A database of news videos for investigating the dynamics of emotion and memory



Rosalie Samide 1 · Rose A. Cooper 1 · Maureen Ritchey 1

Published online: 16 December 2019 © The Psychonomic Society, Inc. 2019

Abstract

Emotional experiences are known to be both perceived and remembered differently from nonemotional experiences, often leading to heightened encoding of salient visual details and subjectively vivid recollection. The vast majority of previous studies have used static images to investigate how emotional event content modulates cognition, yet natural events unfold over time. Therefore, little is known about how emotion dynamically modulates continuous experience. Here we report a norming study wherein we developed a new stimulus set of 126 emotionally negative, positive, and neutral videos depicting real-life news events. Participants continuously rated the valence of each video during its presentation and judged the overall emotional intensity and valence at the end of each video. In a subsequent memory test, participants reported how vividly they could recall the video details and estimated each video's duration. We report data on the affective qualities and subjective memorability of each video. The results replicate the well-established effect that emotional experiences are remembered more vividly than nonemotional experiences. Importantly, this novel stimulus set will facilitate research into the temporal dynamics of emotional processing and memory.

Keywords Emotion · Memory · Dynamic · Videos

Everyday experiences contain a wealth of information, from perceptual details to thoughts and emotions. Cognitive psychology and neuroscience research aims to uncover the basic cognitive and neural processes supporting perception and memory of complex events. Such investigations, however, are faced with a tension between experimental control and ecological validity: in order to isolate a single process, it is often necessary to use very simple stimuli. Yet this approach misses other defining aspects of experience, such as the integration of multimodal streams of information and dynamic changes in these streams—and our reactions to them—over time. Therefore, basic research on cognitive and neural processes can benefit from complementary approaches that incorporate dynamic stimuli to improve generalizability to real-life events.

Natural experiences often include affective elements or states that vary over time. Past research on emotion processing has largely focused on one of two extremes: transient changes in affect elicited by static stimuli, such as words (e.g., Affective Norms for English Words; Bradley & Lang, 1999) or visual scenes, including the well-known International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008) and the Nencki Affective Picture System (Marchewka, Żurawski, Jednoróg, & Grabowska, 2014), or sustained changes in affect elicited by mood inductions or stressors (e.g., Richell & Anderson, 2004; Stroud, Tanofsky-Kraff, Wilfley, & Salovey, 2000). Studies taking the former approach have shown that, in addition to increasing subjective feelings of memory vividness (for a review, see Phelps & Sharot, 2008), emotion strongly influences the way that information is perceived (for a review, see Zadra & Clore, 2011) and the types of details that are recalled (for reviews, see Kensinger, 2009; Yonelinas & Ritchey, 2015). For instance, testing memory for emotional images, such as from the IAPS database, has shown that negative items tend to be remembered in more detail (Kensinger, Garoff-Eaton, & Schacter, 2007), and the presence of emotion increases activity in visual cortex (Kark & Kensinger, 2015; Lang et al., 1998). In scenes that combine emotional items with neutral background information, emotion tends to enhance memory for central emotional information while impairing memory for peripheral scene details (Bisby & Burgess, 2013; Kensinger et al., 2007), suggesting



Maureen Ritchey maureen.ritchey@bc.edu

Department of Psychology, Boston College, Chestnut Hill, MA, USA

that even in static images, the effects of emotion can be quite specific to particular image features.

Although scene stimuli are visually complex, they lack the temporal dynamics of real-life experiences. As such, it remains unclear how the time course of emotional experience affects the processing of complex events. Outside of the emotion domain, naturalistic video stimuli (e.g., clips from movies or television shows) have been increasingly used to investigate the dynamic nature of cognitive and neural processes during perception and subsequent episodic memory retrieval. For example, video stimuli have been used to demonstrate similarities in episodic representation between perception and retrieval (Bird, Keidel, Ing, Horner, & Burgess, 2015; Chen et al., 2017; Oedekoven, Keidel, Berens, & Bird, 2017). Video stimuli also have provided novel insight into how the brain codes for event boundaries and represents timescales (Baldassano et al., 2017). These findings have significantly furthered our understanding of how the brain represents and reconstructs complex events and could not have been ascertained using static images. There has also been a push within social psychology and neuroscience to use naturalistic stimuli (for a review, see Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012) to study, for example, how interpersonal dynamics develop (Curio, Bülthoff, & Giese, 2010; Klin, Jones, Schultz, Volkmar, & Cohen, 2002) and influence our perception and interpretation of the world (Putman, Hermans, & van Honk, 2006). Thus, research across the broad fields of psychology and neuroscience are in need of ecologically valid, dynamic stimuli.

Although research using temporally dynamic stimuli has provided novel insight into cognitive and neural processes of episodic memory, emotional content has not been directly and systematically measured or manipulated in these studies. For example, although the Sherlock database includes valence and arousal ratings collected from several experimenters (Chen et al., 2017), it does not include ratings from a naive sample of subjects over whom ratings could be averaged to increase generalizability to the population. Within the emotion literature, only a limited number of video stimulus databases have been developed to date. The Database for Emotion Analysis Using Physiological Signals (DEAP; Koelstra et al., 2012) comprises minute-long clips from music videos that were rated for valence, arousal, familiarity, dominance, and like versus dislike while electroencephalographic (EEG) data were collected. DEAP is thus optimal for other researchers to use as a comparison against other methods to measure affective states, but music videos are not an ideal proxy for complex, real-life experiences, because they contain music rather than dialogue and often lack a realistic storyline. The MAHNOB-HCI database (Soleymani, Lichtenauer, Pun, & Pantic, 2012) includes short film clips taken from well-known emotion-inducing movies, such as *The Shining* and *Gangs of New York*, also with accompanying EEG recordings. Similarly, the DECAF database (Abadi et al., 2015) includes data on behavioral, magnetoencephalographic, and other physiological responses to stimuli from both DEAP and MAHNOB–HCI. DECAF also contains dynamic annotations indicating the intended evoked emotion made by experts who had viewed the film clips numerous times. Another multimodal stimulus set, the FilmStim database (Shaefer, Nils, Sanchez, & Philippot, 2010), comprises 1- to 7-min-long film segments from movies that had been shown to elicit the basic emotions of anger, sadness, fear, and disgust (Ekman, 1984), as well as amusement, tenderness, and a neutral state.

These existing stimuli datasets have some limitations, which have motivated the present project. First, film clips taken from movies, television shows, or music videos may be processed differently from autobiographical events, because participants are aware that the events depicted are fictional (Abraham, von Cramon, & Schubotz, 2008). Video clips taken from popular media such as movies and television may also be familiar to some participants, which can affect both memory (Tulving, Markowitsch, Craik, Habib, & Houle, 1996; Wilson & Rolls, 1993) and emotion processing (Ishai, Pessoa, Bikle, & Ungerleider, 2004). Thus, when the possibility of familiarity exists, a measure of stimulus familiarity is necessary for the interpretation of memory or emotion results. Additionally, though other video databases have measured emotion, this has typically been rated over short segments, thus averaging over a subsection of the full-length stimulus. A dynamic or continuous measure of participants' emotional experiences while watching stimuli would be useful for future experiments.

The goal of the present study was to develop a new set of natural video stimuli, selected and rated for their affective content, that would enable future research on the dynamics of emotion processing and memory. To improve the ecological validity of this set, we focused on clips from real-life news telecasts, covering stories ranging from terrorist attacks and deadly car accidents to military homecomings and rescued animals. To account for the dynamic nature of the stimuli, participants continuously rated the valence of the video throughout its presentation and then judged how emotionally intense (arousal judgment) and pleasant (summary valence judgment) each video was overall. Furthermore, a basic memory test allowed us to collect subjective memorability ratings for each video, which we hope will be useful for future studies aiming to directly contrast "memorable" and "forgettable" experiences (cf. Bainbridge, 2017). In addition to reporting normative ratings for each video, we investigate the relationship between these measures to provide an in-depth characterization of the stimulus set. Thus, this stimulus set will allow researchers to assess memory for natural, real-life experiences and how the accompanying cognitive and neural processes are dynamically modulated by emotion.



Method

Participants

A total of 100 participants (69 females, 31 males) took part in the present experiment, with 50 participants rating each video. The sample size was determined a priori to be somewhat larger than those used in other studies reporting normative data from video stimuli. For example, the MAHNOB-HCI database (Soleymani et al., 2012) reports data from 27 participants, the DEAP database (Koelstra et al., 2012) reports data from 32, and the DECAF database (Abadi et al., 2015) from 30. Our sample size was chosen before data collection, with the goal of collecting enough normative data to allow researchers to bin videos into negative, neutral, and positive categories for their own experiments. With emotional materials, however, there are often individual differences in participants' affective responses to the stimuli, and thus we also encourage researchers to collect normative ratings from their own sample to confirm our category assignments, especially for participants drawn from populations with different characteristics to those reported here. All participants were between the ages of 18 and 22 years (mean = 19.01, SD = 0.85), had normal or corrected-to-normal vision, and had no current diagnoses or history of psychological or neurological disorders. Informed consent was obtained from all participants. All procedures were approved by the Boston College Institutional Review Board, and participants received course credit for their time

Materials

The videos were gathered from a television news archive, found at https://archive.org/details/tv. The videos were downloaded on the basis of searches for keywords or phrases in the transcript. To get a range of the negative videos, keywords included words such as "disaster," "murder," "poverty," "victim," and "tragic." The neutral videos including words such as "weather," "traffic," "school," "construction," and "business," and the positive videos were found with words including "happy," "celebration," "heartwarming," "surprise," and "uplifting." These videos were then manually filtered to remove any with low-quality resolution, any containing an event similar to other videos, and any reporting highly familiar, international news stories. Videos containing only a verbal description of the event were also excluded; thus, the remaining videos all contained some visual footage of the event. This process resulted in a total of 144 videos that were selected to be used in the norming experiment, 48 of which related to negative keywords, 48 to neutral keywords, and 48 to positive keywords, on the basis of keyword valence judged by the experimenters. All videos were trimmed to be between 20 and 52 s in duration, and to remove any footage at the beginning or end that did not pertain to the central news story. The mean duration was 42.15 s (SD = 7.70). All videos were 640 pixels in width and between 360 and 480 pixels in height.

Procedure

After giving informed consent, participants completed the viewing phase of the experiment (see Fig. 1). The 144 videos were divided into two lists of 72; each list contained 24 videos of each predefined valence category. Each participant viewed the videos from one of the two lists, to limit the length of the experimental session, which lasted approximately 1 h 30 min. Lists were assigned to participants in an alternating fashion. Therefore, 50 participants rated each of the 144 videos. The order of the videos was randomized per participant, and the session was divided into four study-test blocks, each containing 18 videos, to allow for rest breaks. For each study trial, participants first watched the video while adjusting a continuous slider to indicate how pleasant the video was to them at that moment in time, on a scale of extremely unpleasant (coded as 1) to extremely pleasant (coded as 9). This "dynamic valence" slider started at a random position on the scale at the beginning of each trial, and participants were asked to adjust the slider as quickly as possible to reflect their impression of the video and to keep adjusting it as the perceived pleasantness changed. The location of the slider was sampled every 100 ms. At the end of each video, participants were asked an additional four questions: overall valence ("summary valence") on a scale of 1 (extremely unpleasant) to 9 (extremely pleasant), overall arousal ("intensity") on a scale of 1 (not at all intense) to 9 (extremely intense), the familiarity of the video with the options of "I have seen this exact news footage before," "I am familiar with the news story but I have not seen this footage," and "I have not seen or heard of this news story before," and finally if the story depicted in the video was coherent or "easy to follow" with "yes" or "no" response options.

Immediately after watching all 18 videos in the study phase of a block, participants completed a memory test. On each test trial, participants were first shown the first 3 s of the video as a retrieval cue. After this clip, they were asked to rate how vividly they could recollect the auditory details about the video and the visual details of the video, both using the scales 1 (not at all vividly) to 9 (extremely vividly). Participants were instructed to try to remember as much of the video content as possible before responding. They were then shown a scale from 10 to 60 s, with tick marks at 10-s intervals, and were asked to estimate the total duration of the video by moving the slider along the scale and pressing the spacebar to confirm their response. The slider appeared in a random location along this duration scale on every trial. All responses were self-paced.



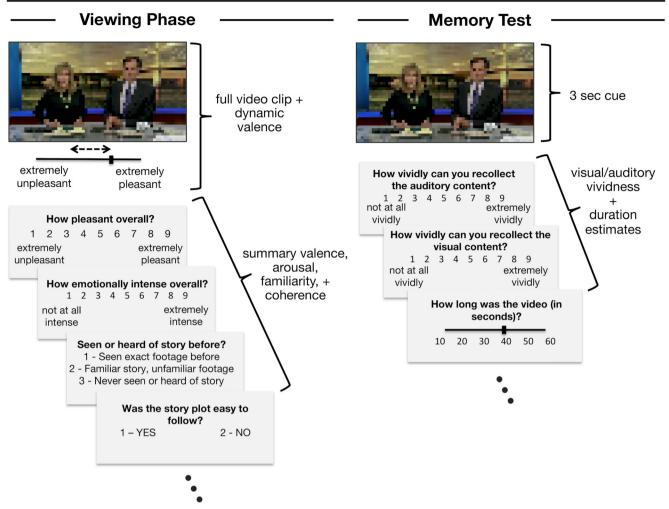


Fig. 1 Task design. During the viewing phase, participants watched news video clips while continuously rating pleasantness. After each clip, they then rated summary pleasantness (valence), emotionally intensity, familiarity with the story, and story coherence. For the memory test,

which was interleaved with the viewing phase in blocks, participants were cued to retrieve each video with a 3-s clip and then judged the vividness of their memory for the auditory content, visual content, and estimated the full video clip length

Results

Videos were excluded prior to the data analysis if more than 25% of the participants rated the video as incoherent (16 videos). Two additional videos were removed from the data analyses as a result of experimenter error: one because its duration was actually significantly longer than 52 s, and one because it showed almost identical news footage to that in a different video. The remaining sample of 126 videos (mean duration = 42.43 s, SD = 7.44 s) are included in all analyses below. All data for the videos can be found at http://www.thememolab.org/paper-videonorming, including the means of all measures collected from participants—dynamic and summary ratings of emotional valence and arousal, ratings of video familiarity, and mean memory vividness and duration estimates. We additionally include information coded by the researchers, including semantic categories, number of

people present, location, and year, as well as short and long unique video descriptions. Although we refrain from directly distributing the videos due to copyright issues, we have provided the URLs for each video and example code that other researchers can use to download the same videos from the TV news archive for use in their own research. All analyses were performed using R version 3.5.0 with RStudio version 1.1.456 (R Core Team, 2018; RStudio Team, 2016). We used linear mixed effects analysis with maximum likelihood estimation, from R's lme4 package (Bates, Maechler, Bolker, & Walker, 2015), to quantify the relationship between variables across videos and participants. In all cases, random intercepts were included for subject and video, and random slopes were also included for subject. The p value for each fixed effect was obtained by a likelihood ratio test of the full model against a null model without the effect in question.



Summary valence and arousal ratings

The summary valence and arousal ratings collected at the end of each video were averaged across participants, and the distribution of these mean ratings can be seen in Fig. 2. We first analyzed the relationship between these ratings across the videos. The mean valence and mean arousal displayed a stereotypical asymmetric V-shaped relation (for a review, see Kuppens, Tuerlinckx, Russell, & Barrett, 2013). To quantify this relationship, we modeled arousal as the dependent variable, with linear and quadratic valence terms as fixed effects. Valence was related to arousal both linearly $[\chi^2(1) = 124.52, p < .001, \beta = -1.22 \pm 0.08$ (standard error)], and quadratically $[\chi^2(1) = 123.38, p < .001, \beta = 1.19 \pm 0.08]$, such that videos at the more negative and positive ends of the valence spectrum were rated as being more intense than neutral videos, but negative videos tended to be the most emotionally intense.

Next we analyzed the variability of summary valence and arousal across participants. To this end, we iteratively correlated the vector of each participant's summary valence ratings for each video with the summary valence ratings of every other participant who had watched the same videos. The mean of these correlations (mean r=.74) reflects a high level of consistency of perceived valence across participants. Repeating this process with arousal ratings revealed that emotional intensity was also stable across participants (mean r=.40), although to a lesser degree than valence. To determine which videos were the most variable in their ratings across participants, we calculated the standard deviation of summary valence (mean SD=2.03) and arousal (mean SD=1.37) across participants for each video. Notably, the standard deviation of the valence and arousal measures did not vary

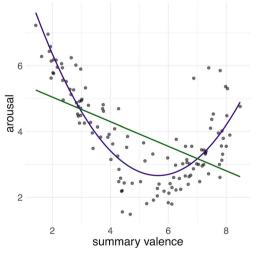


Fig. 2 Each video plotted by mean ratings of arousal (1 = least intense, 9 = most intense) and valence (1 = most negative, 9 = most positive). The green line represents the linear relation between measures; the purple line represents the quadratic relationship between measures

according to their mean values, indicated by multiple regression analyses revealing no significant linear or quadratic relationship between mean summary valence and valence variability (ps > .125), or between mean summary arousal and arousal variability (ps > .145).

To explore the relationship between female and male ratings of summary valence and arousal, we calculated the correlation across videos between mean female and mean male ratings. Summary valence ratings were highly correlated across sex, r(124) = .96, p < .001. Arousal ratings were also highly correlated across sex, r(124) = .92, p < .001, suggesting that females and males rated the videos similarly.

Dynamic pleasantness ratings

We next investigated the properties of the dynamic valence ratings obtained for each video. To examine the time course of these ratings, the slider location data were downsampled to a rating every 500 ms, and the mean rating across participants was calculated for each video at every time point (Fig. 3B). To examine how much the valence of each video varied over time, we then calculated the standard deviation of those time point means for each video (Fig. 3C). The first 5 s of dynamic valence ratings were excluded from all calculations, since the valence slider appeared in a random location at the start of each video and participants took a few seconds to move the slider on the basis of their first impression of the video content. The standard deviation of dynamic valence was strongly correlated with the video arousal ratings, r(124) = .56, p < .001, illustrating that negative and positive videos were associated with more variability in continuous valence ratings over the video time course. To further explore this relationship, we divided the time course of the dynamic valence ratings into thirds (beginning, middle, and end of video) and calculated the standard deviation of the dynamic valence ratings within each third for each video, then averaged these mean standard deviations for each video across participants. We then tested the correlation between the mean standard deviation within each third and the mean arousal for each video. The mean standard deviation within the beginning third of videos was correlated with the mean arousal, r(124) = .25, p < .01. Arousal was not significantly correlated, however, with the standard deviation within the middle third, r(124) = .12, or the end third, r(124) = .03, ps > .10. Together, these findings may reflect the gradual accumulation of information that leads participants to shift from a relatively neutral impression of the video to a more clearly positive or negative impression for the more emotionally arousing videos. The standard deviation of dynamic valence was not significantly correlated with summary valence (r < .20, p > .20).

To analyze the correspondence between dynamic valence ratings and summary valence ratings, we first correlated the mean of dynamic valence ratings with the mean summary



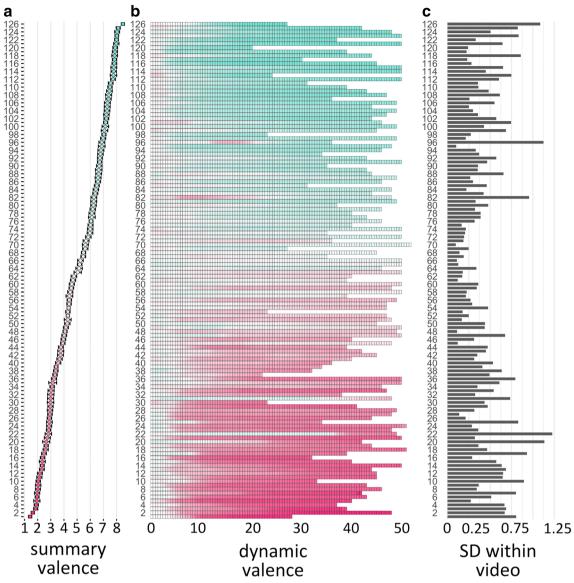


Fig. 3 Valence ratings. Each row represents the data for a single video, arranged in order of (A) ascending mean summary valence. Participants rated summary valence on a scale from 1 to 9 (the mean ratings ranged from 1.43 to 8.48). (B) Mean dynamic valence ratings across participants

at each 500-ms timepoint. Pink represents more negative ratings, and turquoise represents more positive ratings. (C) Standard deviations of the mean dynamic valence ratings over time, representing how much valence was perceived to fluctuate within each video

valence ratings for each video. This revealed a strong relationship, r(124) = .96, p < .001. To examine whether this correlation was driven by the beginning, middle, or end of each video, we again divided each video's mean dynamic time course into thirds and correlated the mean dynamic valence of each third with the mean summary valence. The mean dynamic valence from the beginning third was strongly correlated with summary valence, r(124) = .89, p < .001. This relationship increased for mean dynamic valence within the middle third, r(124) = .95, p < .001, and the end third, r(124) = .99, p < .001, which had the highest correlation between dynamic and summary valence ratings. The peak of the dynamic valence ratings was also calculated across participants for each

video, by centering the valence scale about 0 and taking the maximum absolute value at any time point for each video. Similarly, the peak dynamic valence was highly correlated with the mean summary valence ratings, r(124) = .97, p < .001.

Memorability

Subjective memory vividness We next addressed the data from the memory test, which occurred immediately after video encoding. We first analyzed the relationship between visual and auditory memory vividness ratings and then tested the influence of summary valence and arousal on subsequent



memory vividness. Visual vividness and auditory vividness were highly correlated [r(124) = .96, p < .001], so subsequent analyses use a single composite vividness measure, calculated as the average of visual and auditory vividness on a trial-bytrial basis (see Table 1). To examine the relationship between memory vividness and summary valence and arousal, we tested a linear mixed effects model, with vividness as the dependent variable and fixed effects representing the linear and quadratic effects of summary valence and arousal (see Fig. 4). Summary valence influenced vividness linearly $[\chi^2(1)]$ = 19.28, p < .001, $\beta = -0.33 \pm 0.07$, indicating that positive videos were remembered more vividly than more neutral or negative videos. The quadratic summary valence term was also significant $[\chi^2(1) = 44.31, p < .001, \beta = 0.49 \pm 0.07],$ meaning that highly negative or highly positive videos were remembered more vividly than more neutral videos. Neither the linear nor the quadratic arousal term uniquely contributed to the model fit.

As with summary valence and arousal, we examined the stability of vividness by iteratively correlating each participant's vector of composite vividness ratings across videos with the composite vividness ratings of every other participant who had watched the same videos. The mean of these correlations revealed the stability of video memorability across participants (mean r = .35). Additionally, we again explored the relationship between female and male visual and auditory vividness ratings by calculating the correlation of mean female and mean male ratings across videos. Both visual vividness [r(124) = .90, p < .001] and auditory vividness [r(124) = .91, p < .001] were highly correlated between the sexes.

Temporal memory precision Our final set of analyses focused on memory for video duration. To examine overall accuracy for duration estimation, the mean correlation between the actual and estimated duration was calculated within each participant. Pearson's correlation coefficients were averaged across participants after transformation to Fisher z values (mean z = 0.29, SE = 0.02). A one-sample t test revealed that the true mean was significantly greater than zero, t(97) = 16.62, p < .001, meaning that participants were able to recall video durations with above-chance accuracy. We next asked whether temporal memory was influenced by vividness. To quantify

Table 1 Summary statistics for emotion and memory ratings across videos

nce Arousai	Visual Vividness	Auditory Vividness
3.92	6.94	6.34
1.36	1.20	1.21
3.71	7.15	6.55
1.49	3.00	2.53
7.22	8.77	8.60
	3.71 1.49	3.71 7.15 1.49 3.00

memory precision for temporal duration, we calculated the estimate error as (|actual duration – duration estimate|)/actual duration on a trial-by-trial basis. We then tested a linear mixed effects model with duration estimate error as the dependent variable and a fixed effect representing composite vividness. Vividness predicted duration error [$\chi^2(1) = 26.11, p < .001, \beta$ = -0.12 ± 0.02], indicating that the duration estimates of videos that were remembered vividly were most accurate (lowest error). To test whether valence modulated memory for video duration, we next modeled duration error as the dependent variable, with fixed effects for the linear and quadratic effects of summary valence. Neither the linear nor the quadratic valence term significantly contributed to the model fit. Similarly, to test whether arousal modulated memory for video duration, we again modeled duration error as the dependent variable, with fixed effects for the linear and quadratic effects of arousal. Again, neither the linear nor the quadratic arousal term significantly contributed to the model fit, suggesting that temporal memory precision was not significantly affected by video valence or arousal.

Discussion

Naturalistic video stimuli have been increasingly used to investigate the cognitive and neural processes involved in representing dynamic events. Although emotion is a natural part of real experiences and can strongly influence event processing and memory, previous studies have not systematically manipulated the emotional content of such videos. In this study, we developed a new stimulus set of real news broadcasts and collected ratings of emotionality and subjective memory for each video. The results of our norming study indicated that, as intended, the stimulus set varied in its emotional content and memorability, and that emotional valence predicted the subjective vividness with which the videos were remembered.

In this study, we collected summary ratings of valence and arousal, as well as continuous ratings of valence. The summary ratings showed a pattern similar to that reported for other kinds of emotional stimulus sets (e.g., Marchewka et al., 2014): valence and arousal were significantly related to one another. A linear relationship reflected that videos that were more emotionally negative tended to be the most emotionally intense, consistent with past work documenting a negativity bias for participant ratings of the valence and intensity of emotional stimuli (Kuppens et al., 2013). We also observed a quadratic (V-shaped) relationship, reflecting that arousal was greatest for the most negative and most positive videos, replicating many prior studies of these emotional components (e.g., Lang & Bradley, 2007).

Because of the dynamic nature of the stimuli, we were especially interested in the properties of the continuous



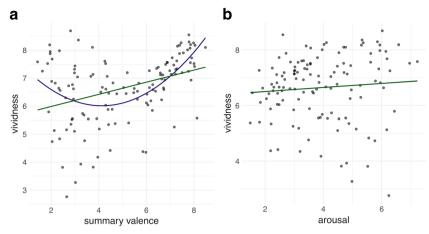
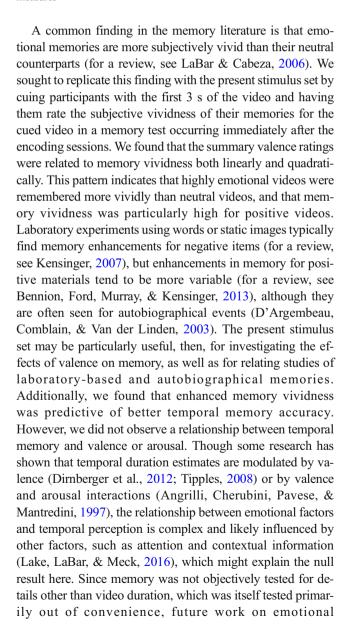


Fig. 4 Each video plotted by mean ratings of (A) composite vividness ($1 = least \ vivid$, $9 = most \ vivid$) by valence ($1 = most \ negative$, $9 = most \ positive$) and (B) composite vividness by arousal ($1 = least \ intense$, $9 = most \ vivid$)

most intense). Green lines represent the linear relations between measures; the purple line represents the quadratic relationship between measures

valence ratings, which reflected the changing emotional experiences of the participants while they viewed the videos. The continuous ratings were strongly related to the summary valence ratings, both in terms of the average rating and the absolute peak (negative or positive) rating. We also found that arousal was related to the standard deviation of the continuous valence ratings. That is, the videos in which the continuous valence ratings changed the most were the most arousing videos. Because this relationship was strongest during the first third of the videos, this pattern may reflect the gradual accumulation of information as participants developed their impression of the more emotional videos. Alternatively, large changes in valence may have been unexpected, and thus amplified emotional arousal.

Though our analyses focused on relationships among the measures, this study also provided some interesting qualitative observations. For example, the video containing the largest deviation in dynamic valence was also the video remembered most vividly. This particular clip showed a young girl standing on a toilet, which initially seems humorous. However, it is later revealed that she is demonstrating what she was taught to do during an active shooter drill at her preschool. The video with the most negative mean summary valence was also the video with the most negative mean peak: a report that white police officers who had beaten a black motorcyclist were acquitted by an all-white jury. Likewise, the video with the most positive mean summary valence also had the most positive mean peak: footage of an excited dog greeting a soldier who had just returned home. This exemplifies the strong relationship between peak valence and summary valence. Overall, though, participants rated the videos as varying widely along the spectrum of valence, which makes these videos ideal for studies examining differential effects of positive and negative emotion. Because valence and arousal were rated as separate measures, these videos can be used to examine differential effects of these variables.





enhancement should examine what aspects of the videos receive the greatest mnemonic benefits.

Several database characteristics should be considered when interpreting the findings described above and deciding whether this stimulus set is appropriate to answer a particular research question. These videos contain third person ("something terrible happened") versus first person ("something terrible happened to me") emotional experiences, and thus perspective and self-relevancy should be taken into account as factors that may modulate memory. Recalling memories from a first-person perspective is related to greater memory vividness, detail, and coherence (Sutin & Robins, 2010), and differentially recruits regions in prefrontal cortex (St Jacques, Conway, Lowder, & Cabeza, 2011). Self-relevant content also tends to be remembered better than self-irrelevant content (Rogers, Kuiper, & Kirker, 1977) and is associated with increased activity in medial prefrontal cortex and cingulate cortex during retrieval (Kelley et al., 2002). However, it should be noted that first-person emotional events are not necessarily self-relevant (e.g., observing something happening to someone else), that third-person emotional events can be selfrelevant (e.g., death of a celebrity one identifies with), and that third-person emotional events are frequently and naturally experienced via news broadcasts (e.g., celebrity death, mass shooting, natural disaster). Thus, these stimuli replicate some naturally occurring emotional experiences, but may not generalize to all emotional experiences.

The effects of auditory (news broadcaster) and visual (headline, crawler) narration should additionally be considered when generalizing to emotional experiences outside of news broadcasts. Though discussion of an event can occur naturally, a newscaster narrative overlaid on previously recorded visual footage (versus narration from a newscaster reporting from the scene) could result in separate visual and auditory narratives that might have some differences in information content. For example, if a participant remembers that a video about an accident contains a wrecked truck, it should be considered whether this information was represented visually or through auditory narrative.

There are also several cultural factors to consider when using this stimulus database. Newscaster narrations and onscreen text are in English, and thus these videos may not be ideal for participants who are not native English speakers. Additionally, participants who are unfamiliar with the format of North American-style news broadcasts may process the videos differently than those who are very familiar—for example, when integrating information from auditory and visual modalities. In addition to the format, the content of news broadcasts are culturally specific and thus emotional reactions and semantic interpretations may be culturally specific as well. Related to this point, the normative data reported here were collected from a relatively homogeneous sample of young adults, and thus these data may not generalize to other

populations. Additionally, our sample was collected on the basis of the convenience of our psychology subject pool, which includes mostly female students 18 to 22 years of age from a university located in an East Coast city in the United States. Although we did not have balanced numbers of males and females, we investigated the correspondence between the male and female ratings by calculating the correlation of female and male ratings across videos. Summary valence, arousal, visual vividness, and auditory vividness were all highly correlated between females and males (all rs > .90). Thus, it does not seem as though there are any large sex differences in the ratings, although it is possible that subtle differences could emerge with a larger, more balanced sample size. In summary, these stimuli and the reported normative data reflect naturalistic emotional experiences, but further work is needed to assess the generalizability to different populations. We strongly encourage researchers to collect additional normative data when studying a population that may be expected to experience the videos differently.

Our video stimuli build upon existing video stimulus databases in several ways. First, this set includes more stimuli than comparable databases. This is particularly important for experiments that require analyses to be focused on a subset of the data—for example, a memory experiment that compares remembered items between valence categories. Additionally, each video can be treated as a discrete event, or videos can be played sequentially like a news broadcast in order to create a longer event that is emotionally variable yet thematically cohesive. We have also provided information about video familiarity and semantic descriptions so that researchers can easily select the stimuli that are most appropriate for their studies. Finally, this video stimulus set is, to our knowledge, the first of its kind to provide dynamic valence ratings of real-world emotional events. As such, it can facilitate examination of the neural and cognitive processes underlying dynamic, complex experiences. For example, it is unclear how changes in emotionality are related to later memory accuracy and vividness. In other words, do "spikes" in emotion enhance memory for only those emotion-eliciting parts of an event, or do moments of emotional transition (e.g., an emotionally negative moment becoming positive) lead to better memory? Is memory for less emotional parts of an event also enhanced, or are neutral moments overshadowed by more salient information and forgotten? In addition to questions about the direct effects of emotion, these stimuli can be used to understand how we can effectively modulate our emotions. For example, emotion regulation strategies have differential effects on both experiential negative emotion and memory for an emotional event (Richards & Gross, 2000), as well as on long-term mental health (Gross & John, 2003). However, the neural mechanisms that underlie healthy versus unhealthy emotion regulation are largely unknown. Answering questions such as these is critical to fully understanding the dynamic effects of emotion and how we can modulate these effects.



Acknowledgements This work was supported in part by grant NIH R00MH103401 (to M.R.) and by a Brain & Behavior Research Foundation NARSAD Young Investigator Grant (also to M.R.). We thank C. Bliss, M. Bluestone, E. Iannazzi, and E. Judge for their assistance with stimulus and data collection. The authors report no conflicts of interest.

References

- Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., & Sebe, N. (2015). DECAF: MEG-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing*, 6, 209–222. doi:https://doi.org/10.1109/TAFFC.2015. 2392932
- Abraham, A., von Cramon, D. Y., & Schubotz, R. I. (2008). Meeting George Bush versus meeting Cinderella: The neural response when telling apart what is real from what is fictional in the context of our reality. *Journal of Cognitive Neuroscience*, 20, 965–976. doi:https:// doi.org/10.1162/jocn.2008.20059
- Angrilli, A., Cherubini, P., Pavese, A., & Mantredini, S. (1997). The influence of affective factors on time perception. *Perception & Psychophysics*, 59, 972–982. doi:https://doi.org/10.3758/ BF03205512
- Bainbridge, W. A. (2017). The memorability of people: Intrinsic memorability across transformations of a person's face. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 706–716.
- Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, 95, 709–721.e5. doi: https://doi.org/10.1016/j.neuron.2017.06.041
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01
- Bennion, K. A., Ford, J. H., Murray, B. D., & Kensinger, E. A. (2013). Oversimplification in the study of emotional memory. *Journal of the International Neuropsychological Society*, 19, 953–961.
- Bird, C. M., Keidel, J. L., Ing, L. P., Horner, A. J., & Burgess, N. (2015). Consolidation of complex events via reinstatement in posterior cingulate cortex. *Journal of Neuroscience*, 35, 14426–14434. doi: https://doi.org/10.1523/JNEUROSCI.1774-15.2015
- Bisby, J. A., & Burgess, N. (2013). Negative affect impairs associative memory but not item memory. *Learning & Memory*, 21, 21–27.
- Bradley, M. M., & Lang, P. J. (1999). Affective Norms for English Words (ANEW): Stimuli, instruction manual and affective ratings (Technical Report No. C-1). Gainesville, FL: University of Florida, NIMH Center for Research in Psychophysiology.
- Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, 20, 115–125. doi:https://doi.org/10.1038/nn.4450
- Curio, C., Bulthoff, H. H., & Giese, M. A. (2010). Dynamic faces: Insights from experiments and computation. Cambridge, MA: MIT Press.
- D'Argembeau, A., Comblain, C., & Van der Linden, M. (2003). Phenomenal characteristics of autobiographical memories for positive, negative, and neutral events. *Applied Cognitive Psychology*, 17, 281–294.
- Dirnberger, G., Hesselmann, G., Roiser, J. P., Preminger, S., Jahanshahi, M., & Paz, R. (2012). Give it time: neural evidence for distorted time perception and enhanced memory encoding in emotional situations. NeuroImage, 63, 591–599. doi:https://doi.org/10.1016/j.neuroimage.2012.06.041

Ekman, P. (1984). Expression and the nature of emotion. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 319–344). Hillsdale, NJ: Erlbaum.

- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and wellbeing. *Journal of Personality and Social Psychology*, 85, 348–362.
- Ishai, A., Pessoa, L., Bikle, C., & Ungerleider, L. G. (2004). Repetition suppression of faces is modulated by emotion. *Proceedings of the National Academy of Sciences*, 101, 9827–9832. doi:https://doi.org/ 10.1073/pnas.0403559101
- Kark, S. M., & Kensinger, E. A. (2015). Effect of emotional valence on retrieval-related recapitulation of encoding activity in the ventral visual stream. *Neuropsychologia*, 78, 221–230. doi:https://doi.org/ 10.1016/j.neuropsychologia.2015.10.014
- Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, 14, 785–794.
- Kensinger, E. A. (2007). Negative emotion enhances memory accuracy: Behavioral and neuroimaging evidence. *Current Directions in Psychological Science*, 16, 213–218. doi:https://doi.org/10.1111/j. 1467-8721.2007.00506.x
- Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, 1, 99–113. doi:https://doi.org/10.1177/1754073908100432
- Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). How negative emotion enhances the visual specificity of a memory. *Journal of Cognitive Neuroscience*, 19, 1872–1887.
- Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Archives of General Psychiatry, 59, 809–816.
- Koelstra, S., Muhl, C., Soleymani, M., Lee, J., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A database for emotion analysis; using physiological signals. *IEEE Transactions* on Affective Computing, 3, 18–31.
- Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, 139, 917–940.
- LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, 7, 54–64. doi:https://doi. org/10.1038/nrn1825
- Lake, J. I., LaBar, K. S., & Meck, W. H. (2016). Emotional modulation of interval timing and time perception. *Neuroscience and Biobehavioral Reviews*, 64, 403–420.
- Lang, P. J., & Bradley, M. M. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 29–46). New York, NY: Oxford University Press.
- Lang, P. J., Bradley, M. M., & Cuthbert, B.N. (2008). International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual (Technical Report A-8). Gainesville, FL: University of Florida, NIMH Center for Research in Psychophysiology.
- Lang, P. J., Bradley, M. M., Fitzsimmons, J. R., Cuthbert, B. N., Scott, J. D., Moulder, B., & Nangia, V. (1998). Emotional arousal and activation of the visual cortex: An fMRI analysis. *Psychophysiology*, 35, 199–210.
- Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014).
 The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. Behavior Research Methods, 46, 596–610. doi:https://doi.org/10.3758/s13428-013-0379-1
- Oedekoven, C. S. H., Keidel, J. L., Berens, S. C., & Bird, C. M. (2017). Reinstatement of memory representations for lifelike events over the course of a week. *Scientific Reports*, 7, 14305.



Phelps, E. A., & Sharot, T. (2008). How (and why) Emotion enhances the subjective sense of recollection. *Current Directions in Psychological Science*, 17, 147–152.

- Putman, P., Hermans, E., & van Honk, J. (2006). Anxiety meets fear in perception of dynamic expressive gaze. *Emotion*, 6, 94–102. doi: https://doi.org/10.1037/1528-3542.6.1.94
- R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/
- Richards, J. M., & Gross, J. J. (2000). Emotion regulation and memory: The cognitive costs of keeping one's cool. *Journal of Personality and Social Psychology*, 79, 410–424.
- Richell, R. A., & Anderson, M. (2004). Reproducibility of negative mood induction: a self-referent plus musical mood induction procedure and a controllable/uncontrollable stress paradigm. *Journal of Psychopharmacology*, 18, 94–101.
- Risko, E. F., Laidlaw, K., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. Frontiers in Human Neuroscience, 6, 143. doi:https://doi.org/10.3389/fnhum. 2012.00143
- Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 35, 677–688. doi:https://doi.org/10.1037/0022-3514.35.9.677
- RStudio Team. (2016). RStudio: Integrated development for R. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/
- Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, 24, 1153–1172. Retrieved from https://sites.uclouvain.be/ipsp/FilmStim/
- Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A multi-modal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3, 42–55.

- St Jacques, P. L., Conway, M. A., Lowder, M. W., & Cabeza, R. (2011). Watching my mind unfold versus yours: An fMRI study using a novel camera technology to examine neural differences in self-projection of self versus other perspectives. *Journal of Cognitive Neuroscience*, 23, 1275–1284. doi:https://doi.org/10.1162/jocn. 2010.21518
- Stroud, L. R., Tanofsky-Kraff, M., Wilfley, D. E., & Salovey, P. (2000).
 The Yale Interpersonal Stressor (YIPS): Affective, physiological, and behavioral responses to a novel interpersonal rejection paradigm. *Annals of Behavioral Medicine*, 22, 204–213.
- Sutin, A. R., & Robins, R. W. (2010). Correlates and phenomenology of first and third person memories. *Memory*, 18, 625–637.
- Tipples, J. (2008). Negative emotionality influences the effects of emotion on time perception. *Emotion*, 8, 127–131. doi:https://doi.org/10.1037/1528-3542.8.1.127
- Tulving, E., Markowitsch, H. J., Craik, F. E., Habib, R., & Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cerebral Cortex*, 6, 71–79.
- Wilson, F. A., & Rolls, E. T. (1993). The effects of stimulus novelty and familiarity on neuronal activity in the amygdala of monkeys performing recognition memory tasks. *Experimental Brain Research*, 93, 367–382.
- Yonelinas, A. P., & Ritchey, M. (2015). The slow forgetting of emotional episodic memories: an emotional binding account. *Trends in Cognitive Sciences*, 19, 259–267. doi:https://doi.org/10.1016/j.tics. 2015.02.009
- Zadra, J. R., & Clore, G. L. (2011). Emotion and perception: The role of affective information. Wiley Interdisciplinary Reviews: Cognitive Science, 2, 676–685.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

