Affective Science
https://doi.org/10.1007/s42761-021-00100-7

METHODS PAPER

EmoCodes: a Standardized Coding System for Socio-emotional Content in Complex Video Stimuli

M. Catalina Camacho[1] · Elizabeth M. Williams[1] · Dori Balser[1] · Ruchika Kamojjala[1] · Nikhil Sekar[1] · David Steinberger[1] · Sishir Yarlagadda[1] · Susan B. Perlman[2] · Deanna M. Barch[1]

Received: 19 October 2021 / Accepted: 21 December 2021

Abstract

Social information processing is vital for inferring emotional states in others, yet affective neuroscience has only begun to scratch the surface of how we represent emotional information in the brain. Most previous affective neuroscience work has used isolated stimuli such as static images of affective faces or scenes to probe affective processing. While this work has provided rich insight to the initial stages of emotion processing (encoding cues), activation to isolated stimuli provides limited insight into later phases of emotion processing such as interpretation of cues or interactions between cues and established cognitive schemas. Recent work has highlighted the potential value of using complex video stimuli to probe socio-emotional processing, highlighting the need to develop standardized video coding schemas as this exciting field expands. Toward that end, we present a standardized and open-source coding system for complex videos, two fully coded videos, and a video and code processing Python library. The EmoCodes manual coding system provides an externally validated and replicable system for coding complex cartoon stimuli, with future plans to validate the system for other video types. The *emocodes* Python library provides automated tools for extracting low-level features from video files as well as tools for summarizing and analyzing the manual codes for suitability of use in neuroimaging analysis. Materials can be freely accessed at https://emocodes.org/. These tools represent an important step toward replicable and standardized study of socio-emotional processing using complex video stimuli.

Keywords Emotion processing · Naturalistic stimuli · Movie-watching · Valence–arousal system

Social information processing (Crick & Dodge, 1994; Lemerise & Arsenio, 2000) is vital for inferring emotional states in others; however, we do not yet know how complex socio-emotional processing is represented in the brain (Dubois & Adolphs, 2016) or the neurodevelopmental basis for affective development (Ruba & Pollak, 2020). Most previous affective neuroscience work has used isolated stimuli such as static images of affective faces or scenes to probe affective processing (e.g., Barch et al., 2012; Hare et al., 2008; Todd et al., 2011). While this work has provided rich insight to the initial stages of emotion processing (encoding cues), activation to isolated stimuli provides limited insight into later phases of emotion processing such as interpretation of cues or interactions between the cues and established cognitive schemas. For instance, a happy face in the context of a person expressing joy in receiving a gift is likely interpreted differently from a happy face in the context of someone taking joy in terrorizing another person. Recent work has highlighted the potential value of using complex video stimuli to probe socio-emotional processing (Cantlon, 2020; Eickhoff et al., 2020; Sonkusare et al., 2019), indicating the need to develop standardized video coding schemas as this exciting field expands. Toward that end, we present a standardized open-source coding system for complex emotional videos (EmoCodes), two coded videos, and a video

Handling editor: Kristen Lindquist

M. Catalina Camacho and Elizabeth M. Williams contributed equally to this work.

✉ M. Catalina Camacho
camachoc@wustl.edu

1 Department of Psychological and Brain Sciences, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, USA

2 Department of Psychiatry, Washington University in St. Louis, 4444 Forest Park Drive, MO 63110 St. Louis, USA

and code processing Python library (*emocodes*). These tools represent another step toward replicable, standardized study of socio-emotional processing using complex video stimuli.

Recent work has already begun to use movie stimuli to examine socio-emotional neural processing in health and disease (Gruskin et al., 2020; Guo et al., 2015; Richardson et al., 2018). In healthy children, neural activation to movie stimuli has been used to capture individual differences in language and math processing (Cantlon & Li, 2013), changes in specific neural networks across age (Richardson, 2018; Richardson et al., 2018), as well as differences in affective processing between adults and children (Camacho et al., 2019; Karim & Perlman, 2017). A study of youth from the Healthy Brain Network found that adolescents with higher depression scores showed less similar activation to their peers during movie-watching (Gruskin et al., 2020). A study of adults with and without depression found that, when shifting from resting state to watching emotional videos, depressed individuals with melancholia had increased connectivity in attentional and interoceptive networks while non-depressed comparison adults showed the opposite effect (Hyett et al., 2015). In an overlapping sample, another analysis found that adults with depression demonstrated less ventromedial prefrontal cortex engagement during a negative film clip as compared to non-depressed adults (Guo et al., 2015).

Across these studies, the approach in characterizing affective content in the videos varied widely, from selecting video clips of predominantly one valence, to using a single metric to capture positive, negative, and intensity information frame-by-frame. The more granular details of the videos used in these studies, such as the video luminance, color vibrance, and audio loudness, are also not uniformly reported, making it unclear to what extent colinear video features may be present in—and inform—each of these findings. For example, if negative content is closely associated with a specific color palette, signal from processing those colors would be represented in those results as well. Having this more granular information can aid in comparisons across studies and across video clips, furthering the utility of movie stimuli for affective research. Finally, broad positive and negative ratings, while statistically useful, do not provide qualitative information as to what specific positive and negative affect is present in the stimuli. Qualitative information about affective stimuli is critical, as humans develop recognition and naming abilities for different specific emotions (i.e., angry, happy, fearful faces/sounds) non-uniformly (Ruba & Pollak, 2020). Further, there is evidence that different emotional categories are likely distinctly processed in the brain in addition to being interpreted differently (Kragel & LaBar, 2016; Kragel et al., 2019). Thus, it is critical that affective neuroscience moves toward not only examining complex stimulus processing, but also carefully characterizing complex affective stimuli to better understand how affective cognition is represented in the brain and how it develops.

To our knowledge, there are no freely available, standardized coding systems for complex affective content presented in audio-visual media. For example, the commonly used Facial Action Coding System (FACS; Ekman et al., 1997) is designed for use in videos or images of real people and does not extend to other media such as cartoons, which are commonly used as stimuli for children. Other studies which have used in-house ratings to create emotion-based regressors in neuroimaging work range from reductive (e.g., coding a single valence determination for each movie frame; Gruskin et al., 2020; Guo et al., 2015; Karim & Perlman, 2017) to capturing broad socio-emotional constructs such as by coding scenes in which characters are engaging in mentalizing (Jacoby et al., 2016) or using a latent variable to capture a wide swath of affective content (Lettieri et al., 2019). While there are advantages to simplifying complex stimuli, reductive approaches do not leverage the full potential of complex movie stimuli for studying individual differences in affective processing. For example, it is common for scenes with social interaction (i.e., with multiple characters) to contain both positive and negative content simultaneously. The differences in processing videos with purely positive, purely negative, or mixed valences are likely of great interest to those studying dysfunctions in emotion processing or affect-biased attention. Further, emotional content may co-occur with other features of videos such as luminance, presence of faces, shifts in color palette, or changes in loudness in some videos more so than others. Characterizing the co-occurrence of these signals is crucial for communicating and understanding imaging findings. Thus, a more granular and standardized system for quantifying affective content is needed.

We therefore present the creation and validation of the EmoCodes video coding system as well as data from two 20-minute videos used in the creation and validation of this coding system. This system produces second-by-second codes for visual, auditory, social, and affective content using an open-source and replicable manual and automated coding system. The EmoCodes system includes both guidance on how to conduct manual frame-by-frame coding of affective content as well as a Python library (*emocodes*) with tools for code processing and automatic, low-level video processing. These codes can be used for behavioral or neuroscientific work, such as regressors in a model of neural activity over time. This system was inspired by the studyforrest coding schema (Labs et al., 2015) and broadened to apply to other forms of video media such as children's cartoons. We demonstrate that the coding provided by this system is both internally reliable across raters and corresponds closely to community ratings of positive and negative content, providing external validation that the codes reflect general community impressions of affective content. While we do not purport to capture the sum total of emotions expressed in complex social interactions, we hope that this tool provides researchers with a means of transparently characterizing and reporting on the emotional content examined in their work. The full manual, codes for the TV episodes used in this validation, and our analysis notebooks can be found at https://emocodes.org/.

Method

EmoCodes Development

Framework Development

Our objective was to create a system that characterized specific features of video stimuli. To that end, we sub-divided the EmoCodes system along two dimensions: automation and objectivity. Features (or "codes") that cannot be automatically derived using software must be manually coded by trained raters and fall into two categories, Objective and Subjective. For each manually coded feature, we aimed for each code to be: (1) concise and reproducible, (2) flexible for researchers of diverse interests, (3) minimally overlapping, and (4) maximally informative given the time spent coding. Overall, the specific codes we present in the EmoCodes system were informed by previous neuroscience work on visual and auditory perception as well as by emotion and developmental theory.

Automatic Codes

Extensive efforts have already been made to automate video annotation of video content such as automatic detection of faces (Jenkins & Burton, 2008), objects (Krizhevsky et al., 2012), and emotional content (Kragel et al., 2019). These tools were developed on realistic stimuli such as live action videos or photographs, however, and it is rare for these algorithms to generalize to abstractions beyond their training data (Ballester & Araujo, 2016). It is therefore expected that these tools do not typically work for cartoon stimuli. Thus, the video features that can be derived automatically from cartoons in a comparable way to live-action videos would be low-level video features, such as brightness and audio loudness, which are agnostic to the type of video being processed. Using the *emocodes* library, low-level features—specifically audio loudness, music tempo, visual salience, color vibrance, luminance, and visual sharpness—can be extracted on a frame-by-frame basis using the ExtractVideoFeatures class from the emocodes.processing.video module. This class wraps the *pliers* (McNamara et al., 2017) and *librosa* (McFee et al., 2021) libraries, organizing the extracted information into a single document. In the context of cartoons, it is possible that any number of these features may be associated with emotional content as animators often manipulate these features to create a clear tone—for example, using shadows or dark colors to denote the arrival of a villain or shifting the color palette to include more blues and purples to denote sad content—thus, it is prudent to test the independence of these features from the affective content of interest when selecting video stimuli for neuroscientific research.

Manual Codes: Selection and Definitions

Video features that need to be manually coded by raters fall into two broad categories: objective or concrete codes (e.g., presence of a face) and subjective or interpretive codes (e.g., if a character is sad). We created lists of codes for each that are of scientific interest and adjusted which category they belonged to with feedback from the independent raters. Because interpretation of emotions is highly variable across individuals (for a discussion, see Barrett et al., 2019), we approached creating definitions as a balance between defining semantic parameters and retaining label intuition. In other words, the goal for each label definition was to make explicit what is meant by each label without implying that other uses of these words are incorrect. Further, we do not rely on stereotyped facial expressions in the definition, but rather the social situations that may elicit each emotion. Please see Table S3 for a full account of the codes that were considered for inclusion in the EmoCodes system. Further, we found that the majority of the affective information in each episode used for development came from the characters (see *External Validation*).

For most codes across the system, a binary choice is forced to enhance agreement across raters and utility across research questions. Binary vectors can be combined to identify more specific content of interest (e.g., combining has_faces with closeup produces a time series of close ups of faces). Each code is completed frame-by-frame. The complete list of codes included in the final manual is included in Tables 1 and 2.

Objective Codes: The objective codes capture non-emotional elements of the video that help contextualize the content coded in the rest of the manual. These codes include character content such as the number of characters on the screen, the presence of faces, the presence of body parts, and the presence of words. This is based on previous work which has found that visual cortex uniquely activates in response to each face and body parts (Grill-Spector et al., 2004; Peelen & Downing, 2007). These codes also characterize the visual composition of the screen such as if it is a close up or wide

Table 1 Objective codes definitions and reliability. Reliability was operationalized as the intraclass correlation (ICC) at the level of each second-by-second rating. Because ICCs require a degree of variability to be useful, the proportion of the whole video which included a non-zero value for each code is included to contextualize these findings

Code label	Definition	Possible code values	Proportion of Video Present		ICC (*each second, entire video*)		ICC range (*each second, 3 min windows*)	
			AHKJ	MLP	AHKJ	MLP	AHKJ	MLP
time_of_day	Denotes if a scene occurs during daylight or night. Any visible daylight was considered "day", including dusk and dawn	0;1	84.6%	86.2%	0.99	0.89	0.95–0.99	0.33–1.00
closeup	If the frame is a closeup or a wider shot. A closeup was defined as one or two characters or objects taking up at least 50% of the screen	0;1	17.3%	37.5%	0.53	0.45	0.12–0.81	0.24–0.77
num_chars	Total number of characters aurally or visually present. Collective characters are counted as 1	{any number}	95.2%	96.2%	0.89	0.75	0.82–0.92	0.51–0.94
collective	This code denotes if the frame includes a group of characters which act as one for the purposes of the scene (e.g., a crowd of fans at a sporting event)	0;1	13.9%	5.8%	0.85	0.87	0.38–1.00	0.33–0.90
has_faces	Presence of any faces or facial features on screen	0;1	91.1%	94.1%	0.71	0.73	0.29–0.84	0.20–0.83
has_body	Presence of any bodies or body parts on screen	0;1	93.5%	93.8%	0.63	0.64	0.15–0.84	0.09–0.84
has_words	Presence of any written words, letters, or grammatical symbols on the screen. This includes credits and title cards as well	0;1	3.8%	3.8%	0.59	0.47	0.28–0.87	0.29–0.63

AHKJ, *All Hail King Julien*, season 1, Episode 2; *MLP*, *My Little Pony*, Season 8, episode 3

shot, capturing if the viewer's attention is being focused on a specific item or character or if there may be more individual differences in attention allocation captured for a given scene. For the complete final list of codes and their definitions, please see Table 1.

Subjective Character Codes: In writing the instructions on how to code the emotion-specific features on a frame-by-frame basis, we endeavored to preserve the natural intuition of emotions as much as possible, disambiguating where necessary to further allow for consistent training and coding. Specifically, we created the manual definitions to create a shared semantic space for each emotion, emphasizing clarity and reproducibility rather than attempting to capture the full range of experiences commonly associated with a given emotion. This process generally involved taking the dictionary definition of each emotion and adding valence–arousal system anchors to help define boundaries. The subjective character codes were divided into two sections. First, global affective information (general codes) was defined utilizing the valence–arousal system to rate the character's overall arousal, positive affect, and negative affect. Second, specific emotions were defined and coded. For each emotion, raters noted if the emotion was cued from the face, body, and/or voice/words (e.g., sad_face, sad_body, sad_verbal). We sought to include basic-level emotions such as angry, happy, and sad in which children tend to learn earlier (Lindquist et al., 2014; Widen & Russell, 2008), as well as complex and nuanced emotions and video feature labels which may be of interest to social and emotion processing researchers. For this initial release of the EmoCodes system, however, we focus on the basic-level emotions present in the video stimuli used for system development. We intend to find appropriate stimuli to develop the other codes with and include other specific emotion codes in a future release.

Internal Consistency Analysis

We conducted internal consistency analysis of the manual codes across a team of raters and two different videos in order to determine reliability of the EmoCodes system. The codes for each video derived from this process are hereafter referred to as "internal ratings" to differentiate from the community ratings used for external validation.

Internal Rating Procedure

A team of six undergraduate research assistant raters applied the coding procedures described in the previous section to two videos: season 1, episode 2 of *All Hail King Julien*, referred to as AHKJ going forward (Owen & Stamboliev, 2014) and season 8, episode 3 of *My Little Pony*, referred to as MLP going forward (Confalone et al., 2018). These episodes were each selected for having plots that were socially driven—such as a misunderstanding or shifting social

Table 2 Subjective character codes definitions and reliability. Reliability was operationalized as the intraclass correlation (ICC) at the level of each second-by-second rating. Because ICCs require a degree of variability to be useful, the proportion of the whole video which included a non-zero value for each code in included to contextualize these findings

Code label	Definition	Possible score values	Proportion of video present		ICC (*each second, entire video*)		ICC range (*each second, 3 min windows*)	
			AHKJ	MLP	AHKJ	MLP	AHKJ	MLP
General Codes								
on_screen	Denotes if the character is visible on screen or audible	0;1	71.2%	76.1%	0.87	0.88	0.81–0.91	0.78–0.94
char_intensity	Captures broad physical arousal level of the character	0;1;2;3	68.9%	75.4%	0.83	0.83	0.76–0.89	0.67–0.87
char_valence_positive	Character is communicating broad positive affect	0;1	37.5%	31.0%	0.71	0.80	0.36–0.90	0.52–0.96
char_valence_negative	Character is communicating broad negative affect	0;1	33.6%	51.0%	0.73	0.78	0.18–0.91	0.35–0.91
Specific Emotions								
c_anger_face	A strong feeling of annoyance, hostility, or displeasure (high arousal, negative). Cued from face	0;1	7.6%	21.1%	0.56	0.77	0.19–0.88	0.33–1.00
c_anger_body	A strong feeling of annoyance, hostility, or displeasure (high arousal, negative). Cued from body	0;1	6.6%	21.8%	0.51	0.77	0.12–0.80	0.10–1.00
c_anger_verbal	A strong feeling of annoyance, hostility, or displeasure (high arousal, negative). Cued from words or sounds	0;1	4.7%	14.9%	0.62	0.75	0.31–0.85	0.54–1.00
c_excite_face	Great enthusiasm or eagerness typically in anticipation of a desired event (high arousal, positive). Cued from face	0;1	21.0%	20.8%	0.56	0.70	0.33–0.87	0.12–0.92
c_excite_body	Great enthusiasm or eagerness typically in anticipation of a desired event (high arousal, positive). Cued from body	0;1	22.7%	20.8%	0.65	0.69	0.18–0.79	0.12–0.97
c_excite_verbal	Great enthusiasm or eagerness typically in anticipation of a desired event (high arousal, positive). Cued from words or sounds	0;1	17.7%	20.8%	0.67	0.63	0.40–0.87	0.09–0.85
c_fear_face	A negative emotional state (typified as flight, fight, or freezing behavior) brought about by an explicit or implicit threat to one's physical or psychological safety (high arousal, negative). Cued from face	0;1	12.0%	7.8%	0.58	0.76	0.14–0.86	0.32–0.94
c_fear_body	A negative emotional state (typified as flight, fight, or freezing behavior) brought about by an explicit or implicit threat to one's physical or psychological safety (high arousal, negative). Cued from body	0;1	12.7%	7.5%	0.51	0.70	0.02–0.77	0.36–0.97
c_fear_verbal	A negative emotional state (typified as flight, fight, or freezing behavior) brought about by an explicit or implicit threat to one's physical or psychological safety (high arousal, negative). Cued from words or sounds	0;1	8.3%	5.2%	0.44	0.53	0.04–0.72	0.07–0.78
c_happy_face	A feeling of pleasure or contentment. To distinguish happiness from excitement, happiness in this coding scheme is a moderate to low arousal positive emotion. Cued from face	0;1	35.6%	27.2%	0.74	0.60	0.34–0.90	0.21–0.93
c_happy_body	A feeling of pleasure or contentment. To distinguish happiness from excitement, happiness in this coding scheme is a moderate to low arousal positive emotion. Cued from body	0;1	37.2%	27.5%	0.77	0.58	0.57–0.94	0.18–0.96

Table 2 (continued)

Code label	Definition	Possible score values	Proportion of video present		ICC (*each second, entire video*)		ICC range (*each second, 3 min windows*)	
			AHKJ	MLP	AHKJ	MLP	AHKJ	MLP
c_happy_verbal	A feeling of pleasure or contentment. To distinguish happiness from excitement, happiness in this coding scheme is a moderate to low arousal positive emotion. Cued from words or sounds	0;1	29.1%	24.1%	0.76	0.55	0.41–0.93	0.22–0.92
c_sad_face	A negative emotion typically brought about by unfavorable events or thoughts (high or low arousal, negative). Cued from face	0;1	10.0%	17.6%	0.86	0.69	0.43–1.00	0.10–0.98
c_sad_body	A negative emotion typically brought about by unfavorable events or thoughts (high or low arousal, negative). Cued from body	0;1	10.2%	17.8%	0.67	0.73	0.21–0.95	0.09–0.98
c_sad_verbal	A negative emotion typically brought about by unfavorable events or thoughts (high or low arousal, negative). Cued from body	0;1	6.5%	11.9%	0.68	0.78	0.40–0.93	0.28–0.98

AHKJ, *All Hail King Julien*, season 1, Episode 2; *MLP*, *My Little Pony*, Season 8, episode 3

relationship—and edited to exclude the opening and closing credits sequences. Frame-by-frame ratings were completed using Datavyu software (Datavyu Team, 2014). Raters were not permitted to discuss or share their coding with each other. Raters were permitted to ask clarifying questions to the authors of the codes (EMW and MCC) who provided semantic clarification but did not give direct guidance on what to rate. This procedure resulted in finalized ratings from 3 to 6 raters for each code—codes requiring revisions (see *Revising Internal Ratings for Consistency*) were redone by the subset of raters still available to work on this project as two students were unable to complete summer hours and one student changed laboratories.

Consistency Analysis

Internal codes were converted to timeseries and resampled to 1 Hz using the CodeTimeSeries class from the emocodes.processing.code module. Since all ratings were discrete, no further transformations were conducted on the data prior to analysis. Consistency for each code was operationalized as the intraclass correlation (ICC; Bartko, 1966), with the code at each instance entered as a separate target and raters assumed to be selected at random from the larger population (Shrout & Fleiss, 1979). In other words, ICCs were calculated as absolute agreement across a random set of raters. To examine if reliability was variable across each video, a sliding window ICC analysis was conducted (180 s windows, 20 s overlap) and the range of ICCs across windows was reported.

Revising Internal Ratings for Consistency

We compared each ICC to a threshold to determine final inclusion in the manual. Our threshold for consistency differed based on whether the code was subjective or objective. For objective codes such as the presence of faces, we required an ICC of at least 0.75 indicating good reliability (Koo & Li, 2016) for inclusion in the final manual. For subjective codes such as the presence of positive content, we required an ICC of at least 0.50, indicating moderate reliability (Koo & Li, 2016). This threshold was determined based on previous literature indicating that categorical distinctions in emotion are often subtle and that people often disagree on what emotions are present even in controlled contexts (le Mau et al., 2021). Indeed, even machine learning algorithms trained to detect differences in more naturalistically presented emotional content rarely exceed 70% accuracy on similar but out-of-sample stimuli and instead typically perform modestly above chance (Cowen & Keltner, 2017; Nwe et al., 2003; Vempala & Russo, 2018; Weidman et al., 2019). Thus, it would be prudent to average across multiple raters rather than rely on a single rater's codes, which is similar to the studyforrest coding procedures (Labs et al., 2015). Codes which did not meet this threshold criterion were reviewed with the raters as a group. If there were discrepancies in how the guidance was interpreted that could be resolved, the guidance was revised, and the codes were independently recoded. If consistency could not be reached, or the guidance could not be edited in such a way that was both reliable between raters and scientifically useful, the code was removed.

Internal Data Processing for Comparison with External Ratings

Summary scores for positive and negative intensity were computed from the internal positive, negative, and intensity codes using two approaches: (1) intensity-modulated and (2) intensity-fixed. To compute intensity-modulated scores, each affective code (positive or negative) is multiplied with the associated intensity score then summing across domains and characters. Intensity-fixed scores were computed by adding valence codes together across domains and characters without including intensity scores. Summary scores were next converted to standard units and resampled to 1 Hz for comparison with the external scores.

External Validation Analysis

We sought to test if codes derived from the EmoCodes system were consistent with emotion perceptions from members of the community. These community ratings are hereafter referred to as "external ratings" to differentiate from the internal EmoCodes ratings.

External Rating Procedure

For a more detailed description of these procedures, please see supplement section A. All study procedures were approved by the Institutional Review Board at Washington University in St. Louis and verbal consent was obtained from each participant. Participants completed personality and demographic questionnaires and rated two episodes of a children's TV show, once each for positive and negative content for a total of four runs of video rating. Ratings (positive first or negative first) and episode order (*My Little Pony* or *All Hail King Julien*) were counterbalanced across the sample. For each rating, participants were asked to use a joystick to make continuous ratings as they watched the show. Participants were given the opportunity to practice before completing the actual ratings. Importantly, each positive and negative content were rated separately, allowing for a given scene to be rated as positive, negative, both, or neither.

Sample Characteristics

Characteristics for the final sample are included in Table S1. Of the twenty-six adults recruited to this study, data from one participant was removed from analyses due to technical difficulties that resulted in a lag between the ratings, video visuals, and video audio. Thus, ratings from twenty-five adults are included in the external validation analysis.

External Data Processing and Consistency Analysis

External ratings from each participant were converted to standard units and resampled to 1 Hz. Ratings-level ICCs were computed across the twenty-five participants with each participant entered as a different random rater and the code at each second as a separate rating.

Consistency Analysis Between Internal and External Ratings

Similarity between the internal and external ratings for each positive and negative content from each episode was operationalized as Spearman's r and mean instantaneous phase synchrony (IPS). IPS was used because of the timeseries nature of the emotional codes, in which context is important for emotional inference. Thus, each code at each frame is not independent from the codes following or preceding it. IPS summarizes the similarity in phase angles between two timeseries (i.e., similarity in *when* changes in each timeseries happen rather than the absolute magnitude of the change). By nature of using the joystick versus using the frame-by-frame coding, the external and internal ratings are inherently different in their coding scales, however, which is why Spearman correlations were reported as well to complement IPS.

Low-level Feature Analysis

Low-level video features automatically extracted using the ExtractVideoFeatures class from the emocodes.processing.video module (luminance, vibrance, motion, audio loudness, audio tempo, visual salience, and visual sharpness) were correlated with external ratings to test if variation in these metrics was associated with variation in general perceptions of emotional intensity.

Results

Internal Ratings Consistency

Code-specific ICCs for the internal ratings made using the EmoCodes system are listed in Tables 1 and 2. The objective and general subjective code reliability metrics were computed after a single round of training (i.e., presenting the contents of the manual) while the emotion-specific subjective were computed after a round of revisions to the codes. This revision in procedure resulted in consistency scores presented in Tables 1 and 2. Other codes which required multiple rounds of revision were ultimately dropped due to poor reliability (please see Table S3 for a complete accounting).

With the exceptions of c_fear_verbal for the AHKJ video and close_up for the MLP video, all codes included in this initial manual release demonstrated moderate to excellent reliability (ICCs > 0.50). Example code timeseries averaged across trained raters for the AHKJ video are shown in Fig. 1. As expected, given the ICC metric, codes that were sparser across the videos had relatively lower reliability scores. The general subjective codes demonstrated good to excellent reliability across the two videos, which include character intensity (MLP ICC = 0.83; AHKJ ICC = 0.83), positive valence (MLP ICC = 0.80; AHKJ ICC = 0.71), and negative valence (MLP ICC = 0.78; AHKJ ICC = 0.73). The ICCs for the emotion-specific codes (0.44–0.86) varied more so than the general codes (0.71–0.88), perhaps as an artifact of their sparser representation within the video (general codes were present 31.0–76.1% of each video; emotion-specific codes were present for 4.7–37.2% of each video).

External Ratings Consistency

External rating ICCs ranged from 0.35 to 0.48 (AHKJ video: Positive = 0.48, Negative = 0.35; MLP video: Positive = 0.40, Negative = 0.46), indicating poor reliability. This is consistent with previous work suggesting that perceptions of emotional content vary between individuals (Cordaro et al., 2018; Feldman Barrett et al., 2006; le Mau et al., 2021; Russell et al., 2003). Mean traces are shown in Fig. 2.

Consistency Between Internal and External Ratings

Consistency between internal and external ratings was highly similar across the two internal summary scoring approaches. Spearman correlations ranged from 0.39 to 0.54 for positive and from 0.53 to 0.56 for negative. Mean phase synchrony scores ranged from 0.59 to 0.62 for positive and from 0.66 to 0.67 for negative. Table 3 reports the specific statistics for each summary approach, valence, and episode. The overlap between the external and internal ratings is shown in Fig. 3. Between these metrics, these findings indicate that the internal and external measures are collinear, capturing similar information in each video.

As a follow-up, we also tested if these similarity metrics changed when a double-gamma hemodynamic response function (6 s to peak with a 12-s undershoot) is convolved with each of the internal and external ratings. The rationale for this test was to determine if this common transformation in neuroimaging work—specifically for functional MRI or functional near-infrared spectroscopy analysis—reduced or exacerbated differences in our internal and external emotion ratings. We found that the inherent smoothing of the signal that occurs during convolution did not affect the similarity between the internal and external ratings (Positive Spearman rs = 0.43–0.54, Negative Spearman rs = 0.53–0.57; Positive IPSs = 0.59–0.62, Negative IPSs = 0.65–0.67).

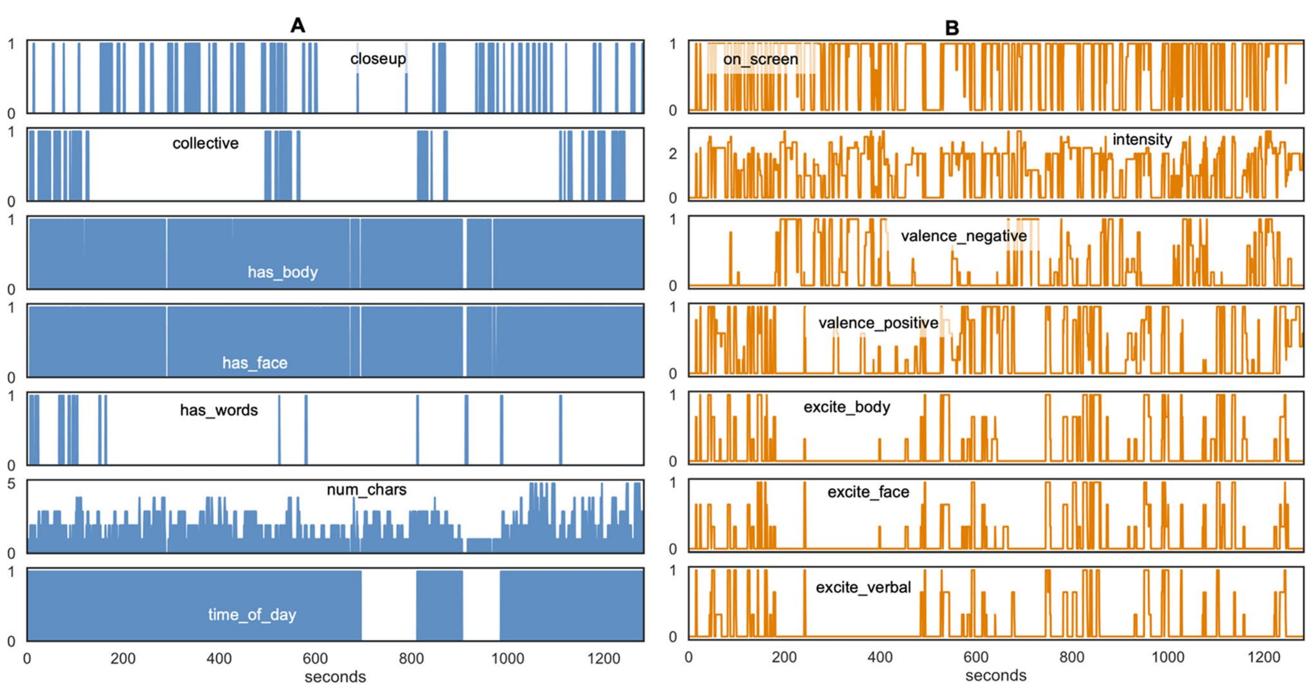


Fig. 1 Completed codes from AHKJ. **A** Area plots of each objective code after consensus across 3 raters. **B** Line plots of the mean general codes and the "excited" specific code from the Subjective Character codes set. AHKJ = All Hail King Julien Season 1 Episode 2

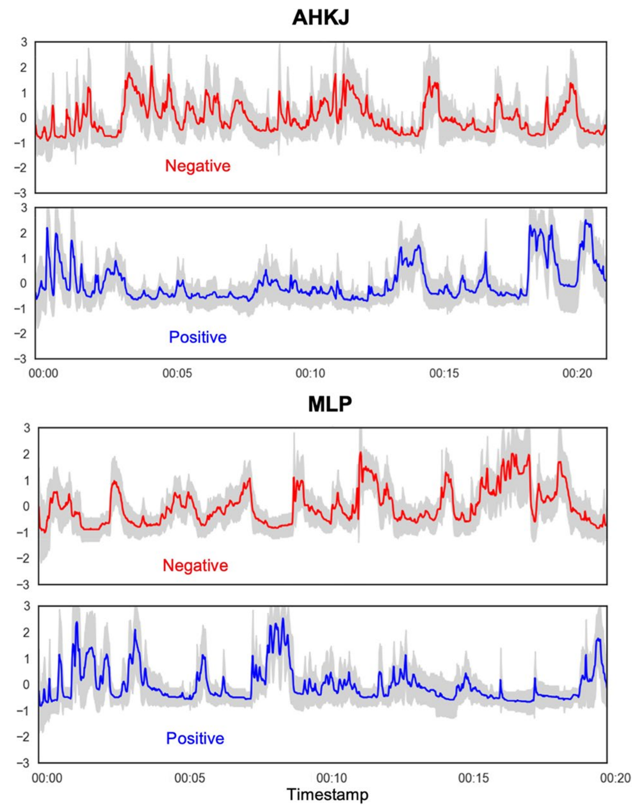


Fig. 2 Mean traces from the external, joystick ratings for each Positive and Negative and each AHKJ and MLP. The gray shading denotes + / − 1 standard deviation from the mean. AHKJ = All Hail King Julien Season 1 Episode 2; MLP = My Little Pony Season 8 Episode 3

Low-Level Feature Analysis

For the AHKJ and MLP videos separately, correlation coefficients between low-level video features and the external positive and negative ratings ranged from − 0.08 to 0.23 and from − 0.06 to 0.11 respectively. For the AHKJ video, the only correlations of greater magnitude than 0.1 were with Positive ratings and included loudness (r = 0.16), vibrance (r = 0.13), and brightness (r = 0.23). For the MLP video, the only correlations of greater magnitude than 0.1 were both with Negative ratings and included vibrance (r = 0.11) and brightness (r = 0.10). Low-level feature plots from the AHKJ video and a correlation heatmap are shown in Fig. 4. The same plots for the MLP video are shown in Figure S2. A close examination of these plots suggests that there may be specific portions of the videos during which luminance and audio loudness are each associated with Positive ratings in the AHKJ video. A similar observation can be made for the Negative ratings and the vibrance and brightness features in the MLP video. While it is clear that in these long videos there are minimal associations between emotional content and low-level video features, these close observations of the plots suggest that it is prudent to quantify and report the collinearity between low-level video features and emotional content.

Table 3 External validation analysis results. Summary positive and negative scores were compared against the corresponding mean external ratings timeseries. Each summary score was computed two ways: intensity-modulated (valence ratings were multiplied by the intensity ratings for each character before summing) and intensity-fixed (valence ratings were simply summed across characters). Statistics indicate significant overlap between changes external and internal scores (moderately high IPS) as well as reasonable correspondence between exact internal and external ratings at each second (Spearman's r). These statistics were largely the same even after convolving the ratings with the double-gamma hemodynamic response function, a common practice in neuroimaging research in order to capture a hemodynamic proxy to neuronal population activation

Video stimulus	Summary scoring approach	Positive		Negative	
		Spearman's r	*Mean IPS*	*Spearman's r*	*Mean IPS*
Raw ratings					
AHKJ	Intensity-modulated	0.39	0.62	0.56	0.66
AHKJ	Intensity-fixed	0.39	0.62	0.56	0.66
MLP	Intensity-modulated	0.54	0.59	0.55	0.66
MLP	Intensity-fixed	0.53	0.59	0.53	0.67
HRF-convolved ratings					
AHKJ	Intensity-modulated	0.43	0.61	0.56	0.65
AHKJ	Intensity-fixed	0.43	0.61	0.56	0.66
MLP	Intensity-modulated	0.54	0.60	0.57	0.66
MLP	Intensity-fixed	0.53	0.59	0.53	0.66

IPS, instantaneous phase synchrony; *AHKJ*, *All Hail King Julien*, season 1, Episode 2; *MLP*, *My Little Pony*, Season 8, episode 3; *HRF*, hemodynamic response function

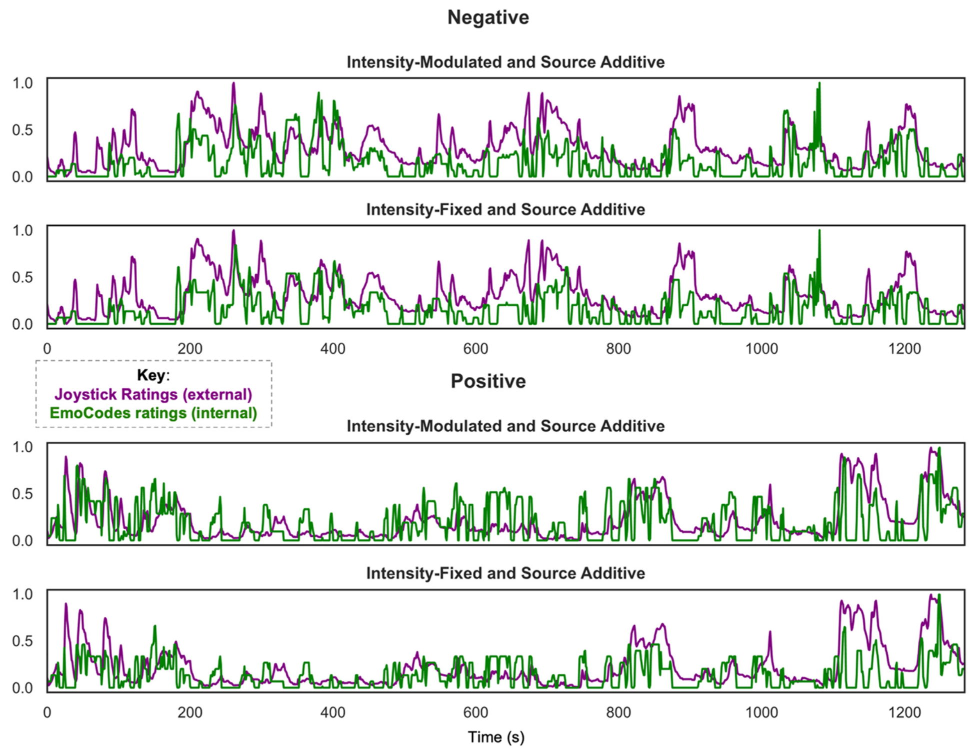
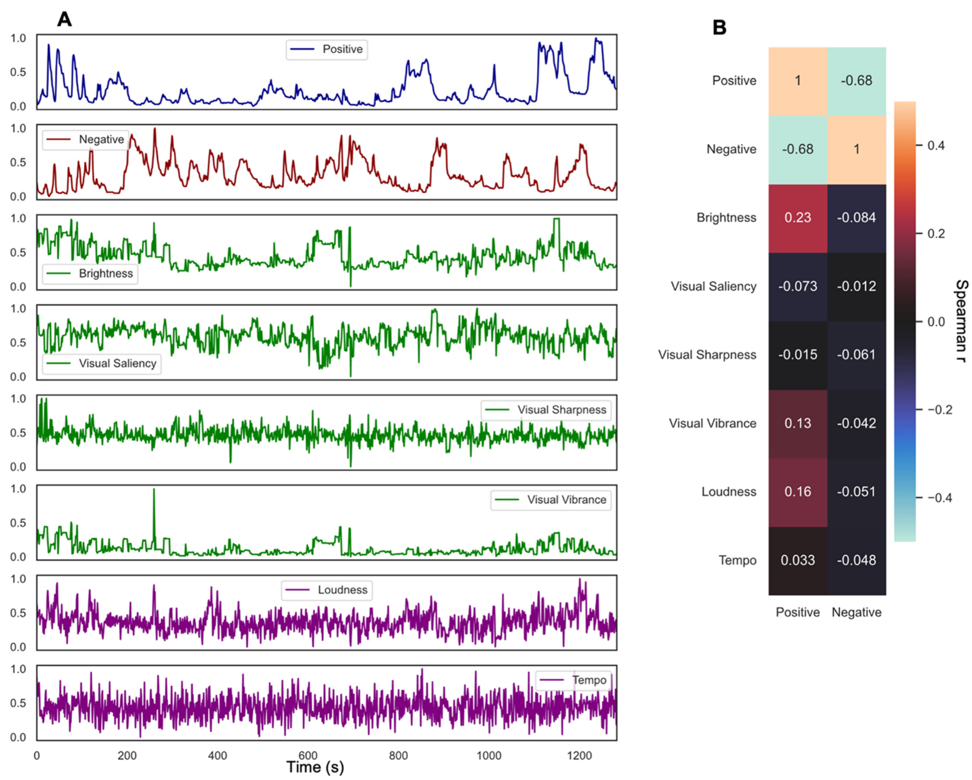


Fig. 3 External validation of the EmoCodes. Each plot below includes the external joystick ratings in purple with the internal EmoCodes ratings overlaid in green for AHKJ. The EmoCodes ratings each computed two ways: Intensity Modulated and Source Additive, in which valence ratings were multiplied by intensity ratings before being summed, and Intensity-Fixed and Source Additive, which did not include the multiplication step. Both methods provided good overlap with the external ratings (mean IPS = 0.62–0.67, Spearman's rs = 0.39–0.56). AHKJ = All Hail King Julien Season 1 Episode 2



	Positive	Negative
Positive	1	-0.68
Negative	-0.68	1
Brightness	0.23	-0.084
Visual Saliency	-0.073	-0.012
Visual Sharpness	-0.015	-0.061
Visual Vibrance	0.13	-0.042
Loudness	0.16	-0.051
Tempo	0.033	-0.048

Fig. 4 Low level video feature analysis. **A** Plots of each of the analyzed low level video features alongside the external joystick ratings for positive and negative content. Visual features are plotted in green and audio features are plotted in purple. **B** Pairwise correlations heatmap between each low-level video feature and each positive and negative external joystick ratings. AHKJ = All Hail King Julien Season 1 Episode 2

Discussion

Overview of the EmoCodes System v1.0

Here, we introduce a video coding system that combines both manual and automatic characterization of video features and emotional content. The coding manual is in two parts, Objective (nonaffective) and Subjective (affective) video coding. For Objective codes, the EmoCodes system demonstrated moderate to excellent interrater reliability at the code level between a random selection of raters. We recommend that these codes be completed by at least two independent raters and reconciled to derive final codes. For the Subjective codes, the EmoCodes system demonstrated moderate to excellent reliability with minimal formal training of the raters. The EmoCodes system also demonstrated good external reliability with minimal training. To maximize reliability, we recommend having at least two independent raters complete these codes and average their ratings to derive final codes. We also recommend that raters for subjective codes be from a similar region or culture to the participants being tested to maximize the likelihood that the ratings represent how the participants are perceiving the video. The *emocodes* Python library has functions to make each of these processing steps trivial. The library also has functions to extract low-level video features as well as to derive statistics to inform which regressors are suitable for further neuroimaging analysis. All EmoCodes training materials (videos, exemplar clips, and the group discussion board), coding materials, the data used in this analysis, and links to the python library and documentation as well as a link to join the EmoCodes group message board are available for free on our website (https://emocodes.org/). Though only a handful of emotion labels are included in this initial release, our goal is to continue to add more labels of interest to future versions of the manual as they are validated on appropriate videos. This coding system provides an important step toward reproducible affective neuroscience research using complex video stimuli by codifying how we characterize complex affective content and can be used to select specific videos before data collection begins or to better characterize videos for already collected datasets alike.

Low-level Video Features Were Not Strongly Associated with Emotional Content

Perhaps surprisingly, low-level video features such as brightness, vibrance, and auditory loudness were not associated with emotional content across the videos. Instead, most variance in the external, community-derived ratings was explained by emotional content coded from the characters themselves. This supports the notion that social context is likely more informative when perceiving emotional content rather than the non-character materials present in these videos. In other words, if overall lighting, auditory volume, or color palette were principal informers of emotional content, they would correlate with the external community ratings more closely. In considering the wide range of video content available; however, it would be wise to quantify and report the associations between affective content and low-level video features as a common practice. For example, in a movie such as *Fantasia*, in which lively animation is set to emotionally evocative music, the tempo of the music may vary more closely with shifts in emotional content than with videos that consist primarily of talking. Further work is needed to tease apart to what degree low-level features communicate emotional content and under which contexts.

Strengths of the EmoCodes System

The EmoCodes System Can Capture both Specific and General Affective Information

By including both global affective codes (valence_positive and valence_negative) and specific affective codes (c_happy_*, c_sad_*, etc.) for each source of affective information in the stimuli, the codes can be used to derive whichever level of affective information is desired for further analysis. Further, researchers who use the global positive and negative scores in research can more transparently report on the contents of the video stimuli by reporting summary statistics of the specific emotional content of the video in methods section or supplement. Considering the diverse neural activation patterns observed across an individual watching different movies (Aliko et al., 2020; Finn & Bandettini, 2021), this will likely be a crucial piece of information to include for future meta-analytic work.

The EmoCodes System Is Flexible

For the manual codes, we have minimized dependencies between specific codes to allow for each to be completed independently. Further, most codes are binary by design, enabling user to combine codes easily (e.g., taking the intersection of has_faces and closeup to isolate close-ups of faces) and therefore maximizing utility. Between the automated and manual codes, the coding system can be used to identify various facets of the video including when characters are positively or negatively interacting, close-ups of specific emotional content, mixed emotion scenes versus scenes with single emotions represented, and many more combinations.

The EmoCodes System Is Designed for Open and Reproducible Science

Both the manual and python library are version-controlled, enabling us to make changes as further research is conducted and allowing users to retain and report specific versions in their work. These resources are free for research purposes under a creative commons by attribution license. We also provide the coded stimuli on the EmoCodes website so that others can employ these episodes in their research without having to re-invest time in coding. Additionally, there is a form on the EmoCodes web page through which researchers can submit completed video codes to be hosted alongside the stimuli we are providing, enabling researchers to select videos that match their research questions before data collection and without investing time coding. We hope that this and other stimulus repositories will enhance reproducibility by reducing duplication of efforts across labs and enhancing methods transparency. Finally, the *emocodes* library can be used to generate a report (SummarizeVideoFeatures class) which can be used to easily compare the content of different videos for selection for a study or can be included in a research article supplement for transparent reporting. This report contains plots of each code and three different metrics quantifying the collinearity between each code providing full transparency of the video contents. The information provided in this report will likely be critical for future meta-analytic work in this area.

Areas for Future Improvement

We are actively working on two areas of the EmoCodes system. First, there were several emotions (for instance, disgust and shame) which did not appear in either video enough to compute consistency between ratings. Similarly, we plan to continue to test the reliability of the specific emotions which did occur sparsely in the presented videos (sadness, fear, and anger) in other media. We are in the process of seeking videos more appropriate for developing these codes with and are excited to include more specific emotion codes in future versions of the coding manual. Second, to enhance accessibility of the tools in the *emocodes* python library, we are currently in the early stages of developing a graphical user interface for the EmoCodes system. This will enable researchers unfamiliar with Python to make use of the coding and video processing tools more readily. Finally, while there is no reason to think that this coding manual cannot be applied to other videos (e.g., live action shows, feature-length movies, or home videos), we did not explicitly include these stimuli in the development of this system. We aim to test compatibility across these stimulus types in the future to harmonize across the ratings provided by the EmoCodes system with outputs from automatic tools already available for processing live action data. Further, we plan to test the full utility of the EmoCodes system for enhancing affective neuroscience research in the future, as examining neural changes in emotion processing is of great interest to the authors. Finally, we emphasize the EmoCodes system as an evolving tool, and we look forward to making additions and changes motivated by research interests and findings. Please refer to the EmoCodes website (https://emocodes.org/) for the most up-to-date version of the coding manual and python tools.

Concluding Remarks

Imaging individuals' neural activity while they watch emotionally dynamic movies is an exciting, rapidly expanding field with limitless potential. Movie-watching during neuroimaging for children is of particular interest, as neuroimaging procedures are more taxing on children than adults (Greene et al., 2018; Vanderwal et al., 2015; Yuan et al., 2009). Most existing manual and automatic movie processing software unfortunately is not designed for children's media such as cartoons, leaving many researchers to develop in-house coding schemes. Thus, the EmoCodes system provides an important step forward in standardized and replicable affective neuroscience research, particularly for developmental work. Created using an open science framework, the system is designed to evolve to incorporate new findings and meet the evolving needs of affective neuroscientists, ultimately benefiting the field.

Acknowledgements The authors thank the participants who provided the external ratings. The authors also thank the undergraduate research assistants who assisted with coding each video and pilot-tested the coding system, Jordana Avigad and Natalia Tovar.

Additional Information

Funding This work was funded by the National Science Foundation [Graduate Research Fellowship DGE-1745038 to MCC] and the National Institutes of Health [HD102156 to MCC; MH124266 to SBP; MH109589 to DMB].

Data availability All data and code are freely available on the EmoCodes website: https://emocodes.org/

Ethics approval and consent to participate Ethics approval was obtained from the Institutional Review Board at Washington University in St. Louis.

Conflict of interest The authors declare no competing interests.

Informed Consent Not applicable.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42761-021-00100-7.

Author contribution MCC: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing-Original Draft, Writing-Review & Editing, Visualization, Supervision, Project Administration. EMW: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing-Original Draft, Writing-Review & Editing, Visualization, Supervision, Project Administration. DB: Conceptualization, Investigation, Writing-Review & Editing, Visualization. RK: Conceptualization, Investigation, Writing-Review & Editing, Visualization. NS: Conceptualization, Investigation, Writing-Review & Editing, Visualization. DS: Writing-Review & Editing, Investigation, Visualization, Project Administration. SY: Conceptualization, Investigation, Writing-Review & Editing, Visualization. SBP: Resources, Writing-Review & Editing. DMB: Conceptualization, Methodology, Resources, Writing-Review & Editing, Supervision, Project Administration, Funding Acquisition.

References

Aliko, S., Huang, J., Gheorghiu, F., Meliss, S., & Skipper, J. I. (2020). A naturalistic neuroimaging database for understanding the brain using ecological stimuli. *Scientific Data, 7*(1), 347. https://doi.org/10.1038/s41597-020-00680-2

Ballester, P., & Araujo, R. M. (2016). On the Performance of GoogLeNet and AlexNet applied to sketches. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*.

Barch, D. M., Gaffrey, M. S., Botteron, K. N., Belden, A. C., & Luby, J. L. (2012). Functional brain activation to emotionally valenced faces in school-aged children with a history of preschool-onset major depression. *Biological Psychiatry, 72*, 1035–1042. https://doi.org/10.1016/j.biopsych.2012.06.009

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest, 20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports, 19*(1), 3–11. https://doi.org/10.2466/pr0.1966.19.1.3

Camacho, M. C., Karim, H. T., & Perlman, S. B. (2019). Neural architecture supporting active emotion processing in children: A multivariate approach. *NeuroImage, 188*,. https://doi.org/10.1016/j.neuroimage.2018.12.013

Confalone, N. (writer), Lu, D. (director), & Myhre, M. (director). (2018, March 31). The Maud couple [streaming show]. In S. Davis, K. Newlands, M. McCarthy, S. Wall, & N. Duboc (executive producers). *My Little Pony*. Canada: Allspark Animation & DHX Studios Vancouver. Accessed at http://www.netflix.com.

Cantlon, J. F. (2020). The balance of rigor and reality in developmental neuroscience. *NeuroImage, 216*, 116464. https://doi.org/10.1016/j.neuroimage.2019.116464

Cantlon, J. F., & Li, R. (2013). Neural activity during natural viewing of sesame street statistically predicts test scores in early childhood. *PLoS Biology, 11*(1), e1001462. https://doi.org/10.1371/journal.pbio.1001462

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion, 18*(1), 75–93. https://doi.org/10.1037/emo0000302

Cowen, A. S., & Keltner, D. (2017). *Self-report captures 27 distinct categories of emotion bridged by continuous gradients*. https://doi.org/10.1073/pnas.1702247114

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*(1), 74–101. https://doi.org/10.1037/0033-2909.115.1.74

Datavyu Team. (2014). *Datavyu: A video coding tool*. https://Datavyu.Org/. Accessed September 1, 2019.

Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in Cognitive Sciences, 20*(6), 425–443. https://doi.org/10.1016/j.tics.2016.03.014

Eickhoff, S. B., Milham, M., & Vanderwal, T. (2020). Towards clinical applications of movie fMRI. *NeuroImage, 217*, 116860. https://doi.org/10.1016/j.neuroimage.2020.116860

Ekman, P., Rosenberg, P. E. E. L., Professor of Psychology Paul Ekman, P. H. D., Rosenberg, E. L., Rosenberg, L. D. P. E. L., & Smith, M. B. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)*. Oxford University Press.

Feldman Barrett, L., Mesquita, L., Barsalou, D., Gilbert, J., Russell, J., Gross, J., Tsai, P., Niedenthal, P., Winkielman, E., Smith, C., Wright, S., Rauch, K., Ochsner, R., Lane, K., Quigley, D., Desteno, P., Devine, D., Prentice, J., Skowronski, R., … & Norem, A. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*(1), 20–46.

Finn, E. S., & Bandettini, P. A. (2021). Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *NeuroImage, 235*, 117963. https://doi.org/10.1016/j.neuroimage.2021.117963

Greene, D. J., Koller, J. M., Hampton, J. M., Wesevich, V., Van, A. N., Nguyen, A. L., Hoyt, C. R., McIntyre, L., Earl, E. A., Klein, R. L., Shimony, J. S., Petersen, S. E., Schlaggar, B. L., Fair, D. A., & Dosenbach, N. U. F. (2018). Behavioral interventions for reducing head motion during MRI scans in children. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2018.01.023

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7*(5), 555–562. https://doi.org/10.1038/nn1224

Gruskin, D. C., Rosenberg, M. D., & Holmes, A. J. (2020). Relationships between depressive symptoms and brain responses during emotional movie viewing emerge in adolescence. *NeuroImage, 216*, 116217. https://doi.org/10.1016/j.neuroimage.2019.116217

Guo, C. C., Nguyen, V. T., Hyett, M. P., Parker, G. B., & Breakspear, M. J. (2015). Out-of-sync: Disrupted neural activity in emotional circuitry during film viewing in melancholic depression. *Scientific Reports, 5*(1), 11605. https://doi.org/10.1038/srep11605

Hare, T. A., Tottenham, N., Galvan, A., Voss, H. U., Glover, G. H., & Casey, B. J. (2008). Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. *Biological Psychiatry, 63*(10), 927–934. https://doi.org/10.1016/j.biopsych.2008.03.015

Hyett, M. P., Parker, G. B., Guo, C. C., Zalesky, A., Nguyen, V. T., Yuen, T., & Breakspear, M. (2015). Scene unseen: Disrupted neuronal adaptation in melancholia during emotional film viewing. *NeuroImage: Clinical, 9*, 660–667. https://doi.org/10.1016/j.nicl.2015.10.011

Jacoby, N., Bruneau, E., Koster-Hale, J., & Saxe, R. (2016). Localizing pain matrix and theory of mind networks with both verbal and non-verbal stimuli. *NeuroImage, 126*, 39–48. https://doi.org/10.1016/j.neuroimage.2015.11.025

Jenkins, R., & Burton, A. M. (2008). 100% Accuracy in automatic face recognition. *Science, 319*(5862), 435–435. https://doi.org/10.1126/science.1149656

Karim, H. T., & Perlman, S. B. (2017). Neurodevelopmental maturation as a function of irritable temperament. *Human Brain Mapping, 38*(10), 5307–5321. https://doi.org/10.1002/hbm.23742

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kragel, P. A., & LaBar, K. S. (2016). Decoding the nature of emotion in the brain. *Trends in Cognitive Sciences, 20*(6), 444–455. https://doi.org/10.1016/j.tics.2016.03.011

Kragel, P. A., Reddan, M. C., LaBar, K. S., & Wager, T. D. (2019). Emotion schemas are embedded in the human visual system. *Science Advances*, *5*(7), eaaw4358. https://doi.org/10.1126/sciadv.aaw4358

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*, 1097–1105.

Labs, A., Reich, T., Schulenburg, H., Boennen, M., Mareike, G., Golz, M., Hartigs, B., Hoffmann, N., Keil, S., Perlow, M., Peukmann, A. K., Rabe, L. N., von Sobbe, F.-R., & Hanke, M. (2015). Portrayed emotions in the movie "Forrest Gump." *F1000Research*, *4*, 92. https://doi.org/10.12688/f1000research.6230.1

le Mau, T., Hoemann, K., Lyons, S. H., Fugate, J. M. B., Brown, E. N., Gendron, M., & Barrett, L. F. (2021). Professional actors demonstrate variability, not stereotypical expressions, when portraying emotional states in photographs. *Nature Communications, 12*(1), 5037. https://doi.org/10.1038/s41467-021-25352-6

Lemerise, E. A., & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social information processing. *Child Development, 71*(1), 107–118. https://doi.org/10.1111/1467-8624.00124

Lettieri, G., Handjaras, G., Ricciardi, E., Leo, A., Papale, P., Betta, M., Pietrini, P., & Cecchetti, L. (2019). Emotionotopy in the human right temporo-parietal cortex. *Nature Communications, 10*(1), 5568. https://doi.org/10.1038/s41467-019-13599-z

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion, 14*(2), 375–387. https://doi.org/10.1037/a0035293

McFee, B., Metsai, A., McVicar, M., Balke, S., Thomé, C., Raffel, C., Zalkow, F., Malek, A., Dana, Lee, K., Nieto, O., Ellis, D., Mason, J., Battenberg, E., Seyfarth, S., Yamamoto, R., viktorandreevichmorozov, Choi, K., Moore, J., … Thassilo. (2021). *librosa/librosa: 0.8.1rc2*. 10.5281/ZENODO.4792298

McNamara, Q., De La Vega, A., & Yarkoni, T. (2017). Developing a comprehensive framework for multimodal feature extraction. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F1296*, 1567–1574. https://doi.org/10.1145/3097983.3098075

Nwe, T. L., Foo, S. W., & de Silva, L. C. (2003). Speech emotion recognition using hidden Markov models. *Speech Communication, 41*(4), 603–623. https://doi.org/10.1016/S0167-6393(03)00099-2

Owen, E. (writer) & Stamboliev, C. (director). (2014). Poll position [streaming show]. In M. Watson & B. Haaland (executive producers). *All Hail King Julien*. United States: DreamWorks Animation Television. Accessed at http://www.netflix.com.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience, 8*(8), 636–648. https://doi.org/10.1038/nrn2195

Richardson, H. (2018). Development of brain networks for social functions: Confirmatory analyses in a large open source dataset. *Developmental Cognitive Neuroscience, 1–0*,. https://doi.org/10.1016/j.dcn.2018.11.002

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications, 9*(1), 1027. https://doi.org/10.1038/s41467-018-03399-2

Ruba, A. L., & Pollak, S. D. (2020). The development of emotion reasoning in infancy and early childhood. *Annual Review of Developmental Psychology, 2*(1), 503–531. https://doi.org/10.1146/annurev-devpsych-060320-102556

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*(1), 329–349. https://doi.org/10.1146/annurev.psych.54.101601.145102

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic stimuli in neuroscience: Critically acclaimed. *Trends in Cognitive Sciences, 23*(8), 699–714. https://doi.org/10.1016/j.tics.2019.05.004

Todd, R. M., Evans, J. W., Morris, D., Lewis, M. D., & Taylor, M. J. (2011). The changing face of emotion: Age-related patterns of amygdala activation to salient faces. *Social Cognitive and Affective Neuroscience, 6*(1), 12–23. https://doi.org/10.1093/scan/nsq007

Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C., & Castellanos, F. X. (2015). Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage, 122*, 222–232. https://doi.org/10.1016/j.neuroimage.2015.07.069

Vempala, N. N., & Russo, F. A. (2018). Modeling music emotion judgments using machine learning methods. *Frontiers in Psychology*, *8*(JAN). https://doi.org/10.3389/fpsyg.2017.02239

Weidman, A. C., Sun, J., Vazire, S., Quoidbach, J., Ungar, L. H., & Dunn, E. W. (2019). (Not) hearing happiness: Predicting fluctuations in happy mood from acoustic cues using machine learning. *Emotion*. https://doi.org/10.1037/EMO0000571

Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development, 23*(2), 291–312. https://doi.org/10.1016/j.cogdev.2008.01.002

Yuan, W., Altaye, M., Ret, J., Schmithorst, V., Byars, A. W., Plante, E., & Holland, S. K. (2009). Quantification of head motion in children during various fMRI language tasks. *Human Brain Mapping, 30*(5), 1481–1489. https://doi.org/10.1002/hbm.20616