Conference on Multimedia Retrieval (ICMR)* (pp. 377–380). New York: ACM. <https://doi.org/10.1145/2911996.2912072>, 2016.
- Sweller, J., van Merriënboer, J. G., & Paas, F. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296. <https://doi.org/10.1023/A:1022193728205>
- Sweller, J., van Merriënboer, J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Tarchi, C., Zaccoletti, S., & Mason, L. (2021). Learning from text, video, or subtitles: A comparative analysis. *Computers & Education*, 160, Article 104034. <https://doi.org/10.1016/j.compedu.2020.104034>

- Wildemann, A., & Hosenfeld, I. (2020). *Bundesweite Elternbefragung zu Homeschooling während der Covid 19 Pandemie: Erkenntnisse zur Umsetzung des Homeschoolings in Deutschland*. Retrieved on 18.02.2021 from https://www.zepf.eu/wp-content/uploads/2020/06/Bericht_HOMEschooling2020.pdf.
- Winde, M., Werner, S. D., Gumbmann, B., & Hieronimus, S. (2020). *Hochschulen, Corona und jetzt?*. Retrieved on 18.02.2021 from <https://www.stifterverband.org/medien/hochschulen-corona-und-jetzt>.
- Wouters, P., Tabbers, H. K., & Paas, F. (2007). Interactivity in video-based models. *Educational Psychology Review*, 19, 327–342. <https://doi.org/10.1007/s10648-007-9045-4>
- Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16, 80–84. <https://doi.org/10.1111/j.1467-8721.2007.00480.x7>
- Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, 130, 29–58. <https://doi.org/10.1037/0096-3445.130.1.29>
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6, 292–297. <https://doi.org/10.1111/j.1467-9280.1995.tb00513.x>