Word Count: 9502

**Neural Effects of Continuous Ratings During Active Engagement Within a Video fMRI Paradigm**

**William J. Mitchell a,** billy.mitchell@temple.edu **\***

**Helen Schmidt a,** helen\_schmidt@temple.edu

**Chelsea Helion a,** chelsea.helion@temple.edu

# 

# 

# 

# 

**a)** **Department of Psychology & Neuroscience**

Weiss Hall, Temple University, 1701 N 13th St. Philadelphia, PA, USA 19122

**\*** Corresponding author.

*E-mail address:* billy.mitchell@temple.edu

*Address:* 717 Weiss Hall, Temple University,

1701 N 13th St. Philadelphia, PA 19122

Phone: (570) 417 – 1579

**PREPRINT**:

**CLASSIFICATIONS**: Psychological and Cognitive Sciences (Social Sciences); Neuroscience (Biological Sciences)

**Abstract (245 / 250 Words):** Continuous self-report ratings offer a high-resolution view of dynamic subjective experiences 1–3. However, concerns about maintaining experiential fidelity have largely limited studies to collecting ratings retrospectively rather than during initial viewing1,2,4. In this fMRI study, we explored how continuously rating film stimuli during initial exposure (i.e., *expressive* active viewing) affects neural activity compared to simply watching the same stimulus without explicit rating (i.e., *reflective* active viewing). Thirty-five participants watched an episode of a murder-mystery program, with one group continuously rating their certainty about a character's guilt or innocence, and another group considering the same question without providing ratings. Participants also recalled the film’s events and character judgments following the rating task. Using parametric modulation, univariate contrasts, and inter-subject correlation analyses, we identified distinct neural signatures associated with the rating process. Expressive rating engaged brain regions involved in attention, interoception, and motor control, including the anterior insula, anterior cingulate cortex, and intraparietal sulcus. In contrast, reflective engagement activated the default mode network, including the precuneus and temporoparietal junction, associated with mind-wandering and social cognition. Inter-subject correlation analysis showed greater neural synchrony among expressive raters in control networks, while reflective non-raters exhibited synchrony in attention and default mode networks. However, there were no differences between raters and non-raters in the proportion of film scenes recalled, or in socioemotional evaluations of the characters. These findings suggest that although rating alters neural activity by engaging attentional and sensory networks, it may not significantly disrupt emotional processing or higher-order social cognition.

**Significance Statement (101 / 120 Words):** Understanding how people think and feel in real-time during dynamic, realistic experiences is crucial for studying social decision-making and emotional processing. This study examines how continuously rating one’s evaluations while watching a video affects brain activity, showing that real-time ratings engage attention and sensory processing without interfering with natural emotional responses or social observation. These findings suggest that real-time self-reporting can be effectively used in brain research to capture authentic human reactions, helping bridge the gap between controlled lab studies and real-world experiences. This approach can enhance the study of complex human behaviors, including decision-making and emotional responses, in feature-rich settings.

## **Introduction**

A persistent problem in quantum mechanics 5, and some have argued, in the pursuit of science more generally 6, is something called “the observer effect.” This posits that the act of observing -- in and of itself -- changes what is being observed or measured. This tension extends beyond waves and particles and is in some respects at the core of psychological science. Is it possible to observe and examine psychological phenomena and subjective experience without substantively changing what is being observed? This presents a particular challenge in more recent years, as psychologists and neuroscientists have pivoted towards the use of more “naturalistic” stimuli 7–11 or experimental contexts 12–15 . While these situations may in some respects more adequately approximate the “real world” relative to more traditional experimental designs, they also introduce a new challenge when it comes to probing and assessing subjective psychological experience.

Tension between experimental control, ecological validity, and optimizing measurement is not new in psychology 16. Concern over the first two has motivated many researchers interested in higher-order cognitive phenomena to use dynamic, feature-rich audio/video stimuli in their research, such as films or television episodes (e.g., [8,10,11](https://www.zotero.org/google-docs/?TBhMtz)). These stimuli better mirror the social world in terms of both complexity (e.g., containing temporally-sensitive narrative structures, nuanced social interactions and emotional information) and cognitive demand (e.g., resolving ambiguities in narrative events, learning context, interpreting dynamic personal relationships and motivations) 17. For instance, a suspenseful drama film may motivate viewers to assess qualities of the various characters and their relationships and to evaluate ambiguous social cues and shifting motivations. These processes are akin to resolving real-world uncertainty in social decision-making (e.g., Is this person trustworthy? Did she mean to act unfairly?). In contrast, a controlled lab task, such as a computerized ultimatum game, isolates decision-making in a simplified, static context, offering participants limited information and clear, quantifiable choices. This approach lacks the richness and unpredictability found in natural social interactions. However, moving away from a segmented trial or block approach presents its own challenges in terms of assessing psychological experience in (roughly) real time. The goal of the present research is to examine the extent to which providing real-time ratings of psychological experience impacts subsequent neural activity, memory for complex narrative content, and higher-order social cognition (i.e., person perception).

**A room with a view: Constraining and rating psychological phenomena during film viewing.** Not all forms of film or television viewing are created equal. For example, passively watching a formulaic procedural crime drama can feel very different than being fully engaged with a suspenseful horror film or unpredictable thriller. Below, we outline two different types of film viewing examined in the present research – *expressive active* and *reflective active* – and conceptually contrast them with passive viewing, or what has been the default approach in the literature.

***Expressive active viewing.*** One solution to modeling subjective is to capture continuous, self-report ratings of a pre-defined subjective construct while engaging with dynamic, feature-rich stimuli – or what we term *expressive active viewing* (See 1,2 for reviews in the context of neuroimaging). Continuous self-report rating approaches have been used extensively beyond neuroimaging as a high-resolution representation of subjective experiences 16,18,19 (but see 3 for a review). These approaches transform a passive viewing experience into an active process by giving subjects an explicit question to consider or instructions to follow (e.g., “How amused do you feel?” in response to stimuli with differing emotional content, “How much do you trust this character?” in response to a suspenseful mystery narrative) while watching the stimulus. These guidelines likely narrow focus and circumscribe cognition 20 relative to passive viewing paradigms. They also afford researchers a window into a specific subjective assessment. Consequently, active viewing paradigms may yield greater experimental control by engaging more deliberative, top-down attention processes than passive, naturalistic viewing 21–23.

Despite the strengths and utility of in-the-moment ratings, the use of this technique has largely stagnated in the neuroimaging literature. This is due to popular interpretations of early studies suggesting that online rating alters neural activity in substantial ways 1,2,4. To date, only a handful of neuroimaging studies that we could find have attempted to capture continuous online self-reported ratings during exposure to a dynamic, feature-rich stimuli 20,24–27. We posit that the broad support for the interpretation that online ratings are inherently problematic in the context of neuroimaging studies may be overstated, as the literature often cited either did not use continuous ratings and dynamic stimuli 28,29 or contrasted significantly different viewing conditions 20,24, confounding the act of rating with differences in instruction (i.e., differences may stem from being given a focus rather than rating itself)*.* While it is true that neural activity captured while continuously rating a stimulus likely differs from passively watching a stimulus with no particular focus or goal, how the act of rating affects neural activity when the focus or goal is kept consistent has not yet been explored.

***Reflective active viewing.*** One approach for constraining cognition without introducing an active rating or updating process is to have participants view a dynamic stimulus in the scanner with specific instructions to narrow focus (i.e., 30,31), or what we term *reflective active viewing.* Through this approach, we can avoid the influence of rating behaviors upon neural data while potentially still capturing a semblance of a subject’s concurrent metacognitive phenomena. The validity of this approach outside of a neuroimaging context is supported by findings that within-participant physiological activity (e.g., skin conductance, heart rate, pulse transmission time, general somatic activity) during initial exposure is significantly correlated with the same metrics during a rewatch while self-reporting ratings 16,32.

Research using this technique has been fruitful, identifying mechanisms through which emotions promote prosociality 4 and neural correlates of both attentional engagement 31 and of humor 33. However, the majority of these studies have relied on post-exposure ratings to probe subjective experience or pooled independent ratings. This technique may be appropriate for normative or gist-level representations of complex experiences 34 but may fail to accurately capture nuances of dynamic or ongoing evaluations as they were experienced by subjects *in the moment*. For example, for a movie with a meaning-changing twist (e.g., “The Sixth Sense”), small features of an interaction or subtle information conveyed by an actor’s facial expressions or verbal tone may be imbued with a different meaning upon recall or rewatch than it was originally assigned during the initial viewing experience. This speaks to the necessity of capturing in-the-moment ratings of experience, and better understanding the implications of this approach for neural activity and downstream cognition. We propose that this type of viewing may be characterized by greater default mode network engagement (i.e., precuneus (pCUN), inferior parietal lobe (IPL), medial prefrontal cortex (mPFC)) without the added pressure of having to rate their evaluations. However, this type of viewing may also lead to greater lapses in attention, or possible forgetting of the active goal.

**Contrasting Expressive Active and Reflective Active Viewing.** In the present research, we test whether *expressive active engagement* -- or collecting in-the-moment continuous self-report ratings of one or more specific questions during initial viewing -- may be a preferable alternative to reflective active engagement. Expressive engagement may be a useful alternative precisely when reflective engagement techniques are limited: 1) when stimuli are long and/or complex, 2) when the subjective experiences we want to study are subtle, intense, or ambiguous, and 3) when retaining the fidelity of the initial response to the stimulus is more important than perhaps altering one’s attention to it. These may be true due to well-documented limitations and biases in attention, memory, and cognitive load 35,36. Maintaining phenomenal fidelity is critical when researching idiosyncratic subject-specific neural signatures (e.g., 26) or associations between subject-specific neural activity and concurrent behavioral outcomes (e.g., 24). Despite concerns regarding the alteration of neural activity, subjects who expressively engage with and passively view stimuli appear overwhelmingly similar in physiological 37 and experiential 20,38 representations of events. This suggests that active introspection imposes minimal penalties upon the fidelity of recorded phenomena, but this has yet to be tested directly in the context of naturalistic neuroimaging.

We propose that expressive active viewing will likely be comprised of two distinct categories of experience. The first is *expressive rating*, in which subjects have the option to express their ratings and an event or information did result in an update to their assessments. Expressive rating likely recruits regions associated with interoception, like the anterior cingulate cortex (ACC) and anterior insula (AI), and quantification, such as the intraparietal sulcus (IPS), in order to maintain awareness of one’s evaluations and to continually pinpoint where those evaluations exist relative to other points on a numeric continuum. The act of rating may also lead people to engage with the narrative more (i.e., superior parietal lobe (SPL)), relative to reflective viewing, to better inform their ratings. The second category is *expressive non-rating,* in which subjects have the option to express their ratings but stimulus events were presumably not salient enough to warrant a rating change (e.g., a seemingly irrelevant or uninformative action is depicted). Even when subjects are not rating in an expressive engagement design, they may be recruiting more task-related circuitry and less default mode circuitry, relative to reflective viewing, to identify events and information salient to rating. These changes in engagement may also lead expressively engaged viewers to demonstrate greater sensory processing (i.e., superior temporal gyrus (STG), occipital lobe (Occ)) and social-emotional responding (i.e., temporoparietal junction (TPJ)/posterior superior temporal sulcus (pSTS), fusiform face area (FFG)), as well.

Previous work contrasting passive and active viewing conditions while undergoing video fMRI found differences in regions associated with attention and introspection of emotions (i.e., ACC, AI, dorsomedial PFC), but reportedly no evidence of differences in regions that the authors noted were responsible for emotion responses 20. Although the studies included likely use the term in a broader sense than how we have used the term thus far, an association test (n studies = 207) of the term ‘rating’ generated using the Neurosynth 39 database found clusters in the left and right medial prefrontal cortex (mPFC), right pregenual and left subgenual ACC, right SPL, right medial temporal pole (mTP), right IPS, and left AI. These regions are common components in the default mode, dorsal attention, and salience networks, and thus, their activation during rating may represent altered levels of attention, interoception, and sensory processing.

**Leveraging inter-subject synchrony to yield new insights.** Traditional univariate contrasts among reflective non-rating, expressive non-rating, and expressive rating would highlight differences in the average magnitude of neural activity that subjects recruit when engaging reflective, focused, and evaluative cognitive states, respectively. However, expressive and reflective engagement may also differ in how consistently neural activity fluctuates over time. As “naturalistic” stimuli has become increasingly more popular in the literature, Expressive raters may demonstrate greater synchrony than reflective non-raters in interoceptive (i.e., ACC, AI) and quantitative (i.e., IPS) regions. If engagement does differ between expressive raters and reflective non-raters, we may also expect less synchrony from expressive raters than reflective non-raters in default mode network associated regions (i.e., pCUN, IPL, mPFC).

**Examining the downstream impact of expressive and reflective active viewing on higher order cognition** To examine the impact of expressive active and reflective active viewing on cognition, it is necessary to collect measures that assess diverse aspects of information processing. Here, we focus on two — memory formation and person perception — that have been extensively studied in the literature. Turning first to memory formation, prior work using naturalistic stimuli has largely focused on passive viewing (i.e., viewing without rating or a concrete goal in mind while viewing), occasionally paired with free recall, or a period wherein participants are tasked with recalling as much of the stimuli as they can, and in whatever order they can. This work has been illuminating in terms of identifying how complex information is structured in memory 40, the extent to which narrative comprehension is shared across individuals 8, and the extent to which neural activity during experience is predictive of neural activity during recall 41. Thus, in the present research, we utilized a free recall paradigm to assess whether there were between subject differences in the proportion of scenes recalled between rated (i.e., expressively viewed) and un-rated (i.e., reflectively viewed) film segments. If there are differences in the proportion of scenes recalled between rated and unrated film segments, this would indicate that the act of rating (rather than just considering an active viewing goal) in and of itself may influence how information is attended to, processed, and later retrieved. We predicted that there would not be differences in the proportion of recalled scenes between the two conditions, given the presence of the same active goal — rating certainty in a social outcome — in both viewing conditions.

Turning next to person perception, we focused on this construct as building a representation of another person is a multi-faceted process. Some work in social psychology indicates that individuals can get a “gist” of an individual relatively quickly 42, and that this gist is predictive of assessments made on longer time scales (i.e., those where individuals have access to more information to integrate into their assessment). Other work has instead argued that attitudes towards others can shift based on relevant information, and that individuals are actively engaged in a cognitive updating process when it comes to assessing others’ traits 43,44. Taken together, this indicates that arriving at a comprehensive assessment of what individuals are like requires some element of observation and information integration, a process that may be disrupted by the act of rating. If there are no differences between reflective and expressive viewing, this could indicate that stable assessments of others are made even in the presence of both an active goal (both reflective and expressive viewing) and when individuals are tasked with actively tracking their position on that goal (expressive viewing).

**The Present Research.** To test these hypotheses directly, in the present research, participants watched video stimuli (a television episode) while being given instructions (evaluate their certainty that a character was guilty or innocent of committing a specific crime). In one half of the episode, participants did not give explicit ratings related to the evaluation (i.e., reflective active viewing). For the other half, participants gave explicit ratings for the evaluation using a bipolar, horizontally-positioned scale positioned below the video stimulus, initially set to 0% certainty, with -100% reflecting complete certainty about the character’s guilt and +100% indicated complete certainty in the character’s innocence. The order of the viewing condition varied across participants, such that some participants expressively viewed (i.e., rated) the first half and reflectively viewed the second, or vice versa. As such, we were able to more directly isolate the neural effects of rating than the previously noted works, both within- *and* between-participants. Following the episode viewing, participants were instructed to freely recall the episode contents while still in the scanner (**Figure 2**). To assess potential differences in person perception as a function of expressive viewing, a subset of participants rated the main characters on a number of dimensions that have been previously used in the literature to assess complex individuals 45. We employed complementary analytic approaches - parametric modulation, whole-brain univariate contrasts, and inter-subject correlation (ISC) analyses - to examine the neural effects of continuous rating. Parametric modulation analysis enabled us to examine how variations in rating activity modulated neural activity. Univariate contrasts allowed us to identify specific brain regions which demonstrate differential activation when different cognitive states are engaged. An ISC approach revealed the consistency of neural temporal dynamics across subjects (i.e., changes over time) within these regions, highlighting shared cognitive processes and temporal dynamics. By integrating these methods, we addressed different aspects of the data to match the multidimensional nature of the task and stimuli. This comprehensive approach enhances the reliability and depth of our findings and provides a comprehensive understanding of the neural mechanisms underlying subjective rating.

**RESULTS**

**Behavioral task performance**

**Participants in both viewing orders exhibited similar levels of rating frequency.** No significant differences were observed between run 1 (mean Run 1 = 22.6 ± 22.7 button presses) and run 2 (mean Run 2 = 25.9 ± 27.6 button presses) regarding the average volume of buttons presses per subject (95% CI = (-21.3 , 14.6), t(27) = -0.4, p = 0.7). On a questionnaire which followed the primary task, all subjects indicated that they felt that they fully understood the task instructions. Only a single subject endorsed having difficulty using the button device, but the cited issue occurred during a task unrelated to the current analysis.

**Subjects reported high engagement and plot comprehension.** To identify potential impediments to stimulus engagement and comprehension of our stimuli given its length and narrative complexity, we collected a series of post-scan self-report measures about participants’ viewing experiences. On a 5-point scale (range = 0-4), ranging from ‘Not at all’ to ‘Extremely’, the median score of all participant ratings indicate that the task was viewed as “very engaging” (one-sample t-test: mean = 2.5, 95% CI = (2.2 , 2.7) , t(33) = 18.3, p < 0.001), that the plot was “not at all difficult” to follow (one-sample t-test: mean = 0.41, 95% CI = (0.15, 0.67), t(33) = 3.2, p = 0.002), and that the audio was “not at all” difficult to understand (one-sample t-test: mean = 0.29, 95% CI = (0.13 , 0.46), t(33) = 3.7, p < 0.001 ). Engagement and plot comprehension difficulties were strongly negatively correlated (r(32) = -0.55, p < 0.001).

**Scene recall did not differ between rating conditions.** Twenty-eight (n = 28) of the thirty-five subjects used in neural analyses completed post-exposure free recall of the stimulus which was able to be transcribed and analyzed. A paired samples t-test found no differences (mean difference = 0.04) in the magnitude of scenes recalled within subject while expressively rating and reflectively not-rating the video (95% CI = (-0.03 , 0.11), t(27) = 1.17, p = 0.3). We could alternatively use a chi-square test to determine if there is a relationship among the proportion of scenes recalled and not recalled while expressively rating and reflectively not rating between subjects. This approach also failed to identify significant differences (χ2(1, N = 28) = 0.85, p = 0.36). It is possible that relying solely on summary statistics to detect differences in multivariate data may overlook more nuanced patterns. To address this, we constructed dissimilarity matrices using Spearman rank correlations across subjects based on scene recall. We then employed bootstrapped permutation testing (n = 5000) to assess whether the strength of these correlations varied depending on which half of the session subjects engaged in expressive rating versus reflective non-rating. Our analysis revealed no statistically significant difference between the rating conditions (mean difference = -0.02, p = 0.20).

**Character representations did not differ between rating conditions.** We again used dissimilarity matrices to determine whether rating condition yielded observable differences in character assessments. Using bootstrapped permutation testing (n = 5000), we did not find any statistically significant differences between the rating conditions (mean difference = 0.01, p = 0.56).

**Neuroimaging results.**

**Increased rating frequency was positively associated with activation of neural circuitry implicated in sensory integration, attention, and self-monitoring.** We used parametric modulation to identify regions sensitive to variability in rating behavior. The frequency of our rating proxy (i.e., button presses per TR) was used as a regressor applied to data from each subject’s expressive engagement run. This revealed significant activation clusters, primarily in the left hemisphere (**Figure 3**). Notable activations included the left postcentral gyrus (PoCG) extending into the precentral gyrus (PrCG), the anterior cingulate cortex (ACC) extending into the mid-cingulate cortex, the Rolandic operculum (ROL) extending into the supramarginal gyrus (SMG), and the supplementary motor area (SMA). Additional activations were observed in the right cerebellum (lobules 4 and 5), the left inferior parietal lobule (IPL), and the left anterior insula (AI). All cluster peak activations were contained within either the somatosensory motor network, salience/ventral attention network A (SVAN A), or dorsal attention network B (DAN B) under the Schaefer-Kong functional parcellation schema 46,47. The clusters observed suggest that rating frequency modulated activity in regions associated with attention and sensory integration (dACC, IPL, ROL), motor control (PoCG, SMA, Cereb), and self-monitoring (dACC, IPL, AI).

**Expressive rating elicits greater activation from attention, sensation, and control regions relative to expressive non-rating and reflective viewing.** To examine the effects of rating during expressive viewing, we conducted two types of contrasts: 1) a contrast comparing rated TRs while expressively viewing to non-rated TRs while expressively viewing (i.e., within subject, within the same stimulus run), and 2) a contrast comparing rated TRs while expressively viewing to non-rated TRs while reflectively engaged (i.e., between subjects, within the same stimulus run). Contrasting all three task components allows us to identify which neural circuitry is engaged when task demands are more (i.e., expressive non-rating) or less (i.e., reflective viewing) similar to expressive rating, thereby offering insights into how the act of explicitly rating subjective experience modulates brain responses when the viewing goal is held constant (**Figure 4**).

Both contrasts indicated significant activations primarily in parietal, frontal, and occipital regions, but more extensive frontal activation was observed within the expressive rating – reflective viewing contrast. Major activations in the expressive rating – reflective viewing contrast included the left superior parietal lobule extending into the inferior parietal lobule, the right superior parietal lobule, and the right angular gyrus. Additional clusters were observed in the left middle occipital gyrus, right supramarginal gyrus, bilateral inferior frontal gyri, and right insula.

Key clusters of the expressive rating-expressive non-rating contrast included the left inferior parietal lobule and the right inferior parietal lobule extending into the supramarginal and angular gyri, the right dorsolateral prefrontal cortex, and the superior parietal lobules bilaterally. Other notable activations were found in the right middle occipital gyrus, middle and inferior temporal lobules, right insula and inferior frontal gyrus, and bilateral anterior cingulate cortex. Major activations in the expressive rating – reflective non-rating contrast included the left superior parietal lobule extending into the inferior parietal lobule, the right superior parietal lobule, and the right angular gyrus. Additional clusters were observed in the left middle occipital gyrus, right supramarginal gyrus, bilateral inferior frontal gyri, and right insula. While precise spatial coordinates varied slightly between contrasts, almost all regions activated by the expressive rating - expressive non-rating contrast were activated by the expressive rating - reflective viewing contrast. However, expressive rating – reflective viewing contrasts uniquely indicated activation in the bilateral fusiform gyri, bilateral hippocampi, and motor regions such as the supplementary motor area and precentral gyrus, none of which achieved significance in the expressive rating - expressive non-rating contrast. Taken together, these results indicate recruitment of attentional, sensory, and motor processes during rating relative to non-rating, even in the context of a similar viewing goal.

**Reflective viewing, relative to expressive rating, elicits greater activation from default mode network.** When examining regions which demonstrated greater activation while not rating, we identified a similar pattern: both contrasts indicated significant activation in default mode network regions, but the reflective viewing -expressive rating differences were more extensive and robust (**Figure 5**). We specifically observed engagement of the bilateral precuneus (pCUN), cuneus (CUN), calcarine cortex, temporoparietal junction (TPJ), middle temporal gyrus (MTG), the temporal poles (TP), and superior temporal sulcus (STS) across both designs. We also found increased activation in regions associated with sensory processing (see Supplemental Materials for full activation details).

**Reflective viewing, relative to expressive non-rating, recruited greater default mode network activation**. The two types of non-rating behavior captured within this study may represent fundamentally different cognitive phenomena. In the former, subjects were able to more passively consider the target question without having to continuously provide any clear signal of their underlying cognitive activity, thus activity may be relatively more heterogeneous. The latter consists of events or periods that subjects determined to be insignificant by not changing their ratings, which should thus consist of relatively more homogeneous cognitive states. To examine how reflective and expressive engagement alter viewing experiences, we contrasted non-rating activity during reflective viewing runs with non-rating activity during expressive viewing runs.

Subjects who reflectively watched the video stimuli demonstrated activation clusters of a greater magnitude in the right temporoparietal junction, right cuneus extending into the precuneus, right insula extending into the inferior frontal orbital cortex, right dorsolateral prefrontal cortex extending into the middle frontal gyrus, and right inferior parietal lobule extending into the supramarginal and angular gyri. In contrast, the inverse comparison revealed a single significant cluster in the left angular gyrus which is considered part of the default mode network A under the Schaefer-Kong functional parcellation schema 46,47. Although many of the structures which appeared significant are typically also considered part of the default mode network (i.e., pCUN, mPFC, IPL), under the Schaefer-Kong functional parcellation schema 46, their peak activations were within control networks B and C as well as part of the salience / ventral attention network B, in the case of the IPS. Regardless, these results (**Figure 6**) may indicate that reflectively engaged watchers demonstrated greater activation of traditional default mode network structures (pCUN, mPFC, IPL) than expressive watchers, even when the latter were not actively providing ratings.

**Having the opportunity to rate may be a distinct mental state, even in the absence of explicitly rating.** Next, we aimed to assess whether being able to rate subjective experience evoked different patterns of neural activity, regardless of whether one was actively rating at the time. To do so, we matched activation clusters in the reflective viewing – expressing rating and expressive non-rating – reflective viewing contrasts by peak voxel coordinates. 11 default mode clusters from the expressive non-rating – expressive rating contrast corresponded with 14 in the reflective viewing -expressive rating contrast. However, only two auditory and one language cluster from the expressive non-rating – expressive rating contrast had counterparts, compared to seven auditory, five visual, and two language clusters in the reflective non-rating – expressive rating contrast. Both contrasts also showed activations in the ventromedial prefrontal cortex (vmPFC), though these activations were again more extensive in the reflective viewing - expressive rating contrast. These findings underscore the consistent involvement of the default mode network in periods of passive engagement. However, these findings also suggest that being in the mental state of rating in and of itself can produce differences in sensory processing, relative to reflective viewing, even when the physical act of rating is not actively happening.

**Rating was associated with increased inter-subject synchronization in control networks, while non-rating was associated with increased synchrony in attention and default mode networks.** The results of our ISC analysis, which examined intra-condition synchrony during expressive rating and reflective viewing, followed trends seen in previous analyses (**Figure 7**). When subjects were reflectively engaged with a stimulus, they demonstrated significantly greater synchrony (i.e., neural dynamics) than expressive raters in the right pCUN (Schaefer-Kong parcellation 225 of 400), and bilateral TPJ (Schaefer-Kong parcellations 108, 311, and 337 of 400). Within the Schaefer-Kong defined functional networks, these regions are part of the default mode (B), salience and ventral attention (B), and auditory networks. However, expressive raters demonstrated greater synchrony than reflective non-raters in the left AI (Schaefer-Kong parcellation 56 of 400) and right IPS (Schaefer-Kong parcellation 248 of 400). Both are considered part of the control network (A).

**Discussion**

The present study aimed to characterize how neural activity differed while continuously rating or not rating a video stimulus under otherwise identical instructional conditions and focal topics. We sought to answer this question comprehensively, examining: 1) differences in average activity magnitude between discrete engagement states, 2) variations in activity that correlated with variations in rating behavior, and 3) differences in intragroup neural synchrony. In doing so, this study extends prior results examining studies that exclusively examined passive viewing (cites), or contrasted passive and expressive viewing without maintaining goal congurency (Hutcherson cite). Based upon this previous work and supplemental priors from an automated meta-analysis, we expected rating to demonstrate increased activity in circuitry associated with X (), Y (), and Z (). ACC, AI, IPS SPL, STG, Occ, TPJ and FFG and show more consistent activity in ACC, AI, and IPS. We also expected reflective non-rating to demonstrate increased activity and more consistent activity in pCUN, IPL, and mPFC. We also expected that being in the mental state of rating – even when not physically doing so – would be associated with increased activation from attention and salience regions.

In line with our hypotheses, we found that rating did recruit greater activation from IPS, SPL, Occ, and FFG, relative to reflective viewing, while expressive non-rating recruited greater activation in IPL, pCun, and vmPFC relative to expressive rating. However, contrary to our hypotheses, we did not find differences between the two conditions in activation of the ACC, AI, or STG. Additionally, we found that non-rating elicited greater pSTS / TPJ activation relative to rating, which was the opposite of what we predicted. Expressive raters did demonstrate more similar activity to one another over time than reflective viewers in portions of the IPS and AI, while reflective viewers demonstrated more intragroup similarity in pCUN and TPJ. No significant group synchrony differences were observed in IPL, mPFC, or ACC, contrary to our hypothesis. While no regions were consistently recruited across all parametric, univariate, and intersubject correlational analyses, portions of the right MTL, bilateral TPJ, right IPS, and right pCun were active in both the univariate and intersubject correlational analyses, though the exact localization of activation clusters differed slightly from analysis to analysis. Left AI was only active in the parametric modulation and intersubject correlational analyses. In general, rating appeared to differentially recruit control, attention, salience, memory, and visual networks while not rating appeared to recruit default mode and auditory networks, with some additional recruitment from visual and salience structures not recruited by rating.

**Interpretation of Results.** Identifying the networks associated with activated neural regions during expressive and reflective engagement enhances our understanding of the specific cognitive and emotional processes involved, links brain activity to behaviors, and improves the ecological validity and applicability of our findings. One interpretation of these results is that rating demanded complex problem-solving and error monitoring via top-down, voluntary attention for sustained periods of time in search of unexpected, decision-relevant stimuli. Reflective viewing instead potentially generated more mind-wandering, mentalizing, and may have promoted broader sensory processing while noting unexpected, decision-relevant stimuli.

Many of these results largely aligned with earlier work contrasting passive viewing with expressive viewing 20. However, the inclusion of the reflective viewing, or holding a consistent goal across both rating and non-rating runs, yielded a distinct pattern of results. For example, activation of the TPJ, TP, and pCUN may suggest differences in social cognition as a product of maintaining an active viewing goal. The extent to which this may be a consequence of social task demands is unclear. Our task demands differed from Hutcherson et al.’s in that our designated focus was an explicitly social question (i.e., subject’s certainty of the guilt or innocence of a character), which may differ from an explicitly emotional question. When assessing a person’s guilt or innocence, there is theoretically a correct answer that one could ascertain from the information provided. As such, there may be more to gain from focusing greater attention upon events to identify relevant, potentially obscured, information and more carefully analyzing character interactions. This might not be equally true for emotions. One’s emotional reaction to the same events, however, may be comparatively subjective and less informed by careful attention or problem-solving. However, more studies analyzing the effects of rating are needed to conclusively determine the effect that different task demands have upon neural activity while rating.

**Additional Findings.** A few unanticipated-but-observed trends might provide additional insight into the cognitive differences underlying expressive and reflective viewing. For example, bilateral hippocampi were significantly activated when rating was compared to reflective, but not expressive, non-rating. This may further support the notion that expressive viewing facilitates deeper engagement with stimulus plot as subjects encode or recall episodic events to inform their ratings, even when not actively changing their ratings. Upon seeing this finding, we examined scene recall accuracy from a surprise free recall task that followed the stimulus to examine memory differences could be observed depending upon whether subjects expressively or reflectively viewed a scene. We did not find a significant difference in the proportion of scenes recalled from expressive viewing runs (M, SD) and reflective viewing runs (M, SD). (*x*2 (1, N = 980) = 1.1004, *p* = 0.294). Though we could not demonstrate memory differences by condition with this analysis, scene recall variability was overall low, which may limit our power to detect significant differences.

Another noted trend was the activation of distinct areas of the control and salience networks within the expressive rating and non-rating contrast, but not within the expressive rating and reflective viewing contrast. This suggests the need for flexible behavioral regulation under changing conditions and bottom-up salience detection when tasked with explicitly rating subjective experience 48,49. This may represent categorically different attentional engagement, as subjects critically evaluate a continuous stream of relevant and irrelevant details to inform their evaluations to an extent not present when ratings do not need to be quantified. However, contradictory evidence may be observed in the non-rating and rating contrasts. As previously noted, most regions activated within the reflective non-rating and expressive rating contrast were also activated within the expressive non-rating and expressive rating contrast with the noted exception of sensory processing regions. Elevated activation in sensory regions might be interpreted as more focused or broader attention to audio or video features of the stimulus while reflectively, not expressively, viewing 22. However, design is confounded with condition (i.e., all expressive-expressive contrasts are within-subject and expressive-reflective contrasts are between-subject) which complicates interpretations of sensory motor differences specifically, as previous studies have shown that sensory-motor regions exhibit low inter-subject but high intra-subject functional connectivity variability relative to association regions 50–52. As such, statistically significant activation under between subject contrasts, but not within subject contrasts, may simply reflect reduced between-subject activation variability.

We hypothesized that expressive rating would yield significantly greater dACC activation than reflective non-rating, under the assumption that it would reflect increased task demands for decision-making and error detection in subjects' personal hypotheses. While we failed to observe this between expressive rating and reflective non-rating, we did observe robust activation of the dACC between expressive rating and non-rating, which may better fit with our hypothesis than initially expected. As previously noted, reflective viewing is a somewhat heterogeneous signal. Any given moment of reflective viewing, whether the subject is experiencing high or low error detection, is behaviorally indistinguishable from any other because we lack ratings (though, it may at some point prove possible to rely upon some yet-undiscovered reliable physiological or neural signature of this type of cognition). On the other hand, periods of expressive non-rating reflect an evaluative process which determined that nothing of sufficient salience had occurred in the past moment. This acts as a much starker contrast to some aspects of expressive rating's neural profile due to greater homogeneity in both conditions. Additional support for this interpretation comes from the parametric modulation analysis which found that increased rating behavior positively correlated with increased dACC activity. The contrast between reflective and expressive non-rating found greater activation of some task-related regions, including the dPFC, IPL, and SMG, under reflective viewing, which also supports this interpretation. This ACC activation also differs notably from what the Neurosynth meta-analysis suggested might activate while rating, as the portions of the ACC that were active in that analysis (i.e., pregenual and subgenual ACC) would reflect altered emotional responses 53. In line with prior work 20, we have little evidence to suggest substantiative differences between emotion responses while expressively and reflectively viewing a stimulus.

**Relevance to Other Research.** By highlighting regions largely affected and unaffected by continuous rating, we have contributed to a shared understanding of the strengths and weaknesses that different approaches to study social and affective neuroscience possess. These results emphasize that activity in many higher-cognition association regions and limbic structures outside of attention networks remain relatively unaffected when rating, while attention and sensory processing appear most widely altered. In other words, we found little evidence to suggest that active online rating alone substantively alters emotion responding or higher cognition, though it may alter what we pay attention to and how engaged we are with the stimulus.

If altered neural activity in a specific region is a substantial concern when using this approach, it may be possible to regress out rating-related activation during pre-processing using a study design similar to what we had used here (i.e., contrasting a rating and non-rating group to the same stimulus). Such statistical adjustments may better isolate the neural responses specific to the social and affective processes they intend to target. These results and this approach may also be of interest to researchers developing computational models of dynamic decision-making, as rating-related activation may be a variable that they wish to represent. Lastly, these results may enhance ROI selection, as the regions we identified as being differentially activated by rating or not rating could inform the selection of regions of interest in future studies on dynamic decision-making.

**Limitations.** Several limitations should be acknowledged in light of the findings of this study. First, the sample size of 35 participants, though adequate for our analyses, may limit the generalizability of our findings. Although we aimed to balance the size of the sample in each condition, fewer subjects were in the reflective-expressive engagement ordered condition (n=15) than in the expressive-reflective engagement ordered condition (n=20) due to the previously noted exclusions. The choice of stimulus, a single television episode, also constrains the generalizability of our findings. Different mediums, genres, emotional tones, or narrative complexities might elicit distinct neural activation patterns 54. While the duration of our selected stimulus is in some ways a strength of the study, as it allowed for greater narrative complexity, it also limited the quantity and type of stimuli which we were able to test this behavior within. The stimulus choice also affected the rate of rating changes. The average number of button presses per minute was only 1.01 and 1.16 in the first and second halves of the episode, respectively, with standard deviations of 8.07 and 8.15 button presses per minute, respectively. While this may simply be representative of the relatively slow speed with which social information is often shared, perceived, and processed 55, tracking another metric, such as certainty of luminance changes in a video, may result in much more varied and rapid rating behaviors, thus potentially increasing the ability of our analyses to discern meaningful neural signals.

Functional MRI itself has inherent limitations, including temporal resolution constraints 56. Although ratings sometimes changed and were sampled at a high rate, all behavioral data must be down sampled to match the imaging rate, or TR, of fMRI, which results in a lower resolution model of complex phenomena due to limits on the temporal precision of our tools. Although the framewise displacement reported by subjects suggested minimal head motion issues, fMRI’s susceptibility to motion artifacts represents another factor limiting the precision of our neural data 57. Additionally, while using dynamic, feature-rich video stimuli may be *relatively* more naturalistic than other approaches to study social and affective phenomena, an MRI still represents a fundamentally artificial environment. Although our stimuli and task mirror some aspects of social observation, they may be less directly social than, for example, dyadic interactions 16,58–60.

The absence of additional comparison tasks, such as a passive viewing condition or a task which elicits high cognitive demand but which is not expressive viewing, limits our ability to isolate neural correlates specific to rating from those related to general cognitive and sensory processing. Also, although our design was able to reduce the confounds between instruction and rating behavior, our univariate contrasts still unfortunately confounded individual variability with condition, as expressive-expressive contrasts were conducted within-subjects, whereas expressive-reflective contrasts were conducted between-subjects. Lastly, the use of a button-box and a 5-pt increment scale may constrain the granularity of participants' responses. More nuanced or continuous rating options could potentially provide richer datasets, capturing finer details of subjective experiences.

**Future Directions.** Future endeavors can build upon these findings in a few ways to promote greater ecological validity in neuroscience research. While past research provided direct comparisons between passive viewing and what we have termed expressive viewing, we are unaware of any direct comparisons between reflective viewing and passive viewing. This comparison may be of interest as reflective viewing could represent a means of reasonably constraining attention and adding more internal validity to naturalistic studies while minimizing the added activation we observed in attention, interoception, and sensory regions during expressive viewing.

It would also be valuable to explore a broader range of stimuli using expressive viewing designs, including varying mediums (e.g., video, audio, text), genres (e.g., comedies, dramas, documentaries), emotional tones (e.g., happy, sad, suspenseful), and narrative complexity, as continuous online ratings may prove to be more or less obtrusive depending upon these factors. Collecting continuous ratings in different contexts would also help to identify whether the observed neural patterns are specific to certain types of stimuli or generalizable across different media. Tasks requiring continuous ratings of non-social aspects, such as visual or auditory features, could help distinguish the neural activity associated with social evaluation from that related to general cognitive and sensory processing.

While this study suggests that expressive engagement alters attention, what remains unclear are the specific details of how attention is altered. Are subjects fixating more on characters? Are they scanning scenes more comprehensively? The analyses and tools that we used here cannot conclusively inform these questions, but complementary tools, such as eye-tracking, employed concurrently with neuroimaging while subjects reflectively and expressively engage with a stimulus, could answer these questions by tracking and comparing gaze behavior 10. Additionally, employing techniques with higher temporal resolution, such as electroencephalography (EEG), could capture rapid changes in neural activity at a rate more commiserate with vision process than fMRI 33. This approach would complement fMRI findings by providing a more detailed temporal profile of the cognitive processes involved in continuous rating. Using a more mobile imaging technique may also allow for the incorporation of more immersive and realistic experiential environments.

Lastly, as the use of naturalistic and feature-rich experimental designs in neuroimaging steadily increases, the value of a formal taxonomy capturing the diversity of these experimental paradigms grows as well. This may be especially beneficial to social and affective neuroscientists, who are increasingly turning to such designs to model their phenomena of interest but whom may experience challenges attempting to identify and build upon existing work. We experienced this ourselves in the development of this manuscript, as disparate literatures with shared interests in continuous rating (e.g., social psychology, neuroeconomics, computer science, etc.) appeared disconnected by differences in terminology and tools. A substantial body of work in this space has already been created. A shared formal taxonomy may reduce siloing of research efforts by creating a common language and provide a robust foundation for this thriving subdiscipline.

**Conclusion**

In this study, we directly compared neural activity of subjects while they either continuously rated or did not rate their evaluations of a specific subjective topic, thus separating rating behavior from differences in task-related instruction. In line with previous research comparing continuous online rating to passive viewing, we found that expressive engagement elicits greater activation and more similar neural dynamics in regions associated with attention, sensory integration, and self-monitoring. Unlike previous research which task subjects with rating emotions and found no differences in emotion responding regions, we also found some evidence to suggest that social processing regions (i.e. FFG, TPJ, TP, pCUN) did demonstrate differential activation as a consequence of this change in focus. This is likely a natural consequence of the demands of the task: having a constant visible reminder of your goal likely motivates subjects to focus more closely on details to inform their ratings than they might otherwise. Nonetheless, these findings underscore the importance of carefully considered study design and the variety of options available to neuroimaging researchers interested in incorporating more dynamic, feature-rich stimuli into their projects.

## **Materials and Methods**

**Participants.** Forty (40) subjects were recruited for a neuroimaging study on decision-making from the greater Philadelphia area between May 2022 and June 2023. Five subjects were excluded for reasons including excessive head motion (1), prior familiarity with the stimulus (1), and technical issues resulting in incomplete data (3). The 35 remaining subjects (N female = 20, N male = 15) ranged in age from 18 to 44 years (median age: 22 years; mean age: 24.5 ± 5.5 years). Eleven (11) subjects reported never having had previous MRI experience before and 5 reported having been imaged five or more times. All other subjects varied in their levels of past MRI experiences. Approximately 54.3% of our sample identified as non-Hispanic white, 5.7% white of Hispanic origin, 31.4% Asian, and 8.6% Black. All participants possessed normal or corrected-to-normal visual acuity, were not color blind, and were free of significant psychological, neurological, and developmental disorders. We did not use handedness as eligibility criteria, but did document it and use it as a regressor where appropriate. Four of the thirty-five eligible subjects identified as left handed and all others were right handed. All participants provided written informed consent as approved by a local Institutional Review Board.

**Task Design.** During the experiment, participants first completed a training exercise to ensure competency using the response device, which was placed on all subjects’ right hand. This practice mirrored the primary task in design. The inclusion of a practice trial is essential for continuous performance tasks, as 61 observed a stabilization effect only after the first run of each of their continuous performance experiments. Participants were then provided information to contextualize the video stimulus that they were about to watch, which was an HBO murder mystery (i.e., Episode 4 of “The Undoing”, HBO television, original airdate: 11/15/20). The stimulus was split into two 22 minute 17 second components, representing the first and second halves of the episode, and were presented to participants across two sequential runs (**Figure 2A**). The episodes halves were presented in chronological order. Prior to exposure, participants were pseudo-randomly assigned to one of two conditions using a dynamic allocation approach in which the probability of condition assignment was determined by the distribution of subjects who had already completed the study in each condition. Subjects were assigned to continuously rate their certainty of a predefined stimulus-specific outcome (i.e., a target character’s innocence/guilt) for one half of the stimulus. They were instructed to watch the other half without rating, but to evaluate the stimulus as if they were providing ratings (i.e., to assess a target character’s innocence/guilt) (**Figure 2B**). When rating, a bipolar, horizontally-positioned scale was visualized below the video stimulus. The initial position of scale was set to 0% certainty. Pressing a button with the index finger incremented the scale by 5% closer towards the left pole (i.e., 100% certain of guilt) and pressing a button with the middle finger incremented the scale by 5% closer towards the right pole (i.e., 100% certain of innocence). Of the final sample, twenty (20) subjects rated the first half of the stimulus and fifteen (15) subjects rated the second half using the handheld device. Following the episode viewing task, participants completed an additional functional run in which they gauged the certainty of a non-social predefined stimulus-specific outcome (the visual luminance of the image). The purpose of this task is beyond the purview of this manuscript, but will be characterized in future work.

Following rating tasks, subjects completed a surprise free recall of the stimulus while undergoing fMRI. Subjects were instructed to describe everything that they can remember from the episode in as much detail as possible, even if a detail might not seem important. They were instructed to recall events in chronological order, but to return to any details that they later remember if they were forgotten. Lastly, subjects were instructed to speak for at least 10 minutes (self-timed), but that the recall could end whenever they choose. The maximum possible recall duration was established at 20 minutes (mean dur: 12 minutes 13 seconds, sd dur: 04 minutes 42 seconds). Instructions were presented and free recall responses were recorded via Matlab. Subjects subsequently completed task-related and individual difference measures outside of the scanner on Qualtrics. This included thirteen assessments of four prominent characters (e.g. “How agreeable was character X?”) measured on a 0 – 100 continuum. These assessments were based on a validated measure of common dimensions individuals use to represent others 45. These character assessments were completed by twenty-four (n = 24) of thirty-five subjects, as eleven subjects had completed the study before their addition to the protocol. All scripts associated with this task are publicly available at https://github.com/wj-mitchell/active\_viewing.

**Experimental Display and Rating Acquisition.** Software and hardware options available to researchers specifically designed for collecting continuous self-report ratings are numerous and constitute a rich topic of research on their own (62 contains a useful summary of these efforts). We designed a novel script programmed in Python [v3.8.13] 63 using the PsychoPy [v2021.2.3] 64 python library to capture our ratings. This choice provided flexibility to customize components present in the experimental session and ensured, due to using open-source software, that the code could be readily shared, replicated, and operated on any other Python-compatible computer. We chose to provide subjects with an MR-safe handheld button box (Psychology Software Tools five-button response unit) to provide their ratings rather than a joystick, as during piloting we found the joystick (Psychology Software Tools four-button joystick) to be more susceptible to generating inaccuracies in ratings (e.g., overshooting a target rating; accidentally changing ratings when not intending to, etc.). Additionally, by incrementing rating values only upon release of the button, rather than continuously for as long as the button is pressed, we hoped to more clearly delineate inflections in neural activity associated with rating or button pressing. All subjects used their right hand to provide ratings. Left handed subjects did not report any difficulties using the right-hand button box. Ratings were sampled at the stimulus’s average framerate (24 Hz).

**Audio Delivery.** Audio for the experimental task was presented through OptoAcoustics OptoActive sound-canceling headphones. To ensure clear and audible audio during MRI scanning, we analyzed the noise frequencies inherently generated by the MRI machine during imaging. We compared these frequencies with those in our audio stimuli and used Adobe’s Premiere Pro to shift any competing audio to non-competing frequency bands. This adjustment preserved the integrity of the audio experience for the subjects while minimizing interference from MRI noise. Presentation volume was adjusted to a comfortable level for each participant based upon subject feedback during a training exercise which featured royalty-free city noises played at a median volume which matched the median of our stimulus. Subjects could request volume changes between runs as needed. The visual elements of the experimental setup were projected on an MRI-compatible, out-of-bore screen using a Hyperion Projector.

**fMRI Pre-Processing.** We first converted all MRI data from DICOM to BIDS-formatted NIfTI files using heudiconv [v0.11.3] 65. Neuroimaging data was preprocessed with the standard fMRIPrep [v20.2.6] pipeline 66 within a Docker [v19.03.12] container to maintain generalizability. Motion outliers were assessed using the FSL Motion Outlier Tool 67, which defines outlier thresholds as the 75th percentile plus 1.5 times the interquartile range. TRs identified as outliers were incorporated into the GLM using regressor-based censoring. If greater than 15% of TRs that compose a trial are outliers, the trial was excluded from analyses. One subject was excluded according to this standard. Head motion was generally ideal, with 99.9% of all analyzed TRs (98.1% including the excluded subject) falling within an acceptable range.

For the ISC analysis, additional preprocessing was performed using nltools [v0.4.7] 68. Data were smoothed using a 6mm gaussian kernel and despiked using nltools find\_spikes function. Covariates of the data, including motion translations and rotations, were then regressed upon the neural data before it was parcellated into 400 unique functionally defined regions of interest using the 2022 17-network Schaefer-Kong Atlas 47. While the Schaefer-Kong Atlas is available in resolutions from 100 to 1000 parcels, 400 parcels is widely used as a standard due to previous work suggesting that the human cortex can be divided into 300 to 400 unique functional regions 69. It should be noted that MVPA analyses like ISC, which are sensitive to the voxel-level patterns that spatial smoothing could distort, are robust to the standard gaussian kernel size that fMRIPrep applies during spatial smoothing 70.

**Univariate Analysis.** FSL's [v6.0.5.1] FEAT [v6.0.0] 67 was used to perform a univariate parametric modulation analysis and contrast analyses between conditions (rating vs. non-rating). For rated runs, three (3) three-column event files were constructed. The first event file denoted all TRs in which rating changes occurred and did not include a parametric regressor. The second event file denoted all TRs in which no rating changes occurred and did not include a parametric regressor. The final event file modeled every TR and included a z-standardized parametric regressor: the number of button presses (i.e. rating changes) each subject performed within each TR (2s) over the 22m17s duration of the video.

Importantly, the regressor was based solely on the volume of button presses, regardless of the direction (i.e., index or middle finger button; left or right) or the corresponding certainty values. This approach ensured that our analysis isolated the neural activity associated with the act of making a decision (i.e., pressing a button) rather than subjects’ subjective certainty levels. By focusing exclusively on the number of button presses, we aimed to capture the cognitive and motor processes involved in rating itself.

Non-rated runs featured a single three-column event file with a single event that featured an onset and duration that corresponded to the entire length of the stimulus presentation. Each event file constituted a separate explanatory variable (EV) at the first level and was convolved with the standard FSL Double-Gamma HRF Temporal derivatives and filtering were applied, but no thresholding was used at this level. Data were then re-registered using the recommended technique for data preprocessed with fMRIPrep and analyzed in FSL 71.

A subset of the standard fMRIPrep confound output was used as nuisance regressors and included three-dimensional head motion translations, rotations, and their first derivatives, framewise displacement and censored head motion outlier timepoints. Additional confounds included cosine calculations to adjust for scanner drift 72 - an especially important adjustment for long duration stimuli 73 - as well as the first five temporal and anatomical components identified by fMRIPrep, which account for time- and spatial-related physiological confounds, respectively. Some stimulus features may confound with or/and elicit neural responses unrelated to rating, including image luminance, volume, character speech, and presence of faces. As such, volume-by-volume annotations of these features were also included as confounds. The luminance of each video frame was calculated using the imread function from the OpenCV [v4.10.0.82] 74 python library and averaged within each TR. The average volume in decibels within each TR was calculated using the librosa [v0.10.2] 75 Python library. The presence of speech and faces were manually coded moment-to-moment by a trained human annotator and confirmed by researcher review. Additional confirmation regarding the presence of faces or speech within each TR was achieved using Whisper [v1.1.10] 76 and the face\_recognition [v1.3.0] 77 Python library, which aligned with manual annotations. All stimulus-related confounds were z-scored.

At higher-level FEAT analyses, additional mean-centered EVs were created to adjust for subject-level confounds (i.e., handedness and sex) and stimulus. Contrasts between conditions and the parametric effect of button presses within the rating run were specified using FSL's Mixed Effects FLAME 1, with cluster-wise significance thresholds set following Woo et al.’s recommendations (z = 3.29, p < 0.001) 78. These included contrasts between rating and not rating both within-subject (i.e., contrasting the neural activity of subject in moments when they were actively rating a stimulus and moments when they were not actively changing their ratings) and between subjects (i.e., contrasting the neural activity of subjects who watched a stimulus via expressive active engagement versus subjects who watched the same stimulus via reflective active engagement).

**Intersubject Correlation Analysis.** Intersubject correlations were calculated using the parcel-wise approach that nltool’s isc and isc\_group functions 68 employed in Python. These functions correlate the time series of each ROI's activity within each participant with the average time series of that same ROI across all other participants, or all other participants within their group in the case of isc\_group. This yields a coefficient (the median correlative value, as recommended by Chen et al. 79) representing how similar neural activity patterns are in that ROI among that sample. To assess the significance of differences in neural synchrony between groups (i.e., expressive raters and reflective non-raters) within each run, we used subject-wise bootstrapping, which creates a new pairwise similarity matrix with randomly selected subjects for each group to generate a null distribution. Statistical significance is computed as the proportion of observations from the null distribution which are greater than the absolute value of the observed ISC difference relative to the total number of bootstrap samples (n = 10000), following the percentile method outlined by 80. This method is a more conservative test of statistical significance than permutation testing 79. To combine p-values from multiple tests (i.e., across runs), we used Fisher's method, which sums the logarithms of the individual p-values and compares the result to a chi-squared distribution to determine overall significance. Adjustments for multiple comparisons were then made using the Bejamini-Hochberg procedure to maintain the false discovery rate below 0.001.

**Activation Labeling.** After completing analyses, thresholded z-statistic maps and r-statistic maps were annotated using the automated anatomical atlas (AAL) 81, which provided probabilistically determined anatomical labels for each significant cluster. These labels were supplemented with the Schaefer-Kong atlas 46,47, which consists of 400 functionally-defined cortical parcellations and denotes which of 17 networks 49 each region predominantly participates within. When labeling was ambiguous or unavailable, the anatomical label in question was entered as a term in Neurosynth and the activation peak of the meta-analysis compared to the activation peak of the cluster in question. Additionally, certain specialized anatomical regions which are widely recognized within the neuroscience community, but which are not used in either of the atlases (e.g., temporoparietal junction) were confirmed using this technique.

**Free Recall Transcription**. Of the thirty-five subjects who completed the study, twenty-eight (n = 28) provided audio recordings that could reliably be transcribed. The audio of three subjects suffered technical issues which corrupted the files and audio from four other subjects was of low-quality and could not be reliably transcribed. Transcription was completed in stages. Audio recordings were first transcribed to text using a locally-hosted version of OpenAI’s Whisper [v1.1.10] 76 library in Python [v3.8.13] 63. These textual representations were then independently appraised and revised by a team of trained reviewers familiar with the stimulus and whom reviewed annotations of the stimulus generated by a separate team of reviewers in line with standards described by Chen et al.8. Each transcription was reviewed by at least two individuals who were required to reach a consensus. Any discrepancies between reviewers’ initial appraisals were flagged and discussed until a consensus could be reached or the questionable segment was censored. Two reviewers then independently segmented recall and categorized each segment according to which scene they believe that it was describing. We determined that there were 35 distinct scenes present in the stimulus. Again, any discrepancies between reviewers’ initial appraisals were discussed in a meeting facilitated by the corresponding author until a consensus on segmentation and categorization could be reached between reviewers. Training materials and standard operating procedures are available publicly at https://github.com/wj-mitchell/active\_viewing.

**Memory and Character Assessment Analyses.** Analyseswere conducted in R [v4.3.1] 82 using R Studio [v2023.09.0] as the IDE. For recall analyses, a dataframe containing a row for each scene and a column for each subject was constructed with a binary value in each cell denoting whether each scene was recalled. Complete pairwise Spearman-rank correlations were calculated among each pair

For each scene, we calculated correlations of z-scored ratings, separately within each condition (first-half raters vs. second-half raters). These within-condition correlations assess the consistency in rating patterns among participants who rated the same segments. To investigate the similarity in rating patterns across conditions, we used a form of RSA by correlating the similarity matrices for “same-condition” (both raters watched the same scenes) and “different-condition” comparisons. This RSA-based approach tests whether rating patterns are more similar within conditions than between them, reflecting consistency or divergence in interpretation across different scene orders. We performed within- and between-condition correlation analyses to assess similarity in rating patterns. An RSA approach was used to compare same-condition and different-condition correlations, providing a measure of representational similarity between conditions. To examine recall rates in relation to the rated/not-rated condition, we conducted a chi-square test. This test evaluates whether scenes that were rated in one half of the experiment are more likely to be recalled than scenes that were passively viewed. The chi-square test provides insight into whether active engagement influences memory encoding and recall. A chi-square test was conducted to compare recall rates for scenes participants actively rated versus those they passively viewed. This analysis examined the influence of active engagement on memory recall within the study. For participants who experienced both rating and non-rating conditions, we used a paired t-test to compare recall rates across rated and non-rated scenes. This approach accounts for within-subject differences in recall rates, allowing us to determine whether engagement type significantly affects memory. Paired t-tests compared recall rates between rated and non-rated scenes, focusing on within-subject differences to account for individual recall variability. This test assessed the impact of active rating on memory recall. To further validate the RSA analysis and check for statistical significance, we used a permutation test with 5,000 iterations. This non-parametric test shuffles condition labels across participants, recalculating the mean similarity difference between matching and non-matching conditions for each permutation. This yields a p-value reflecting the likelihood of observing the actual mean difference under random label assignment. To ensure robustness, we conducted a permutation test (5,000 iterations) by shuffling condition labels, recalculating mean similarity differences for each iteration. This approach provides a non-parametric p-value, indicating whether the observed condition similarity difference is statistically significant.

**Open Access Statement.** A detailed outline and scripts associated with pre-processing, analyses, and visualizations are publicly available at https://github.com/wj-mitchell/active\_viewing.

**Acknowledgments and Funding Sources**

**References**

1. Saarimäki, H. Naturalistic Stimuli in Affective Neuroimaging: A Review. *Front. Hum. Neurosci.* **15**, 675068 (2021).

2. Jääskeläinen, I. P. *et al.* Behavioral Experience-Sampling Methods in Neuroimaging Studies With Movie and Narrative Stimuli. *Frontiers in Human Neuroscience* **16**, (2022).

3. Ruef, A. M. & Levenson, R. W. Continuous Measurement of Emotion:The Affect Rating Dial. in *Handbook of Emotion Elicitation and Assessment* (eds. Coan, J. A. & Allen, J. J. B.) 286–297 (Oxford University PressNew York, NY, 2007). doi:10.1093/oso/9780195169157.003.0018.

4. Nummenmaa, L. *et al.* Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 9599–9604 (2012).

5. Mazza, D., Kolecki, J. C. & Scott, T. M. The Observer in Modern Physics Some Personal Speculations. *National Aeronautics and Space Administration* https://www.grc.nasa.gov/www/k-12/Numbers/Math/Mathematical\_Thinking/observer.htm (2000).

6. Gleiser, M. The reach of the scientific method is constrained by the limitations of our tools and the intrinsic impenetrability of some of nature’s deepest questions. *Nature* **557**, S20–S21 (2018).

7. Chang, L. J. *et al.* Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Sci. Adv.* **7**, eabf7129 (2021).

8. Chen, J. *et al.* Shared memories reveal shared structure in neural activity across individuals. *Nat Neurosci* **20**, 115–125 (2017).

9. Finn, E. S., Corlett, P. R., Chen, G., Bandettini, P. A. & Constable, R. T. Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nat Commun* **9**, 2043 (2018).

10. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject Synchronization of Cortical Activity During Natural Vision. *Science* **303**, 1634–1640 (2004).

11. Hasson, U., Furman, O., Clark, D., Dudai, Y. & Davachi, L. Enhanced Intersubject Correlations during Movie Viewing Correlate with Successful Episodic Encoding. *Neuron* **57**, 452–462 (2008).

12. Cliver, K. G. *et al.* Temporal memory for threatening events encoded in a haunted house. *Cogn. Emot.* 1–17 (2024) doi:https://doi.org/10.1080/02699931.2024.2338962.

13. Mobbs, D. *et al.* When Fear Is Near: Threat Imminence Elicits Prefrontal– Periaqueductal Gray Shifts in Humans. **317**, 6 (2007).

14. Stasiak, J. E. *et al.* Physiological arousal guides situational appraisals and metacognitive recall for naturalistic experiences. *Neuropsychologia* **180**, 108467 (2023).

15. Tashjian, S. M., Fedrigo, V., Molapour, T., Mobbs, D. & Camerer, C. F. Physiological responses to a haunted house threat experience: Distinct tonic and phasic effects. *Psychological Science* **33**, 236–248 (2022).

16. Levenson, R. W. & Gottman, J. M. Marital interaction: physiological linkage and affective exchange. *Journal of Personality and Social Psychology* **45**, 587–597 (1983).

17. Nastase, S. A., Goldstein, A. & Hasson, U. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage* **222**, 117254 (2020).

18. Fredrickson, B. L. & Kahneman, D. Duration Neglect in Retrospective Evaluations of Affective Episodes. *Journal of Personality and Social Psychology* **65**, 45–55 (1993).

19. Jeremy Peterman & Peterman, J. N. The ‘program analyzer’: a new technique in studying liked and disliked items in radio programs. *Journal of Applied Psychology* **24**, 728–741 (1940).

20. Hutcherson, C. A. *et al.* Attention and emotion: Does rating emotion alter neural responses to amusing and sad films? *NeuroImage* **27**, 656–668 (2005).

21. Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S. & Keysers, C. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Sciences* **16**, 114–121 (2012).

22. Posner, M. I. & Petersen, S. E. The Attention System of the Human Brain. *Annu. Rev. Neurosci.* **13**, 25–42 (1990).

23. Sonkusare, S., Breakspear, M. & Guo, C. C. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences* **23**, 699–714 (2019).

24. Borja Jimenez, K. C. *et al.* Changes in brain activity following the voluntary control of empathy. *Neuroimage* **216**, 116529 (2020).

25. Lehne, M. *et al.* Reading a suspenseful literary text activates brain areas related to social cognition and predictive inference. *PLOS ONE* **10**, (2015).

26. Sawahata, Y., Komine, K., Morita, T. & Hiruma, N. Decoding humor experiences from brain activity of people viewing comedy movies. *PLOS ONE* **8**, (2013).

27. Wallentin, M. *et al.* Amygdala and heart rate variability responses from listening to emotionally intense parts of a story. *NeuroImage* **58**, 963–973 (2011).

28. Lieberman, M. D. *et al.* Putting Feelings Into Words Affect Labeling Disrupts Amygdala Activity in Response to Affective Stimuli. *Psychological Science* **18**, 421–428 (2007).

29. Taylor, S. F., Phan, K. L., Decker, L. R. & Liberzon, I. Subjective rating of emotionally salient stimuli modulates neural activity. *NeuroImage* **18**, 650–659 (2003).

30. Lahnakoski, J. M. *et al.* Synchronous brain activity across individuals underlies shared psychological perspectives. *Neuroimage* **100**, 316–324 (2014).

31. Song, H., Finn, E. S. & Rosenberg, M. D. Neural signatures of attentional engagement during narratives and its consequences for event memory. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2021905118 (2021).

32. Gottman, J. M. & Levenson, R. W. A valid procedure for obtaining self-report of affect in marital interaction. *Journal of Consulting and Clinical Psychology* **53**, 151–160 (1985).

33. Axelrod, V. *et al.* Intracranial study in humans: Neural spectral changes during watching comedy movie of Charlie Chaplin. **185**, 108558–108558 (2023).

34. Fayn, K. *et al.* Full throttle: Demonstrating the speed, accuracy, and validity of a new method for continuous two-dimensional self-report and annotation. *Behavior Research Methods* **53**, 1–15 (2021).

35. Gazzaley, A. & Nobre, A. C. Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences* **16**, 129–135 (2012).

36. Schooler, J. W. & Engstler-Schooler, T. Y. Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology* **22**, 36–71 (1990).

37. Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H. & Gross, J. J. The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion* **5**, 175–190 (2005).

38. Valentin Wagner *et al.* Effects of continuous self-reporting on aesthetic evaluation and emotional responses. *Poetics* **85**, 101497 (2020).

39. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. NeuroSynth: a new platform for large-scale automated synthesis of human functional neuroimaging data. *Front. Neuroinform.* **5**, (2011).

40. Baldassano, C. *et al.* Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* **95**, 709-721.e5 (2017).

41. Lee, H. & Chen, J. Predicting memory from the network structure of naturalistic events. *Nat Commun* **13**, 4235 (2022).

42. Ambady, N. & Rosenthal, R. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin* **111**, 256–274 (1992).

43. Cone, J., Mann, T. C. & Ferguson, M. J. Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. in *Advances in experimental social psychology.* 131–199 (Elsevier Academic Press, San Diego, CA, US, 2017). doi:10.1016/bs.aesp.2017.03.001.

44. Ferguson, M. J., Mann, T. C., Cone, J. & Shen, X. When and How Implicit First Impressions Can Be Updated. *Curr Dir Psychol Sci* **28**, 331–336 (2019).

45. Thornton, M. A. & Mitchell, J. P. Theories of Person Perception Predict Patterns of Neural Activity During Mentalizing. *Cereb Cortex* **28**, 3505–3520 (2018).

46. Kong, R. *et al.* Individual-Specific Areal-Level Parcellations Improve Functional Connectivity Prediction of Behavior. *Cerebral Cortex* **31**, 4477–4500 (2021).

47. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex* **28**, 3095–3114 (2018).

48. Power, J. D. *et al.* Functional Network Organization of the Human Brain. *Neuron* **72**, 665–678 (2011).

49. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology* **106**, 1125–1165 (2011).

50. Kong, R. *et al.* Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *Cerebral Cortex* **29**, 2533–2551 (2019).

51. Laumann, T. O. *et al.* Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron* **87**, 657–670 (2015).

52. Mueller, S. *et al.* Individual Variability in Functional Connectivity Architecture of the Human Brain. *Neuron* **77**, 586–595 (2013).

53. Etkin, A., Egner, T. & Kalisch, R. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences* **15**, 85–93 (2011).

54. Hasson, U. *et al.* Neurocinematics: The Neuroscience of Film. *Projections* **2**, 1–26 (2008).

55. FeldmanHall, O. & Shenhav, A. Resolving uncertainty in a social world. *Nat Hum Behav* **3**, 426–435 (2019).

56. Logothetis, N. K. What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878 (2008).

57. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).

58. Reilly, J. *et al.* Dynamics of Language Use and Alignment in Different-Age Conversation Partners. *PsyArxiv* (2023) doi:https://doi.org/10.31234/osf.io/3crua.

59. Sievers, B., Welker, C., Hasson, U., Kleinbaum, A. M. & Wheatley, T. Consensus-building conversation leads to neural alignment. *Nat Commun* **15**, 3936 (2024).

60. Yeomans, M., Boland, F. K., Collins, H. K., Abi-Esber, N. & Brooks, A. W. A Practical Guide to Conversation Research: How to Study What People Say to Each Other. *Advances in Methods and Practices in Psychological Science* **6**, 25152459231183919 (2023).

61. Teresa Jacobson Kimberley *et al.* Reliability of fMRI during a Continuous Motor Task: Assessment of Analysis Techniques. *Journal of Neuroimaging* **18**, 18–27 (2008).

62. Girard, J. M. & Wright, A. G. C. DARMA: Software for dual axis rating and media annotation. *Behavior Research Methods* **50**, 902–909 (2018).

63. van Rossum, G. Python tutorial. (1995).

64. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav Res* **51**, 195–203 (2019).

65. Halchenko, Y. *et al.* nipy/heudiconv: Zenodo https://doi.org/10.5281/zenodo.5557588 (2021).

66. Esteban, O. *et al.* Poldracklab/Fmriprep: 1.0.0-Rc5. Zenodo https://doi.org/10.5281/ZENODO.996169 (2017).

67. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790 (2012).

68. Chang, L., Eshin Jolly, Cheong, J. H., Burnashev, A. & Chen, A. cosanlab/nltools: 0.3.11. Zenodo https://doi.org/10.5281/ZENODO.2229813 (2018).

69. Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J. & Coalson, T. Parcellations and Hemispheric Asymmetries of Human Cerebral Cortex Analyzed on Surface-Based Atlases. *Cerebral Cortex* **22**, 2241–2262 (2012).

70. Hendriks, M. H. A., Daniels, N., Pegado, F. & Op de Beeck, H. P. The Effect of Spatial Smoothing on Representational Similarity in a Simple Motor Paradigm. *Front. Neurol.* **8**, 222 (2017).

71. *How to Use FEAT While Skipping Registration*. (Youtube, 2017).

72. Friston, K. J., Frith, C. D., Turner, R. & Frackowiak, R. S. Characterizing evoked hemodynamics with fMRI. *NEUROIMAGE* **2**, 157–165 (1995).

73. Power, J. D., Schlaggar, B. L. & Petersen, S. E. Studying Brain Organization via Spontaneous fMRI Signal. *Neuron* **84**, 681–696 (2014).

74. Bradski, G. The OpenCV Library. (2000).

75. McFee, B. *et al.* librosa: Audio and music signal analysis in python. in *Proceedings of the 14th python in science conference* vol. 8 (2015).

76. OpenAI. Whisper. (2023).

77. Ageitgey, A. face-recognition. (2023).

78. Woo, C.-W., Krishnan, A. & Wager, T. D. Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage* **91**, 412–419 (2014).

79. Chen, G. *et al.* Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. *NeuroImage* **142**, 248–259 (2016).

80. Hall, P. & Wilson, S. R. Two Guidelines for Bootstrap Hypothesis Testing. *Biometrics* **47**, 757 (1991).

81. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289 (2002).

82. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing (2022).

**Figure Legends**

A diagram of a variety of different views

Description automatically generated with medium confidence

Figure 1. Conceptual overview.

(ADD IN CHARACTER AND RECALL ASSESSMENT FIGURES)

A group of brain images

Description automatically generated with medium confidence

A group of images of a brain

Description automatically generated

Figure 4. Expressive versus reflective rating.

A group of images of a brain

Description automatically generated

Figure 5. Reflective versus expressive rating.

A diagram of the brain

Description automatically generated

Figure 6. Comparing non-rating related activation between reflective and expressive viewing.