Check for updates

OPEN ACCESS

EDITED BY Catie Chang, Vanderbilt University, United States

REVIEWED BY

Matthew Elliott Sachs, Columbia University, United States Xin Di, New Jersey Institute of Technology, United States

*CORRESPONDENCE Talia Brandman talli.brandman@gmail.com

SPECIALTY SECTION

This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

RECEIVED 30 May 2022 ACCEPTED 19 October 2022 PUBLISHED 09 November 2022

CITATION

Brandman T, Malach R and Simony E (2022) Retrospective behavioral sampling (RBS): A method to effectively track the cognitive fluctuations driven by naturalistic stimulation. *Front. Hum. Neurosci.* 16:956708. doi: 10.3389/fnhum.2022.956708

COPYRIGHT

© 2022 Brandman, Malach and Simony. This is an open-access article distributed under the terms of the Creative Commons Attribution License

(CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Retrospective behavioral sampling (RBS): A method to effectively track the cognitive fluctuations driven by naturalistic stimulation

Talia Brandman^{1*}, Rafael Malach¹ and Erez Simony^{1,2}

¹Department of Brain Sciences and Azrieli National Institute for Human Brain Imaging and Research, Weizmann Institute of Science, Rehovot, Israel, ²Faculty of Engineering, Holon Institute of Technology, Holon, Israel

Everyday experiences are dynamic, driving fluctuations across simultaneous cognitive processes. A key challenge in the study of naturalistic cognition is to disentangle the complexity of these dynamic processes, without altering the natural experience itself. Retrospective behavioral sampling (RBS) is a novel approach to model the cognitive fluctuations corresponding to the time-course of naturalistic stimulation, across a variety of cognitive dimensions. We tested the effectiveness and reliability of RBS in a webbased experiment, in which 53 participants viewed short movies and listened to a story, followed by retrospective reporting. Participants recalled their experience of 55 discrete events from the stimuli, rating their quality of memory, magnitude of surprise, intensity of negative and positive emotions, perceived importance, reflectivity state, and mental time travel. In addition, a subset of the original cohort re-rated their memory of events in a followup questionnaire. Results show highly replicable fluctuation patterns across distinct cognitive dimensions, thereby revealing a stimulus-driven experience that is substantially shared among individuals. Remarkably, memory ratings more than a week after stimulation resulted in an almost identical timecourse of memorability as measured immediately following stimulation. In addition, idiosyncratic response patterns were preserved across different stimuli, indicating that RBS characterizes individual differences that are stimulus invariant. The current findings highlight the potential of RBS as a powerful tool for measuring dynamic processes of naturalistic cognition. We discuss the promising approach of matching RBS fluctuations with dynamic processes measured via other testing modalities, such as neuroimaging, to study the neural manifestations of naturalistic cognitive processing.

KEYWORDS

retrospective, behavioral sampling, naturalistic stimulation, cognitive dynamics, cognitive fluctuations, memory, individual typicality, naturalistic perception

Introduction

Can we describe the sequence of cognitive processes engaged while watching a movie? Everyday experiences are dynamic, complex and multisensory in nature, supported by the interaction and regulation of simultaneously activated processing mechanisms (Hasson et al., 2004; Sonkusare et al., 2019). A fundamental challenge in the study of real-life cognition lies in tracking the cognitive processes that govern one's continuously-changing experience of unfolding events. In recent years, a multitude of neuroimaging research groups has ventured beyond reductionist paradigms to mimic reallife experiences, such as watching a movie or listening to a story, while recording brain activation (e.g., Spiers and Maguire, 2007; Simony et al., 2016; Chen et al., 2017; Zadbood et al., 2017; Baldassano et al., 2018; Eickhoff et al., 2020; Brandman et al., 2021; Finn and Bandettini, 2021). Characterizing the fluctuations in cognitive states is key to understanding the fluctuations in neural and physiological states recorded during naturalistic stimulation.

There are three crucial problems in characterizing the cognitive dynamics driven by naturalistic stimuli as movies and stories. Unlike reductionist manipulations that isolate a specific target mechanism by onset [e.g., face perception: Farah et al. (1998); tone prediction: Bendixen et al. (2009)], the multidimensional nature of naturalistic stimulation drives many simultaneous mechanisms of varying time-scales (Hasson et al., 2008). These can include a wide range of perceptual, semantic and emotional processes, social inference and memory encoding (e.g., McGurk and MacDonald, 1976; Loftus et al., 1978; Yeshurun et al., 2017). Furthermore, as information is continuously changing throughout unfolding events, we cannot predict how each of these simultaneous mechanisms will fluctuate throughout the movie. Thus, the first problem is disentangling the multitude of simultaneous processes, and the second problem is tracking their fluctuations across time. In addition, as we are interested in the natural, unaltered, human experience, our third problem is the act of measurement itself, which may affect in real-time the exact process we are trying to measure. This is widely known as the measurement problem, which, analogous to its origin in quantum physics, stipulates that "the mental content is invariably altered when the attention is concentrated on any special feature of it" (Bohr, 1958; Malach, 2007). How then, can we disentangle simultaneous cognitive processes activated during movie viewing, and measure their fluctuations along the time-course of the movie, without interfering with the natural ongoing experience?

To tackle these challenges posed by naturalistic stimulation, we developed a new approach termed retrospective behavioral sampling (RBS), which models the cognitive dynamics of human experience across a variety of cognitive dimensions. The central premise of this approach is based on previous work using autobiographical interview techniques (Berntsen, 2002; Levine et al., 2002), demonstrating that we can gain insight into one's subjective past experience by reactivating their memory of it in retrospect. Such reactivation has been shown to evoke neural activity containing representations of the original experience (Gelbard-Sagiv et al., 2008; Chen et al., 2017; Zadbood et al., 2017; Norman et al., 2019). Following this logic, RBS is designed to reconstruct the cognitive state throughout the time-course of a dynamic stimulus, by deconstructing the experience into discrete events for reactivation. Following uninterrupted stimulation, the memory of each event is reactivated, and the (retrospective) subjective experience at the moment of the event (i.e., how one remembers the experience) is systematically reported across an array of cognitive dimensions.

We examined the effectiveness and reliability of RBS as a method for modeling cognitive dynamics driven by naturalistic stimuli such as movies and stories. To clarify, the dynamics captured by RBS describe changes in cognitive states across discrete events throughout the movie, not continuous momentto-moment fluctuations within an event. The results presented below show that RBS patterns are informative and highly replicable across groups and testing sessions, and that individual differences are largely stable across stimuli. We assess the strengths and the limitations of RBS, and discuss its potential for matching cognitive processes with neural dynamics, individual clinical markers and other testing modalities.

Materials and methods

Retrospective behavioral sampling began with uninterrupted stimulation using movies or stories, followed by a behavioral questionnaire collecting a systematic report of the cognitive experience corresponding to each event in the movie. The details of the approach are presented in Figure 1. To reactivate one's experience of a specific event, it was described by a textual reminder, with time taken to silently recall all details of the event and the subjective experience it had evoked. Thereafter, the experience was rated on a series of cognitive scales measuring the quality of memory for the event, how surprising it was at the time, the intensity of positive and negative emotions triggered by the event, and perceived importance of the event to the main storyline. In addition, introspective rating questions were aimed at modeling one's state-of-mind around the time of the event. These included the extent to which one felt immersed in the movie or reflective in thought, at the time leading into the event and later as triggered by the event. Finally, mental time-travel referred to the extent to which such reflections concerned the past or the future.

Participants

Fifty-three participants were included in the study, in two groups: Group I included 26 participants (13 female, ages



yellow line illustrates where the example Event #31 falls within each of the rating time-courses.

 33.08 ± 10.17) based in the USA. Group II included 27 participants (15 female, ages 34.48 ± 8.33) based in the UK. All participants were native English speakers with fluent writing abilities, and had intact hearing and eyesight (or corrected eyesight). Pre-screening further included a high approval record of at least 95% on previous web-surveys submitted on the survey platform, and technological compatibility criteria (see Section "Experimental apparatus"). Two additional participants, not included above, had completed the experiment, but were excluded from analysis due to suspected cheating. This was

determined by zero variance across measures on one or more events, reflecting repetition of the same rating response (usually "1") throughout a long consecutive series of questions.

Stimuli

The stimuli for this study included four short documentary movies and an auditory story. Two of the movies, "Human Body" (1:03 min; BBC) and "Pockets" (3:08 min; Pilgrim Films), were adapted from the stimulus set of the Human

Connectome Project (HCP) (Cutting et al., 2012; Van Essen et al., 2013). The former was a narrated clip depicting people in extraordinary circumstances (e.g., a blind rock climber), and the latter was a street-interview clip in which passersby presented and explained the content of their pockets. Two additional movies were narrated nature documentaries adapted from Planet Earth II (BBC), depicting Great Bowerbirds nesting and mating strategies (4:38 min), and Hyenas coexisting with humans in the city of Hara in Ethiopia (5:07 min). Finally, the auditory piece was a real-life story, "Pie Man" (7:18 min; Jim O'Grady), recorded at a live storytelling performance ("The Moth" storytelling event, New York City), and adapted from the stimulus set used in earlier work (Simony et al., 2016). The choice of stimuli was aimed at maximizing variability on all tested cognitive dimensions, for example, by including tight and lose storylines, positive, negative and neutral emotional content, positively and negatively surprising events, etc. The objective was to minimize the chance of confounding multiple measures such that they are indissociable [e.g., confounded emotional intensity and surprise in a suspenseful Hitchcock video; Brandman et al. (2021)].

Experimental procedure

Experimental procedures were approved by the institutional review board (IRB; approval reference # 533-2) of the Weizmann Institute of Science.

Individuals who matched the pre-screening criteria on the web-survey platform were directed to an introductory page with a description of the task and technical compatibility requirements. Those who chose to continue and confirmed compatibility proceeded to the experiment. Participants then went through two additional screening steps to make sure that they could view the videos in their intended dimensions and that they could hear and see the stimuli well with their current setting (see Section "Experimental apparatus").

After passing all screening stages, participants were presented with the four videos in random order, and the auditory story presented last. Only after all stimuli were viewed participants were directed to the RBS questionnaire, in which they responded to each stimulus in turn, by order of initial stimulation. For each stimulus, participants first typed an openended free recall of the entire viewing experience, including all they could remember from the movie/story and their own reactions when watching. The free recall was followed by a series of rating questions corresponding to the same movie/story, in which each time, participants were reminded of a single event in the movie/story and rated how well they remembered it, understood it, how much it surprised them, the intensity of any positive and negative emotions it triggered, how important the event was to the main storyline of the movie/story, how much they were reflecting in thought versus immersed in stimulation at the beginning and at the end of the event, and how much the event prompted them to think about the past or the future. Altogether, participants completed five freerecall sessions, corresponding to the five stimuli, and 495 rating questions, corresponding to nine measures X 55 sampled events.

Upon completing the RBS questionnaire, participants were asked to report their age and gender, and were asked whether they had been exposed to any of the five stimuli before participating in this experiment.

Experimental apparatus

The experiment was conducted in an online web-survey written in HTML/CSS and JavaScript, and carried out on survey platforms Amazon Mechanical Turk (Group I) and Prolific.com (Group II). Technological compatibility screening included a computer operated with Windows, Mac, Linux or UNIX, Wifi connection sufficient for downloading ~200 MB video content, working sound device, and a screen large enough to view the videos at fixed presentation size. Videos were presented at 200 mm × 112.5 mm, and participants were instructed to maintain a distance of about 30 cm from the screen. Screen size was validated in an interactive procedure, in which participants increased their browser window size until a button appeared allowing them to proceed with the experiment. Video presentation quality was validated with a practice video sample followed by catch questions to confirm participants both saw and heard the sample well. Prior test stimulation, participants were instructed to view every video and listen to the auditory story in full, without pausing, rewinding or skipping ahead. Pauses were recorded and viewing time was monitored to assure full and continuous viewing of each stimulus.

The retrospective behavioral sampling questionnaire

We developed the RBS questionnaire to track the temporal fluctuations in cognitive states along the stimuli timeline. To achieve this, we densely sampled events along the stimulus timeline at intervals of around 15–30 s. The main consideration for selection of events within these interval timeframes was that they could be coherently described, i.e., that something distinct was happening that could be used as a reminder to recall the event. Throughout the four movie clips, 33 events were sampled, and 22 more events were sampled throughout the auditory story, resulting in a total of 55 events.

The questionnaire was blocked by stimulus, such that after viewing all stimuli consecutively, participants completed the entire RBS questionnaire by order of stimulus viewing. Each stimulus block began with a typed free recall of the entire movie or story, including the subjective experience of the participant

while viewing. The main objective of the free recall was to verify that web-based participation had included actual viewing and attending to the stimulus being played. After free recall, participants received instructions for event rating, in which they were asked to refer specifically to each event, meaning no more than a few seconds before or after the described moment in the movie/story. General instructions for the event rating phase were given as follows: "In the next pages, you will be prompted on particular events in the movie/story. Each time, you will receive a description of an event. For example: narrator: "Lions can see well in the moonlight. This one is on the lookout for prey". (1) You will be asked to recall the precise moment in the movie/story-no more than a few seconds before and after this event (not the entire scene)-and rate the quality of your memory of it. (2) You will be asked to answer a series of questions about your experience in that moment of the movie/story. Read the instructions of each question carefully, and use the full range of the scale between 1 and 7 to provide a response that most accurately describes your experience, as you remember it, at the time of watching/hearing the particular event in question. The answers should describe your personal experience while watching/listening, as accurately as you can remember."

Following these general instructions for event rating, participants began the event-rating phase of the block. Each time participants would be probed on a single event from the movie/story and rate it on all tested measures in a sequence of multiple-choice questions. All sampled events for the movie/story were rated. Events were probed in random order.

Each event began with a textual reminder. For example: *Mustached man eating little sweets: "Yes it's salty, and... it's licorice"* (all event reminders are listed in **Supplementary Table 1**). This screen appeared for 30 s, in which participants were instructed to silently recall the event in their mind, including everything they could remember about what happened and their own subjective viewing experience. The exact instructions were: "Please take a moment to recall this exact moment of the movie, what happened, what you saw, what you heard, what were your own thoughts, emotions and/or physical sensations at that moment, etc." The objective of the silent recall was to bring the earlier experience of viewing the event into mind. When the time was up, the screen automatically changed to the first rating question.

Each rating question appeared in single page, with the event reminder kept at the top of the page during the entire rating sequence. No time limit was given for response. Upon response (key press between 1 and 7), participants were automatically forwarded to the next question.

Rating questions Memory

Before rating any specific type of experience, participants were asked to rate their quality and detail of memory for the recalled event. The exact phrasing was: "How would you rate your quality of memory for this event as compared to other events? Consider the full range of the scale and provide an accurate response between 1—no memory at all—to 7—best and most detailed remembered event in this video."

Comprehension sanity check

"How would you rate your comprehension of this event? Consider the full range of the scale and provide an accurate response between 1—did not understand what was happening to 7—understood perfectly what was happening."

Surprise

"How surprising was the event? Consider the full range of the scale and provide an accurate response between 1—did not surprise me at all—to 7—no other event in this video surprised me this much."

Negative emotion

"How much negative emotion did this event cause you? Consider the full range of the scale and provide an accurate response between 1—no detectable negative emotion—to 7 more intensely negative than any other in this video."

Positive emotion

"How much positive emotion did this event cause you? Consider the full range of the scale and provide an accurate response between 1—no detectable positive emotion—to 7 more intensely positive than any other in this video."

Importance

"How important was this event for understanding the main storyline of this video? Consider the full range of the scale and provide an accurate response between 1—insignificant—to 7 more important than any other event in this video."

Initial reflectivity

"When this event came up, at what state of mind did it "catch" you? How much were you reflecting on things either related or unrelated to the movie immediately before this event happened? (such as recounting previous events or personal memories, predicting what would happen next, analyzing the plot or characters, or any other type of thought). Consider the full range of the scale and provide an accurate response between 1—completely immersed in video, no other thought—to 7 most reflective moment in this video."

Triggered reflectivity

"How did this event affect your state of mind? How much did this event make you reflect on things either related or unrelated to the movie immediately after it happened? (such as recounting previous events or personal memories, predicting what would happen next, analyzing the plot or characters, or any other type of thought). Consider the full range of the scale and provide an accurate response between 1—completely immersed in video, no other thought—to 7—most reflective moment in this video."

Mental time travel

"What did this event make you think of - did it send your thoughts into the past or into the future? In other words, how much did it remind you of past events or make you predict what would happen next? Consider the full range of the scale and provide an accurate response between 1—vivid recollection of past events—to 7—vivid prediction of future events."

Follow-up testing

A subset of 23 participants from Group I re-rated their memory of all events between 1 and 2 weeks following original testing. During follow-up testing, participants did not view the stimuli again, and were directly taken to a stripped-down version of the RBS questionnaire. In each block, they completed the typed free recall of the entire stimulus, and the silent recall and memory rating for every event, as in the original questionnaire, but were not asked any further questions.

Data processing

Trial exclusion

In the case that a participant had reported previous exposure to a stimulus, i.e., that the experiment had not been their first viewing of a certain movie/story, only the events corresponding to that stimulus were excluded from that participant's data. Previous exposure was reported for the Bowerbird video in three participants and for the Hyena video in 6 participants. In addition, in the case that a participant presented poor understanding of a certain stimulus, all events corresponding to that stimulus were excluded from their data. Particularly, one participant rated their comprehension between 1 and 3 (out of 7) for 16 out of the 22 events in the auditory story, whereas the same participant's comprehension ratings for the movie events were high. We therefore excluded the story events from this participant's results. Notably, unremembered events (memory rating of 1) were reported at an average of less than one event per subject, and were not excluded from the analysis.

Sanity checks

Following trial exclusion as described above, comprehension ratings were examined as a sanity check, to ensure overall coherency of sampled events and their textual reminders. Means across subjects of event comprehension suggest that all events and probes were sufficiently coherent, with comprehension of 5.74 (out of 7) on average across all events, and the lowest mean comprehension of 4.83 on Event #9.

Scoring

To prepare rating data for statistical testing, z-scoring was performed along the entire dataset of each participant separately, by subtracting the mean and dividing by standard deviation.

Individual typicality

The typicality index for each participant was defined by the following equation, as the Euclidean distance (*d*) between the feature vector of each participant (*q*) and that of the group (*p*), given by the square root of the sum of squared deltas for each element *i* of *q* and *p* between 1 and *n*. Euclidean distance was scaled by the square root of vector length (*n*).

$$d(p,q) = \frac{\sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}}{\sqrt{n}}$$
(1)

Results

We analyzed participants' retrospective responses to 55 events throughout the movies and story, which were rated on scales of Memory, Surprise, Negative Emotion, Positive Emotion, Importance, Initial Reflectivity, Triggered Reflectivity, and Mental Time Travel.

How effective is retrospective behavioral sampling for dynamic cognitive tracking?

We set out to assess the efficacy of RBS as a reliable and informative method for cognitive tracking of dynamic natural experience. To this end, we tested the replicability of the sampled cognitive dynamics across independent groups and testing sessions, as well as their stability by group size, and informativeness across time.

First, we examined whether RBS is generally informative of differences between events and between measures. Indeed, we found significant differences between sampled events, which further varied as a function of cognitive measure. This was tested in repeated-measures ANOVA of ratings, with event (1 through 55) and cognitive measure (memory, surprise, negative emotion, positive emotion, importance, initial reflectivity, triggered reflectivity, mental time travel) as within-subject factors, which revealed main effects of event [$F_{(54,2430)} = 19.23$, P < 0.001], and measure [$F_{(7,315)} = 105.99$, P < 0.001], as well as an interaction between them [$F_{(378,1.7e4)} = 11.68$, P < 0.001]. These effects indicate that ratings significantly differed across the 55 events, in a pattern that varied by the type of cognitive measure.

Next, we asked whether these cognitive dynamics, given by the time-courses of ratings across events, can be replicated across the two independent groups of participants ($N_1 = 26$, $N_2 = 27$). Correlation analysis revealed robust replicability of cognitive dynamics across groups in all tested measures (FDR corrected *P* < 0.001). Specifically, mean ratings were correlated across the 55 events between the two groups at R = 0.9 for Importance, Positive Emotion, Negative Emotion, and Surprise, R = 0.8 for Memory, R = 0.7 for Triggered Reflectivity and Mental Time Travel, and R = 0.5 for Initial Reflectivity. Detailed correlation values are presented in **Figure 2A**.

To further examine the stability of these results as a function of group size, we calculated the correlations of randomly assigned groups sampled 100 times for each group size, ranging from 1 to 23 participants per group (of the 46 participants with full trial data). This bootstrap analysis revealed high stability, with mean correlations reaching R > 0.4 with as few as three participants for Surprise, Memory, Importance, Negative Emotion and Positive Emotion, and with four participants for Mental Time Travel, relative to moderate stability of Initial and Triggered reflectivity, which reached R > 0.4 with 17 and 10 participants, respectively. Detailed correlation values are presented in Figure 2B.

Finally, to assess test-retest reliability of RBS, we examined the replicability of memory ratings in a subset of participants (N = 23) who had rated their memory of all events a second time, between 1 and 2 weeks after stimulation. We found strikingly high replicability across the two rating sessions, with mean ratings of Memory correlated across the 55 events between the two testing sessions at R = 0.9 (P < 0.001) (**Figure 3**). We note that, while this confirms test-retest reliability, it does not rule out potential effects of initial reporting on memory ratings in the follow-up session.



FIGURE 2

Replicability and stability of cognitive dynamics. (A) Comparison of mean ratings across subjects between Group I–Mturk subjects (green; N = 26), and Group II–Prolific subjects (blue; N = 27). Dynamics were significantly correlated between groups in all tested cognitive measures (FDR corrected P < 0.001). Ratings are presented as mean and SEM across subjects. (B) Stability of group dynamics as a function of group size in a 100-iterations bootstrapping procedure. Each iteration randomly assigned 23 participants to each of two independent groups and tested the correlation across group means for each group size between 1 and 23. Bootstrapping results are presented as mean and SEM across sampling iterations.



How are cognitive measures related to one another?

To examine the relationship among the different cognitive measures, we constructed their pair-wise correlation matrix for each participant, by correlating across the 55 events between each pair of cognitive measures. The group means of within-subject correlations are presented in **Figure 4A**. In addition, we calculated the correlations between each pair of measures across the two independent groups of participants, such that each measure in one group was correlated with the second measure in the second group. The symmetrized matrix of correlations between groups is presented in **Figure 4B**. In both correlation matrices, the four largest collinearities among cognitive measures were observed between Initial Reflectivity and Triggered Reflectivity, Positive and Negative Emotion (inverse), Memory and Importance, and Memory and Surprise (FDR corrected P < 0.001).

These collinearities raise the question of redundancy, i.e., whether any one measure could be fully explained by its shared variance with other measures. To test this, we computed the partial correlations across measures and groups. This was done by correlating each single measure in Group I with the same measure in Group II, while controlling for correlations between Group I and all other measures in Group II. Partial correlations were calculated in both directions, i.e., between Group I and II while controlling for all other measures in Group II and vice versa, then averaged across the two results. Mean partial correlations: Importance R = 0.83 (FDR corrected P < 0.001), Surprise R = 0.79 (cor. P < 0.001), Memory R = 0.73 (cor.



P < 0.001), Positive Emotion R = 0.66 (cor. P < 0.001), Negative Emotion R = 0.64 (cor. P < 0.001), Mental Time Travel R = 0.47 (cor. P = 0.002), Initial Reflectivity R = 0.19 (cor. P = 0.596) and Triggered Reflectivity R = 0.17 (cor. P = 0.674). Thus, only collinearities among measures of Reflectivity (both initial and triggered) explained the full variance driving replicability of these two measures across groups.

How effective is retrospective behavioral sampling for studying individual differences?

To test the potential use of RBS for the study of individual differences, we calculated participants' (a)typicality index,

defined as the scaled Euclidean distance between his or her rating pattern and that of the group mean (**Figure 5A**; Ramot et al., 2020). Thus, the typicality index provides an assessment of how different an individual is in his or her unique pattern of response relative to the group. For example, a rating time-course similar to the group mean would be indexed as typical—given by a *low* typicality index, whereas a rating pattern more different from the group mean would be indexed as less typical—given by a *high* typicality index (**Figure 5B**). We note that the results reported below did not change when, instead of using the full group mean, typicality scores were calculated against the group mean excluding the target subject.

To test whether individual typicality captured by RBS dynamics is a reliable marker of individual differences, we asked whether it is stimulus-invariant, i.e., whether it would be stable across different types of narrative stimuli–even when they differ in content and modality. To test this, we calculated the typicality index for movies (33 events) and, separately, the typicality index for the auditory story (22 events), for each participant with complete trial data (N = 46). The distribution of typicality indices is illustrated in **Figure 5C** in reduced dimensionality. We then measured the correlation across participants' typicality indices between the two types of stimuli, revealing a robust correlation of R = 0.78 (P < 0.001), which thereby explains 61% of the variance across individuals ($R^2 = 0.61$). As visible in **Figure 5D**, all participants showed stable typicality across stimuli.

While the above typicality analysis harnessed the full variability of the temporal response patterns across individuals, we additionally calculated the specific typicality indices corresponding to each cognitive measure separately (see **Supplementary Figure 1**). Correlations across stimuli varied by measure: Mental Time Travel R = 0.77 (FDR corrected P < 0.001), Memory R = 0.75 (cor. P < 0.001), Negative Emotion R = 0.72 (cor. P < 0.001), Triggered Reflectivity R = 0.62 (cor. P < 0.001), Initial Reflectivity R = 0.62 (cor. P < 0.001), Positive Emotion R = 0.49 (corrected P < 0.001), Importance R = 0.18 (cor. P = 0.255), Surprise R = 0.10 (corrected P = 0.498).

Discussion

Retrospective behavioral sampling was developed as a novel approach to reconstruct the fluctuations in cognitive states driven by naturalistic stimuli such as movies and stories. With RBS, we aimed to disentangle simultaneous cognitive processes, and model their fluctuations along the time-course of stimulation, without interfering with the direct natural experience. Results clearly show that RBS is both effective and reliable in characterizing these cognitive fluctuations. RBS captured changes in the cognitive state across events, which differed as a function of cognitive measure. These cognitive dynamics were highly replicable across independent groups of participants, indicating a substantial and reliable stimulusdriven experience. RBS was highly effective in tracking these shared dynamics, exhibiting striking replicability even with small sample sizes, as well as across separate testing sessions. Furthermore, RBS was informative of individual differences in subjective experience, showing consistency of individual typicality across different narratives and stimulus modalities.

Disentangling simultaneous cognitive processes

Cognitive dynamics, measured by RBS rating timecourses, varied across the eight tested cognitive measures, consistently beyond individual variability. Furthermore, the specific dynamics corresponding to each cognitive process, separately, exhibited robust reliability. Together, these results suggest that RBS can effectively dissociate the temporal dynamics of several cognitive processes within a single questionnaire. Note, however, that the ability to dissociate these processes does not rule out the existence of collinearities among them, as well as non-linear interactions. In the current data, we see the largest collinearities among memory, surprise and importance-consistent with concepts of predictive processing, i.e., that deviance from expectation strengthens encoding (Rescorla and Wagner, 1972; Schultz, 1998), as well as between positive and negative emotion, and between initial and triggered reflectivity. Nevertheless, after controlling for all collinearities, the resultant unique component of each measure that is replicable across groups, remained high for all measures other than reflectivity. This suggests that the two reflectivity measures could be collapsed into a single question in future, with no expected loss of information.

Measuring stimulus-driven cognitive dynamics

As we have seen, rating time-courses exhibited robust replicability across independent groups of participants, suggesting that stimulus-driven experience encompasses typical patterns of cognitive dynamics that are shared across individuals. This can be viewed as analogous to the shared component of response found in stimulus-driven neural fluctuations, when correlating neuroimaging responses across individuals using dynamic inter-subject correlation techniques (Hasson et al., 2004; Simony et al., 2016). Such shared neural dynamics have been previously shown to reflect high-level cognitive representations that are shared across individuals [e.g., narrative interpretation: Nguyen et al. (2019); social inference: Ramot et al. (2020)]. In the current study, the shared element of cognitive experience was found to be effectively and reliably measures by RBS along the stimuli



timeline, with minimal sample size. This makes RBS an ideal tool to generate predictor variables that can be applied to the analysis of fluctuations in a variety of testing modalities (Simony and Chang, 2020). In essence, using RBS, we can associate cognitive processes with any other dynamic process corresponding to the same stimulus, e.g., as measured with neuroimaging, electrophysiology, dynamic physiological indices or eye-tracking.

Showcasing its potential, our previous work using an earlier (provisional) version of RBS, uncovered correlations between default mode network (DMN) coactivations and experienced surprise, pointing to an important role of the DMN in predictive processing of movie narratives (Brandman et al., 2021). The discovery was made possible by comparing the RBS surprise ratings collected in a web-based experiment, with neural dynamics from earlier functional magnetic resonance imaging (fMRI) studies (Chen et al., 2017; Zadbood et al., 2017), thereby matching cognitive and neural fluctuations across independent groups of participants. This highlights an important advantage of RBS, namely, that it does not require that the same participants undergo both RBS and neuroimaging. Thus, RBS can be used to extract cognitive regressors for an entirely independent dataset, allowing for its immediate implementation to investigate the cognitive processes reflected in brain signals recorded in earlier neuroimaging and electrophysiological studies. The only prerequisite is that RBS be collected for the same stimuli as in the dataset to be matched. In addition, using a similar approach, we could compare the dynamics of human

experience, as measured with RBS, with computer inference, as given by artificial neural-network output to the same stimuli.

Individual differences in cognitive dynamics

Retrospective behavioral sampling rating patterns were indicative of the extent to which an individual's reported experience was typical relative to the common experience of the group. Individual typicality was found to be stable across movies and story stimuli, thereby tapping into stimulus-invariant individual response tendencies. These results highlight the functional power of coding through similarity distances, which have been previously shown to explain perceptual differences as well as brain specializations (Davidesco et al., 2014; O'Connor et al., 2017; Finn et al., 2020; Ramot et al., 2020; Levakov et al., 2021; Malach, 2021).

When examining each cognitive measure on its own, we found stimulus-invariant typicality for all but surprise and importance, yet no clear relationship or trade-off was observed between stability of individual typicality and the robustness of group dynamics (i.e., replicability). At this point, we refrain from inferring about long-term cognitive traits from these data, as further research will be needed to assess whether RBS typicality is informative of stable trait information beyond a single session (i.e., by splitting stimulus sets by testing days).

The current findings call for two potential approaches to examine individual differences utilizing RBS individual typicality. The first approach is applicable when neural (or physiological) responses are recorded during sensory stimulation, e.g., movie viewing, prior to applying RBS to the same stimuli. In this case, RBS can be used to match individual cognitive typicality to its corresponding neural typicality for the same movie, such as the distance between the participant and the group in the activation of a region of interest [intersubject correlation: Hasson et al. (2004)], or its coactivation with another region (Simony et al., 2016; Simony and Chang, 2020). This type of comparison does not rely on stimulusinvariant information, and thus can potentially relate cognitive processes that lack that invariance to simultaneous neural processes, in a manner that reflects subjective fluctuations in the neurocognitive state. Thereby, RBS typicality can be used to explore the neural mechanisms related to subjective processing of specific cognitive dimensions, which may be canceledout in the common group dynamics. The second potential application of RBS typicality extends to stimulus-invariant information. In this case, RBS typicality, being a special form of representational distance measure, could be matched with any marker of individual differences, in any modality, regardless of stimulation (e.g., Grossman et al., 2019). For example, future research could examine the diagnostic value of behavioral naturalistic-stimulation settings, by comparing RBS typicality with performance scores on clinical questionnaires or tasks. A similar approach using representational distances has been successfully applied in previous clinical research (Hahamy et al., 2015).

The measurement trade-off: Real-time versus retrospective reporting

Retrospective behavioral sampling was designed to reconstruct the cognitive processes evoked by naturalistic stimulation, without interfering with the natural experience as it is happening. This is fundamentally different from methods of real-time tracking, such as the continuous rating of emotional intensity while watching a movie (Raz et al., 2012, 2016), or real-time interruption of the movie at fixed intervals to collect emotional ratings (Jaaskelainen et al., 2008). Rather, one of the central goals of RBS was to allow for passive uninterrupted stimulation. Thereby, we avoid the measurement problem, by which the introspective attention and reflection upon one's own experience, required for explicit reporting in real time, risks altering the experience itself (Bohr, 1958; Goldberg et al., 2006; Malach, 2007). Instead of interrupting the natural experience, retrospective reporting relies on one's memory of it. Particularly, RBS was inspired by previous work showing that one's subjective past experience can be reactivated through memory, in great detail that goes beyond episodic recall of the facts to include emotional and introspective perceptions (Berntsen, 2002; Levine et al., 2002). Accordingly, in RBS, the experience of each event is first reactivated in a silent-recall stage, followed by the rating of subjective experience. In addition, a practical advantage of RBS over real-time reporting is the wider scope of information that can be recovered. Unlike real-time reporting, which is limited in the number of simultaneous processes to which one can attend and report on at a given moment, RBS probes a variety of distinct processes that are each recalled from memory in turn.

On the other hand, the reliance on memory for retrospective reporting comes at a price, namely, that we are measuring the memory of an experience, not the experience itself. Hence, in RBS, positive emotion is the memory of feeling good, and surprise is the memory of being surprised, etc. The gap between experience and the memory of it has been the main rational for advocates of real-time rather than retrospective paradigms (Kahneman and Riis, 2005; Schwarz, 2007). By this logic, in the lack of an associated real-time measure, we cannot disentangle RBS real-time validity from memory distortion. Yet RBS was developed with the idea in mind of matching retrospective reports with real-time neuroimaging dynamics. Critically, when matching RBS with any real-time recorded measure, such as neuroimaging, their shared variance across events will represent a real-time component of experience. In this approach, the resultant real-time component will not be contaminated by interruptions to stimulation. By contrast, real-time cognitive reporting by neuroimaging participants will require interference with stimulation during neuroimaging recording, thereby contaminating both measurement modalities in a way that is irrecoverable. We therefore argue that, considering all aspects of the measurement trade-off, RBS is beneficial for achieving high validity of neurocognitive processes, while maximizing the ecological nature of real-time experience.

Open questions for future testing

We note a few open questions that could not be directly addressed with the current data. First, following on the previous section, without an external real-time comparison we cannot measure the difference between the memory of experience measured by RBS, and the real-time experience. Future studies could compare RBS ratings with recent techniques of realtime measurement (e.g., Chang et al., 2021). Second, we cannot rule out the influence of the event reminders themselves on participants' responses [see Loftus and Palmer (1974)]. While this is unlikely, given strong emotional responses even to neutral event reminders (i.e., events that seem neutral when isolated from context, e.g., Event #29; Supplementary Table 1), such potential impact of the cue cannot be ruled out. This may be an inherent limitation of any cued-recall task, as we cannot reliably differentiate the cognitive response to the cue from the recalled memory in participants' reports, as these cannot be separately tested without substantially changing the task from a recall task to a judgement/tagging task. Nevertheless, when matching RBS reports with any real-time measurement (neuroimaging, electrophysiology, etc.), the shared component would represent the stimulus-driven real-time response. Third, we point out that, in RBS, events are pre-defined by the experimenter based on temporal intervals and coherency, potentially missing other moments of the movie that were significant to participants. In a related approach, Spiers and Maguire (2006) assessed retrospective experience from memory by replaying a video of the recalled experience, at a pace determined by the participants, while collecting open-ended recall of their cognitive state. This technique could be of potential use in pinpointing significant moments in the movie, which could be compared with RBS time-courses. Lastly, the RBS questionnaire is long, thereby limiting the number of events that can be tested in a single session. Future studies could assess the necessity, or minimal length, of the silent recall for reporting reliability, which could reduce the time spent on each event in the questionnaire.

Conclusion

Retrospective behavioral sampling is shown to be an informative and reliable method to track natural cognitive

dynamics, successfully tackling the central challenges of naturalistic stimulation; RBS effectively disentangles a wide range of simultaneous cognitive processes, depicts them along the time-course of stimulation, and does this without interfering with the real-time experience. The robust replicability of RBS time-courses makes it a powerful tool to characterize the shared stimulus-driven cognitive experience. These shared dynamics can be applied to ascribe specific cognitive processes to neural and physiological time-courses driven by the same stimuli. RBS was further shown to carry stimulus-invariant information on individual differences, which could be tested against neural and clinical individual markers to study idiosyncratic neurocognitive states and traits. In sum, RBS is a promising new tool to advance the study of naturalistic cognition and its neural and physiological manifestations.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board (IRB) of the Weizmann Institute of Science. The participants provided their written informed consent to participate in this study.

Author contributions

ES: funding acquisition. TB: experimental design, data collection, data analysis, and visualization. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Israel Science Foundation grant 1458/17 to ES.

Acknowledgments

We thank Sharon Pinto and Ronen Pinto for web development supporting the survey distribution and data aggregation and export.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

Baldassano, C., Hasson, U., and Norman, K. A. (2018). Representation of Real-World Event Schemas during Narrative Perception. *J. Neurosci.* 38, 9689–9699. doi: 10.1523/JNEUROSCI.0251-18.2018

Bendixen, A., Schroger, E., and Winkler, I. (2009). I heard that coming: Eventrelated potential evidence for stimulus-driven prediction in the auditory system. *J. Neurosci.* 29, 8447–8451. doi: 10.1523/JNEUROSCI.1493-09.2009

Berntsen, D. (2002). Tunnel memories for autobiographical events: Central details are remembered more frequently from shocking than from happy experiences. *Mem. Cognit.* 30, 1010–1020. doi: 10.3758/bf03194319

Bohr, N. (1958). Atomic physics and human knowledge. Hoboken, NJ: Wiley.

Brandman, T., Malach, R., and Simony, E. (2021). The surprising role of the default mode network in naturalistic perception. *Commun. Biol.* 4:79. doi: 10.1038/ s42003-020-01602-z

Chang, L. J., Jolly, E., Cheong, J. H., Rapuano, K. M., Greenstein, N., Chen, P. A., et al. (2021). Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Sci. Adv.* 7:eabf7129. doi: 10.1126/sciadv.abf7129

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* 20, 115–125. doi: 10.1038/nn.4450

Cutting, J. E., Brunick, K. L., and Candan, A. (2012). Perceiving event dynamics and parsing Hollywood films. J. Exp. Psychol. Hum. Percept. Perform. 38, 1476– 1490. doi: 10.1037/a0027737

Davidesco, I., Zion-Golumbic, E., Bickel, S., Harel, M., Groppe, D. M., Keller, C. J., et al. (2014). Exemplar selectivity reflects perceptual similarities in the human fusiform cortex. *Cereb. Cortex* 24, 1879–1893. doi: 10.1093/cercor/bht038

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

Farah, M. J., Wilson, K. D., Drain, M., and Tanaka, J. N. (1998). What is "special" about face perception? *Psychol. Rev.* 105, 482–498.

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

Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., et al. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *Neuroimage* 215:116828. doi: 10. 1016/j.neuroimage.2020.116828

Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., and Fried, I. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* 322, 96–101. doi: 10.1126/science.1164685

Goldberg, I. I., Harel, M., and Malach, R. (2006). When the brain loses its self: Prefrontal inactivation during sensorimotor processing. *Neuron* 50, 329–339. doi: 10.1016/j.neuron.2006.03.015

Grossman, S., Gaziv, G., Yeagle, E. M., Harel, M., Megevand, P., Groppe, D. M., et al. (2019). Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. *Nat. Commun.* 10:4934. doi: 10.1038/s41467-019-12623-6

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum. 2022.956708/full#supplementary-material

Hahamy, A., Behrmann, M., and Malach, R. (2015). The idiosyncratic brain: Distortion of spontaneous connectivity patterns in autism spectrum disorder. *Nat. Neurosci.* 18, 302–309. doi: 10.1038/nn.3919

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* 28, 2539–2550. doi: 10.1523/JNEUROSCI.5487-07.2008

Jaaskelainen, I. P., Koskentalo, K., Balk, M. H., Autti, T., Kauramaki, J., Pomren, C., et al. (2008). Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. *Open Neuroimage J.* 2, 14–19. doi: 10.2174/ 1874440000802010014

Kahneman, D., and Riis, J. (2005). Living, and thinking about it: Two perspectives on life. Sci. Well Being 1, 285-304.

Levakov, G., Faskowitz, J., Avidan, G., and Sporns, O. (2021). Mapping individual differences across brain network structure to function and behavior with connectome embedding. *Neuroimage* 242:118469. doi: 10.1016/j.neuroimage. 2021.118469

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., and Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychol. Aging* 17, 677–689.

Loftus, E. F., and Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *J. Verb. Learn. Verb. Behav.* 13, 585–589. doi: 10.1080/17470218.2016.1237980

Loftus, E. F., Miller, D. G., and Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. J. Exp. Psychol. Hum. Learn. 4, 19–31.

Malach, R. (2007). The measurement problem in consciousness research. *Behav. Brain Sci.* 30, 516–517.

Malach, R. (2021). Local neuronal relational structures underlying the contents of human conscious experience. *Neurosci. Conscious.* 2021:niab028. doi: 10.1093/nc/niab028

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. Nature 264, 746-748. doi: 10.1038/264746a0

Nguyen, M., Vanderwal, T., and Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *Neuroimage* 184, 161–170. doi: 10.1016/j.neuroimage.2018.09.010

Norman, Y., Yeagle, E. M., Khuvis, S., Harel, M., Mehta, A. D., and Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science* 365:eaax1030. doi: 10.1126/science.aax1030

O'Connor, D., Potler, N. V., Kovacs, M., Xu, T., Ai, L., Pellman, J., et al. (2017). The Healthy Brain Network Serial Scanning Initiative: A resource for evaluating inter-individual differences and their reliabilities across scan conditions and sessions. *Gigascience* 6, 1–14. doi: 10.1093/gigascience/giw011

Ramot, M., Walsh, C., Reimann, G. E., and Martin, A. (2020). Distinct neural mechanisms of social orienting and mentalizing revealed by independent measures of neural and eye movement typicality. *Commun. Biol.* 3:48. doi: 10. 1038/s42003-020-0771-1 Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., et al. (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cogn. Affect. Behav. Neurosci.* 16, 709–723. doi: 10.3758/s13415-016-0425-4

Raz, G., Winetraub, Y., Jacob, Y., Kinreich, S., Maron-Katz, A., Shaham, G., et al. (2012). Portraying emotions at their unfolding: A multilayered approach for probing dynamics of neural networks. *Neuroimage* 60, 1448–1461. doi: 10.1016/j. neuroimage.2011.12.084

Rescorla, R. A., and Wagner, A. R. (1972). "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non reinforcement," in *Classical conditioning II: Current research and theory*, eds A. H. Black and W. F. Prokasy (New York: Appleton-Century-Crofts), 64–99. doi: 10.1177/ 0269881120954052

Schultz, W. (1998). Predictive reward signal of dopamine neurons. J. Neurophysiol. 80, 1–27. doi: 10.1152/jn.1998.80.1.1

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. Sci. Real Time Data Capture Self Rep. Health Res. 11:26.

Simony, E., and Chang, C. (2020). Analysis of stimulus-induced brain dynamics during naturalistic paradigms. *Neuroimage* 216:116461. doi: 10.1016/j. neuroimage.2019.116461 Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., et al. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* 7:12141. doi: 10.1038/ncomms12141

Sonkusare, S., Breakspear, M., and Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends Cogn. Sci.* 23, 699–714. doi: 10.1016/j. tics.2019.05.004

Spiers, H. J., and Maguire, E. A. (2006). Thoughts, behaviour, and brain dynamics during navigation in the real world. *Neuroimage* 31, 1826–1840.

Spiers, H. J., and Maguire, E. A. (2007). Decoding human brain activity during real-world experiences. *Trends Cogn. Sci.* 11, 356–365.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., et al. (2013). The WU-Minn Human Connectome Project: An overview. *Neuroimage* 80, 62–79. doi: 10.1016/j.neuroimage.2013.05.041

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., et al. (2017). Same Story. Different Story. *Psychol. Sci.* 28, 307–319. doi: 10.1177/0956797616682029

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., and Hasson, U. (2017). How We Transmit Memories to Other Brains: Constructing Shared Neural Representations Via Communication. *Cereb. Cortex* 27, 4988–5000. doi: 10.1093/ cercor/bhx202