Title: Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing

Authors: James W. Antony[1], Thomas H. Hartshorne[2], Ken Pomeroy[3], Todd M. Gureckis[4], Uri Hasson[1], Samuel D. McDougle[5], and Kenneth A. Norman[1]

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ

[2]Woodrow Wilson School of Public & International Affairs, Princeton University, Princeton, NJ

[3]www.kenpom.com

[4]Department of Psychology, New York University, New York, NY

[5]Department of Psychology, Yale University, New Haven, CT

Corresponding author:

Dr. James Antony Princeton Neuroscience Institute

jantony@princeton.edu Washington Road

Phone: (262) 347-8224 Princeton, NJ 08544, USA

Pages: 59

Figures: 6

Abstract length: 150 words

Abstract

Surprise signals a discrepancy between predicted and observed outcomes. It is theorized to segment the flow of experience into discrete perceived events, drive affective experiences, and create particularly resilient memories. However, the ability to precisely measure naturalistic surprise has remained elusive. We used advanced basketball analytics to derive a quantitative measure of surprise and characterized its behavioral, physiological, and neural effects on human subjects observing basketball games. We found that surprise served to segment ongoing experiences, as reflected in subjectively perceived event boundaries and shifts in neocortical neural patterns underlying belief states. Interestingly, these effects differed by whether surprising moments contradicted or bolstered current predominant beliefs. Surprise also positively correlated with pupil dilation, processing in subcortical regions associated with dopamine, game enjoyment, and, along with these physiological and neural measures, long-term memory. These investigations support key predictions from event segmentation theory and extend theoretical conceptualizations of surprise to real-world contexts.

Introduction

As events in the world unfold, the brain rapidly adjusts its predictions of what will happen next. Of course, our predictions are not always correct – when they are inaccurate, we experience surprise (i.e., unsigned prediction error[1–5]). Surprise is theorized to be critical for learning and memory[6,7], updating our beliefs about the structure of the world[8], and demarcating events in the continuous flow of time[9]. Moreover, although people typically prefer certainty about outcomes that are instrumental for survival[10], in domains with non-instrumental information, such as narratives, music, and sports, people tend to prefer violations of their expectations[11–13], suggesting that surprise is often a rewarding affective experience.

Although surprise has been elegantly operationalized in laboratory experiments, in more naturalistic settings it is difficult to precisely characterize. Most experiments measure surprise in the context of discrete temporal units (e.g., trials) involving repeated sensory cues, rather than as a probabilistic belief state that is continuously updated over time[14]. Here, we leveraged naturalistic stimuli (video clips of basketball games) in which we could quantify how people continuously update their predictions about an outcome (which team will win). First, we describe and validate our model of surprise. Then, we show how surprise relates to perceived event boundaries, the segmentation of neural event states, and pupil dilation. Finally, we show that surprise predicts both subjective enjoyment and neural signatures of reward and leads to improved long-term memory for events.

Results

Calculating and validating surprise during basketball viewing

Self-ascribed basketball fans (*N* = 20, 6 female) underwent eye tracking and fMRI scans while watching the final five minutes of nine games from the 2012 men's NCAA® college basketball ("March Madness") tournament (Fig 1A). Subjects reported their preferred team (if any) and enjoyment after each game and freely recalled the games from memory after each set of three. We operationalized predictions using a "win probability" metric derived from an expert basketball analyst that was constant within possessions of the game and updated after each score or possession change (Fig 1A, B, top; Fig S1). These predictions were used to compute punctate values of surprise, defined as the absolute value of the derivative of the win probability time course. Note that surprise refers to prediction errors (differences between expected and actual outcomes) that can[15], but do not always[1], lead to belief updates (how much beliefs change as a result); here we operationalize surprise to indicate both a model-based estimate of the prediction error an observer should experience at possession boundaries and how much they should update their beliefs going into the next possession.

To validate that subjects represented these probabilities in some form, at the end of the experiment we presented them with a test featuring scenarios from different 2012 tournament games that they had not previously viewed (Fig 1C). In this test, subjects rated the likelihood that the home team would win for every possession (starting with five minutes left) when given the score, the time remaining, and which team had the ball. The distribution of probabilities in this test resembled those from viewed games (Fig S2). Indeed, subject responses correlated highly with the win probabilities specified by the algorithm (mean $r = 0.82$, range = 0.69 to 0.92, $p < 0.001$), validating its use for computing surprise.

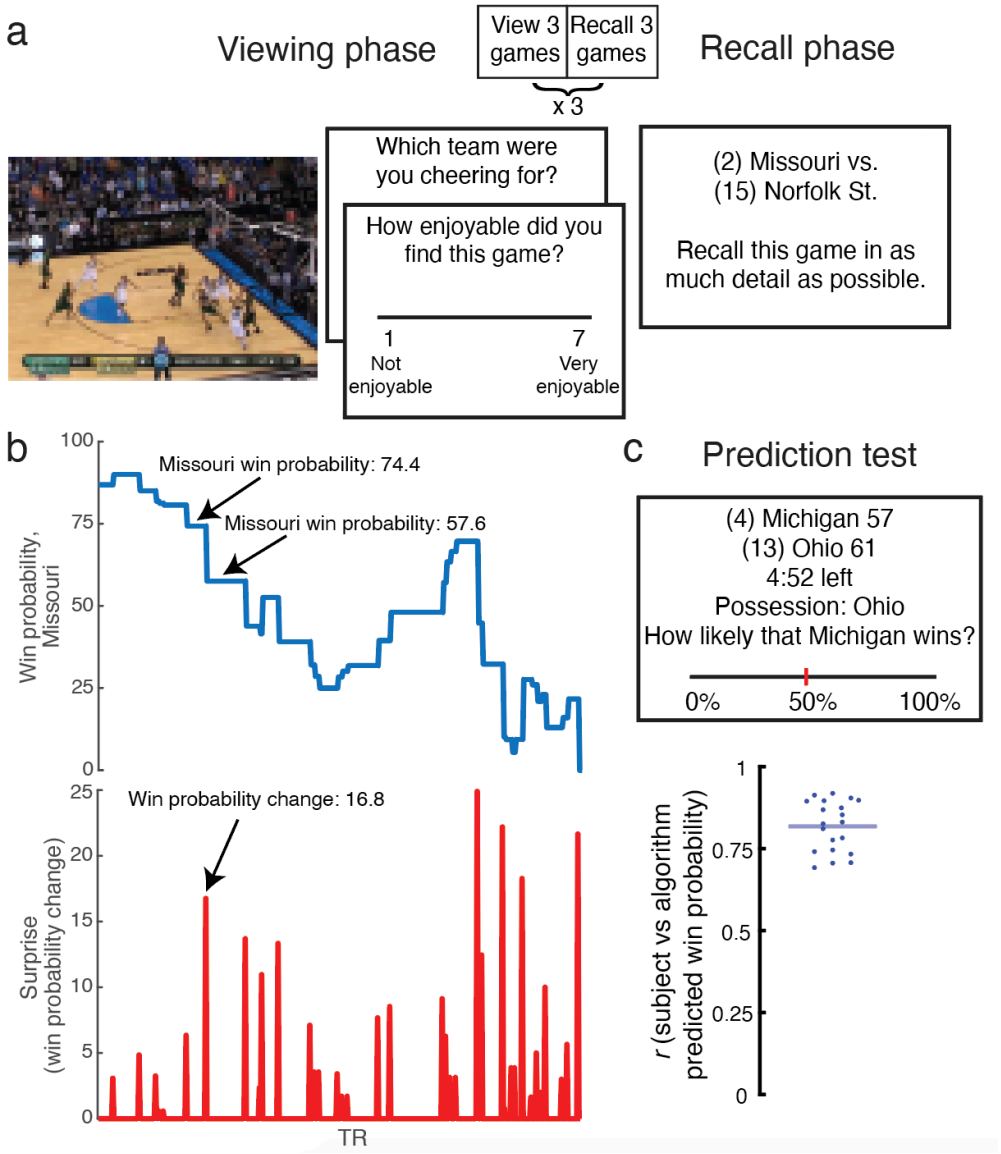

Fig 1. Experimental procedure and surprise calculation and validation. a) Subjects undergoing eye-tracking and fMRI viewed and recalled nine NCAA basketball games. Game screenshot was blurred here for copyright reasons. b) (top) Win probability metrics (used to model subjects' predictions) from an established algorithm chart the likelihood of one team winning across time points in the scanner (TRs). (bottom) Surprise was defined as the unsigned derivative of this time course. c) (top) Subject predictions were tested on scenarios from different games given at the end of the experiment. (bottom) Subject predictions correlated highly with the algorithm.

Surprise increases subjective event segmentation, the segmentation of neural states, and pupil dilation

Event segmentation theory (EST) posits that humans naturally segment their ongoing stream of experience and create internal event models to predict upcoming events[16,17]. Violations of these predictions (i.e., surprises) are thought to create event boundaries, which are reflected by subjective segmentation of continuous perception[16], pronounced shifts in neural states[18], and physiological changes such as pupil dilations[19–23]. Here, we investigated how surprise relates to these measurements.

First, we wanted to examine how a separate group of basketball fans perceived game events and how these judgments related to surprise. These fans (N = 15, 8 female) watched the games outside of the scanner and responded when they perceived the ending of "game units" at the coarsest level that was meaningful to them (see *Methods* for instructions). Overall, subjects responded anywhere from 5 to 150 times (66 ± 11.7) across the 9 games (Fig 2A). Based on these responses, we computed a *subjective boundary agreement* score for each possession boundary (e.g., score or change in possession), quantifying the agreement across subjects (ranging from 0 for no agreement to 1 for perfect agreement) that a game unit ended at that possession boundary. Crucially, supporting the predictions from EST outlined above, subjective boundary agreement correlated with model-based surprise across possession boundaries (r = 0.25, p = 0.01, permutation test shuffling possessions within each game) (Fig 2C).

We next intuited that, while subjects can represent these predictions probabilistically (Fig 1C), they could also binarize their predictions[24], holding a predominant belief of the likely

winner (e.g., team X currently has >50% chance of winning). If so, new evidence that is consistent with this belief (e.g., team X scores) should differ from evidence that is inconsistent with this belief (e.g., team Y scores); inconsistent evidence requires more updating of one's beliefs regarding the future state of the world, and thus should be more likely to trigger segmentation (Fig 2B). Thus, we next correlated subjective boundary agreement across possession boundaries separately with belief-consistent and belief-inconsistent surprise (Fig 2C). The correlation was significant for belief-inconsistent but not belief-consistent surprise, and the two significantly differed (inconsistent: $r = 0.40$, $p < 0.001$; consistent: $r = -0.06$, $p = 0.8$; difference: $r = 0.46$, $p < 0.001$, permutation tests shuffling possessions within each game). Taken together, these results suggest that event segmentation increases with surprise and, rather than simply demarcating probabilistic changes in belief, segmentation is especially robust when new information conflicts with the predominant belief.

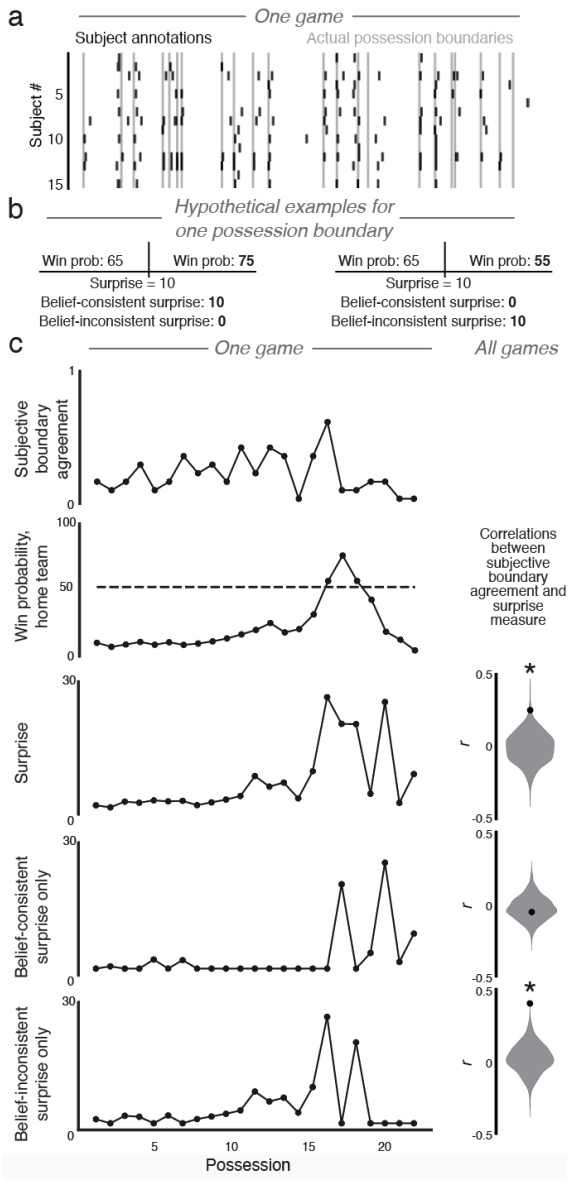

Fig 2: Perceived event boundaries correlate with surprise measures. a) Each subject's annotations demarcating game units are shown in rows (black ticks) amidst actual possession boundaries for one game (gray). (b) Hypothetical examples show how surprise, belief-consistent surprise, and belief-inconsistent surprise were calculated. (c) (left) Subjective boundary agreement (the proportion of subjects who marked an event boundary at each possession boundary) is plotted for one game against win probability for the home team, surprise, belief-consistent surprise, and belief-inconsistent surprise at those boundaries (right). Subjective boundary agreement was significantly correlated with both surprise and belief-inconsistent surprise. *: $p < 0.05$.

We next attempted to capture signatures of the segmentation of continuous experience in neural data, leveraging Hidden Markov models (HMMs) to analyze BOLD responses in fMRI[18,25]. HMMs are data-driven algorithms that probabilistically segment data into stable states and discrete shifts between those states, making them useful for aligning events in the world and events in the brain. Here we used HMMs to address three questions: First, do HMM state transitions naturally align with actual game possession boundaries? Second, following from the hypothesis that surprise leads to event segmentation, does surprise predict more frequent HMM state transitions across possessions and does greater cumulative surprise lead to more neural event states across games? Finally, are HMM state transitions more likely to occur at moments of belief-inconsistent versus belief-consistent surprise? We addressed these questions using neural activity from primary visual cortex (V1), precuneus, and ventromedial prefrontal cortex (vmPFC) as *a priori* regions-of-interest (ROIs). We predicted that V1 would track small time-scale changes with clear sensory features[18,26], like possession changes, and could also be modulated by higher-level processes[14,27]. For precuneus, a node in the default mode network, we predicted state changes would occur on a longer time-scale and reflect subjective segmentations, as was found previously[18]. Finally, we predicted that representations in vmPFC, a higher-level region involved in abstract inference and the representation of abstract states[28–31], would change on the longest time-scale and more frequently with higher surprise, reflecting changes in the broader narrative of the game.

In agreement with our prediction, the time course of neural state transitions in V1 was significantly correlated with the time course of ground-truth possession changes, which involve substantial changes in visual features (Fig 3B; $r = 0.18$, $p < 0.001$, via permutation tests

preserving possession lengths within each game). This effect was not found in our high-order ROIs, precuneus (r = 0.07, p = 0.26) and vmPFC (r = 0.04, p = 0.51).

Next, we asked whether surprise leads to more neural event states both across games and possessions. Across games, we correlated the mean surprise per minute in each game and the cross-validated best-fitting number of states per minute (averaged across subjects) in each neural region. This relationship was significant in vmPFC (r = 0.73, p = 0.03, via permutation tests that shuffle games) but not in V1 (r = 0.58, p = 0.16) nor precuneus (r = 0.38, p = 0.30). Representational shifts did not correlate with possible confounding factors such as the number of possessions per minute (vmPFC: r = -0.42, p = 0.65) nor the number of camera angle changes per minute (vmPFC: r = -0.10, p = 0.85), and the relationship with mean surprise per minute remained significant in a regression model controlling for these factors (p = 0.04, via permutation tests that shuffle games) (Fig 3B).

To investigate the effects of surprise at the finer temporal resolution of possessions, we first calculated (for each region of interest, and for each possession boundary) HMM *state transition agreement*, operationalized as the proportion of subjects showing at least one HMM state transition in the 15-s window spanning the possession boundary (±7 s around the boundary). Then, we correlated this value with surprise across possession boundaries. In contrast to the first analysis in this section, which looked at whether HMM state transitions were more likely to occur at possession boundaries (vs. other time points), this analysis "zooms in" on possession boundaries and asks whether the occurrence of HMM state transitions at these time points is modulated by surprise. We found that vmPFC state transition agreement at possession boundaries was significantly correlated with surprise (r = 0.21, p = 0.02, via

permutation tests that shuffle surprise across possessions within each game), but this relationship was not significant in V1 (r = 0.11, p = 0.19) nor precuneus (r = 0.16, p = 0.12) (Fig 3D). Therefore, vmPFC pattern shifts occur more frequently with greater surprise.

We next asked if HMM state transition agreement differentially correlated with belief-consistent versus belief-inconsistent surprise across possession boundaries. Correlations were significant for belief-inconsistent surprises in all three regions (V1: r = 0.29, p = 0.008; precuneus: r = 0.24, p = 0.02; vmPFC: r = 0.25, p = 0.02), but they were not significant for belief-consistent ones (V1: r = -0.12, p = 0.34; precuneus: r = -0.04, p = 0.89; vmPFC: r = 0.03, p = 0.58; difference between belief-inconsistent and belief-consistent: V1: r = 0.41, p = 0.01; precuneus: r = 0.29, p = 0.04; vmPFC: r = 0.22, p = 0.12) (Fig 3D, lower right). These results accord with neural predictions of EST[9], specifically that vmPFC holds higher-level predictive representations, and that surprising outcomes – especially those that counter the current belief and thereby increase uncertainty – "reset" neural representations of event models[32]. Finally, we asked whether HMM state transition agreement correlated with subjective boundary agreement across subjects, given that both measures correlated with surprise (and specifically belief-inconsistent surprise) from the underlying model. We found that HMM state transition agreement in precuneus (r = 0.31, p < 0.001) and V1 (r = 0.27, p = 0.001) correlated with subjective boundary agreement, whereas vmPFC did not (r = 0.02, p = 0.66).

In sum, neural states in these three regions appear to change on different time-scales, and they are most strongly modulated by different variables. V1 state changes occur rapidly and successfully track possession boundaries overall, but also become more likely at surprising event boundaries. This may be a consequence of modulation from higher-level regions[27].

Precuneus changes occur at a moderate rate and appear especially well-tuned for tracking subjective event boundaries[18,33]. Lastly, vmPFC changes occur sparsely (less than once per minute), but they occur preferentially at surprising moments and, in particular, when events conflict with the current belief about the likely game winner. This result is consistent with vmPFC "resetting" during moments of behavioral uncertainty[34] or salient changes in environmental structure[32,35,36].

Additionally, we performed exploratory analyses asking how HMM state transitions correspond to the above surprise measures in other parts of the neocortex by repeating these across-possession analyses in 48 bilateral cortical parcels from the Harvard-Oxford Brain Atlas (Fig S3, full details in Supplementary Table 1). These results were largely consistent with those above: The frontal medial cortex parcel, which best overlaps our vmPFC ROI, was among the regions showing the strongest correlations for both surprise (r = 0.20, p = 0.14, permutation test shuffling possessions within each game; after cuneal cortex) and belief-inconsistent surprise (r = 0.24, p = 0.04; after cuneal cortex, precuneus, posterior parahippocampal gyrus, posterior medial temporal gyrus, temporo-occipital inferior temporal gyrus, and lingual gyrus). The precuneus showed strong correlations with belief-inconsistent surprise (r = 0.26, p = 0.02) and subjective boundary agreement (r = 0.28, p = 0.001). The angular gyrus also strongly aligned with subjective boundary agreement (r = 0.21, p = 0.005), in line with previous accounts[18,33].

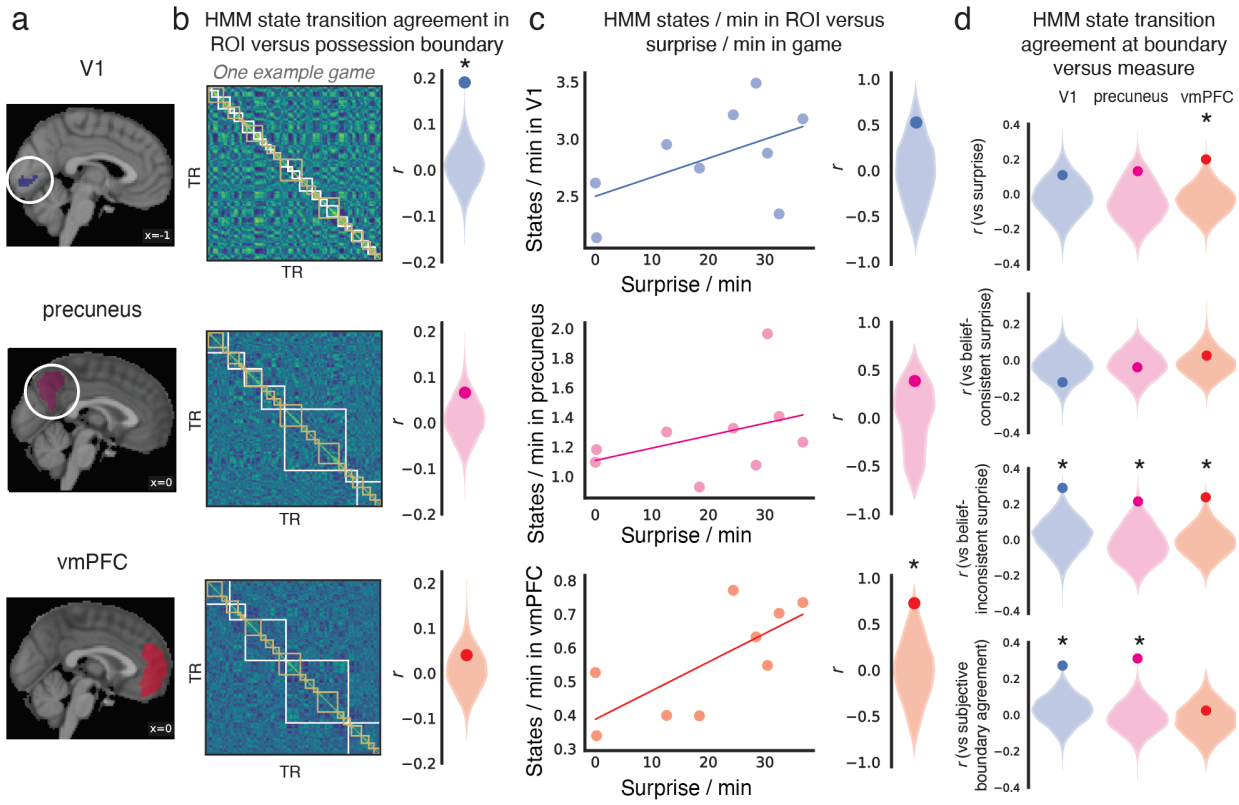

Fig 3: Surprise predicts neural event boundaries. a) V1 (top), precuneus (middle), and vmPFC (bottom) ROIs are shown. b) (left) TR-by-TR, 2-D spatial correlation matrices in each ROI plotted over time reveal off-diagonal, “block-like” structure relating to different brain states demarcated by Hidden Markov model (HMM) boundaries (white). Actual possession boundaries are shown in gold. (right) Correlations between the time series of HMM state transitions and the time series of true possession boundaries are plotted, compared to null distributions. This relationship was significant in V1. c) Correlations across games (dots) show the number of states per minute in each region versus surprise per minute (left), which was significant in vmPFC (right). d) Across possession boundaries, HMM state transition agreement (the proportion of subjects showing a HMM state transition in the time window around the boundary) was correlated with surprise, belief-consistent surprise, belief-inconsistent surprise, and subjective boundary agreement for each ROI, compared to null distributions. Surprise was significantly (positively) correlated with HMM state transition agreement in vmPFC. Belief-inconsistent (but not belief-consistent) surprise was significantly correlated with HMM state transition agreement in all three regions. Subjective boundary agreement correlated significantly with HMM state transition agreement in V1 and precuneus.

Previous accounts have linked event segmentation and surprise with pupil dilation[19–23], which we investigated next. There are challenges analyzing pupil dilation with video stimuli, as pupil area measurements with conventional eye trackers differ by the gaze location of the eye[37], decrease with global and local visual luminance[38], and increase with salient sounds[39]. We addressed the first challenge by normalizing the measurements within x- and y-coordinate bins according to the gaze location (Fig 4A). We addressed the second and third challenges by including the following sensory variables for each second of the game broadcast in a linear model relating surprise to pupil dilation: global luminance of the entire video, local luminance surrounding the eye location, and the auditory envelope from the broadcast (Fig 4B).

On average, subjects' pupils dilated following possession changes ($p < 0.001$; Fig 4C). More importantly, if pupil dilation reflects event segmentation, pupil dilation across an event boundary should scale with surprise at that boundary. To address this, we asked how surprise, along with the above sensory variables, predicted pupil area change across a boundary using a linear mixed-effects model with subject as a random effect. In line with EST, surprise significantly predicted pupil area change ($p < 0.001$; Fig 4D). Both luminance metrics were also significantly related to dilation ($p = 0.002$, both), but the auditory envelope was not ($p = 0.18$). Surprise effects did not significantly differ based on whether it was consistent ($p = 0.002$) or inconsistent ($p = 0.009$) with the predominant belief (difference: $p = 0.53$). Therefore, pupil dilation modulations differed somewhat from the observed neural effects, reflecting surprise in a manner invariant to the predominant belief.

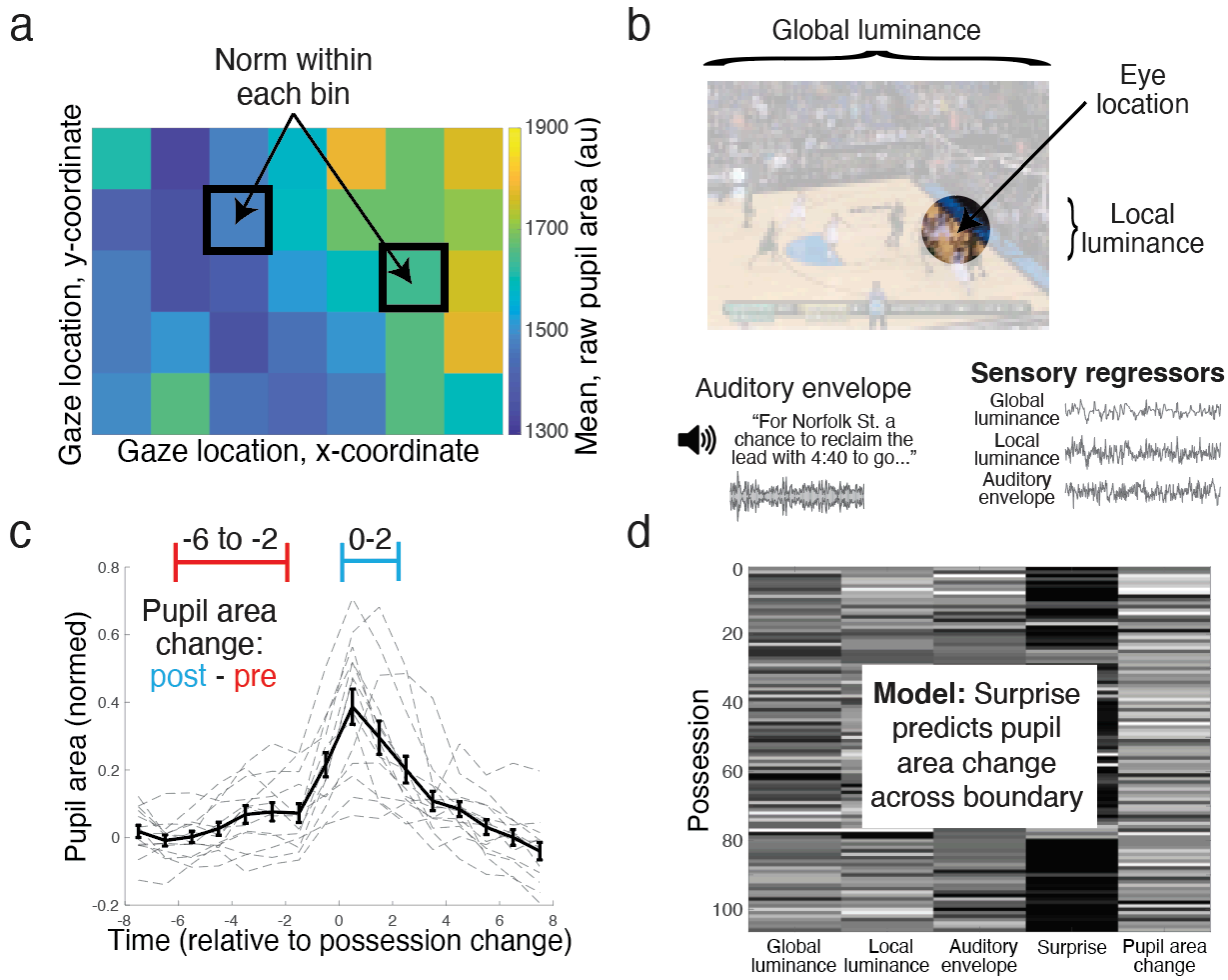

Fig 4. Surprise increases pupil dilation across possession boundaries. a) Confounds due to gaze location were addressed by splitting the screen into x- and y-coordinate bins and norming the raw pupil area values within each bin. b) (top) Luminance for each TR was calculated as an average for the entire screen (global luminance) as well as for a 2° radius around the current location of the eye (local luminance). (bottom left) The auditory envelope was averaged within each TR based on the broadcast volume. (bottom right) These three variables were included in analyses related to pupil area changes. c) Pupil area (averaged for each TR) increased after possession changes. Pupil area change, computed by subtracting a post- (0 to 2 s) minus pre-possession (-6 to -2 s) time period, significantly increased across the boundary. d) Variables entering a mixed-effects model are shown for one subject. This model showed that surprise was significantly related to pupil area change.

Surprise correlates with enjoyment and neural activity in reward-related regions.

We next examined if and how surprise relates to subjective enjoyment, as well as activity in neural regions associated with reward. Casual fans typically prefer games with high uncertainty and surprise to combat boredom[13,40]. Indeed, in many circumstances, subjects find tasks to be less boring when events are not perfectly predictable[41,42]. Accordingly, subjects' enjoyment ratings given after each game correlated significantly with that game's mean level of surprise ($r = 0.81$, $p = 0.008$; Fig S4). We asked whether subjects valued surprise in the absence of direct experience – that is, would they be more excited to watch games merely expected to contain more surprise? At the end of the experiment, subjects rated their excitement to watch games from the following year's tournament (2013), starting from five minutes remaining. We computed future expected surprise for each of these games by finding games from the 2012 corpus that had similar win probabilities at five minutes remaining and then summed the total amount of surprise remaining in those games. All subjects preferred watching games with higher expected surprise in the future (mean $r = 0.60$, range: 0.12 to 0.92, $p < 0.001$; Fig S4). Thus, viewers appear to value both the experience and the expectation of surprise.

In addition to considering enjoyment across a full game, we also investigated the neural effects of surprise on a shorter time scale. Reward signals are intimately linked with the activity of dopamine neurons in regions of the brainstem such as the ventral tegmental area (VTA), as well as targets of those neurons, particularly the nucleus accumbens (NAcc). Classically, the VTA[43,44] and NAcc[12,45,46] respond strongly when rewards are larger or earlier than expected (i.e., reward prediction errors, or RPEs). However, the VTA can respond more broadly to variables other than reward[47], including sensory PEs[48–50], unexpected events[51,52], aversive PEs[53], changes

in hidden belief states[31,54], reward expectation[55,56], advance information[57], and stimulus-stimulus learning[58], all in the absence of (or controlling for) reward.

Subjects reported having a team preference in approximately half of the games (48 ± 6%) (Fig 5A). Therefore, in addition to looking at unsigned surprise in all games, for games with a stated team preference, we can ask whether neural activity was modulated by the valence of the outcome, or "signed" surprise. We initially characterized activity in NAcc and VTA (Fig 5B) for each subject using two regressors in separate general linear models (GLMs) - unsigned surprise for all games, and signed surprise for games with a stated preference (Fig 5C). We predicted the NAcc would respond to RPEs as signed surprise, whereas the VTA could respond either to RPEs or more broadly to unsigned surprise. In NAcc, activity was significantly correlated with signed surprise ($p = 0.005$, *t*-test of GLM betas against zero) but not unsigned surprise ($p = 0.81$), consistent with the standard RPE model (Fig 5D). By contrast, a more complex pattern emerged from VTA, which responded significantly to unsigned surprise ($p < 0.001$) and marginally to signed surprise ($p = 0.08$). Of note, the VTA responded significantly to both belief-consistent and belief-inconsistent surprise separately (both $p < 0.001$), but not differentially ($p = 0.72$).

To tease apart whether the signed response in NAcc was driven by positive and/or negative events, and whether the response in VTA actually increased to both positive and negative events separately, we included two more GLMs with signed regressors reflecting only positive or only negative events (which we modeled as positive to reduce confusion). NAcc responded positively to positive events ($p = 0.003$) but not to negative events ($p = 0.41$), whereas VTA responded both positively to positive events ($p = 0.001$) and positively to negative

events (p = 0.009). Importantly, given the hemodynamic lag inherent to the fMRI signal, our GLM analyses do not allow us to disentangle whether the VTA responses were solely driven by the possession outcome or a scaled anticipatory process that ramps shortly before the outcome (e.g., when a basketball shot is in the air) (Fig S5). Lastly, control analyses showed that unsigned VTA and signed NAcc responses were better fit using our surprise-based model versus a binarized, unscaled model that reflected binary valence (Fig S6). In sum, we replicated the classic finding that NAcc responses scaled with RPEs using a novel naturalistic stimulus, whereas VTA responses increased with unsigned naturalistic surprise.

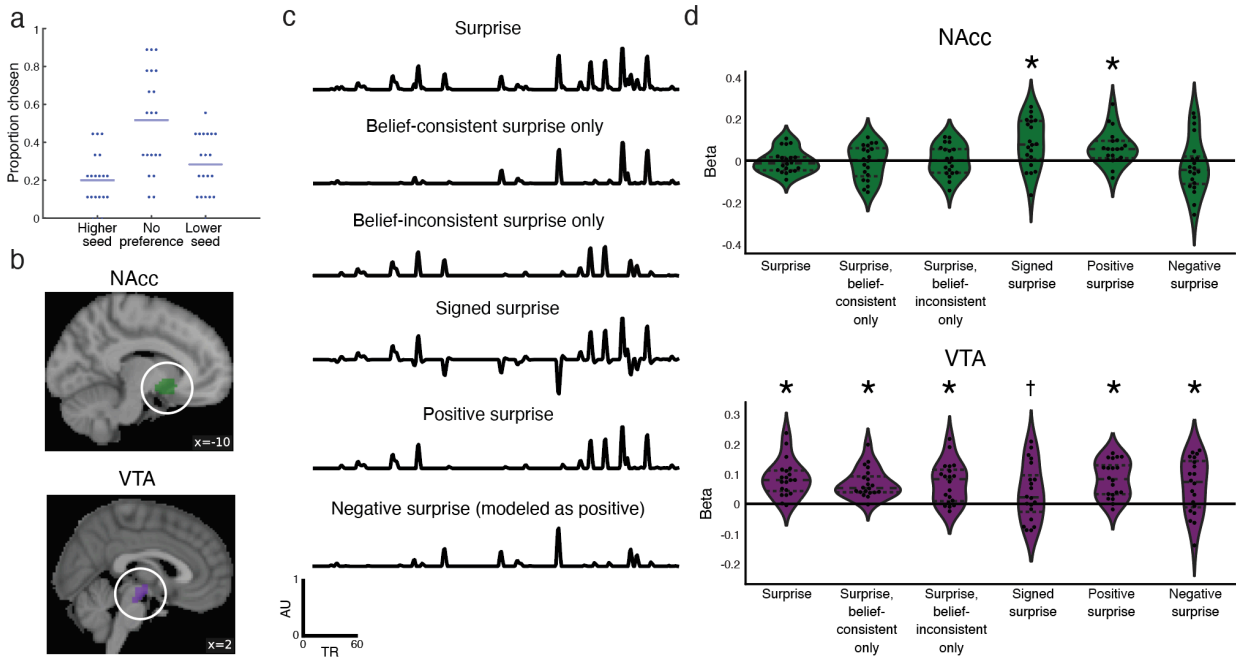

Fig 5: Surprise correlates with neural activity in reward-related regions. a) Proportions of games in which subjects preferred the higher seed, lower seed, or had no preference showed that subjects had a preference in approximately half of the games. b) NAcc and VTA ROIs. c) Depictions for one game of different models included in general linear model analyses (GLMs) for how surprise relates to neural activity in reward-related regions. (top three) For all games, we modeled unsigned surprise, as well as belief-consistent and belief-inconsistent surprise. (fourth from top) For games with a preference, we modeled signed surprise, which was positive when the preferred team scored and negative when the non-preferred team scored. (bottom two) For these games, we also modeled only the positive and negative events. d) The NAcc responded only in a signed fashion to positive events (top). The VTA responded positively to unsigned surprise, including a similar response to belief-consistent and belief-inconsistent surprise and a strong response to positive and negative events independently (bottom). Dotted lines indicate quartiles of the distribution. *: $p < 0.05$; †: 0.05 $p < 0.1$

Surprise, pupil area change, and neural event boundaries positively predict memory

EST posits that the segmentation of ongoing experience into discrete events enhances memory for information near the event boundaries[2,59–61]. Given that surprise helps create event boundaries[9] and enhances memory in laboratory settings[2,3,5,62], we asked how it and other factors described above predict long-term memory in our naturalistic paradigm.

We assessed memory by computing the number of possessions subjects recalled with enough specific details that they could be readily identified by an independent rater (Fig 6A). Overall, subjects recalled few possessions (12.0 ± 2.8 out of 157), likely due to high interference from the similar content in the clips. Additionally, subjects' recall tended to show temporal contiguity[63], preferentially transitioning forward from one possession to the next (Fig 6B).

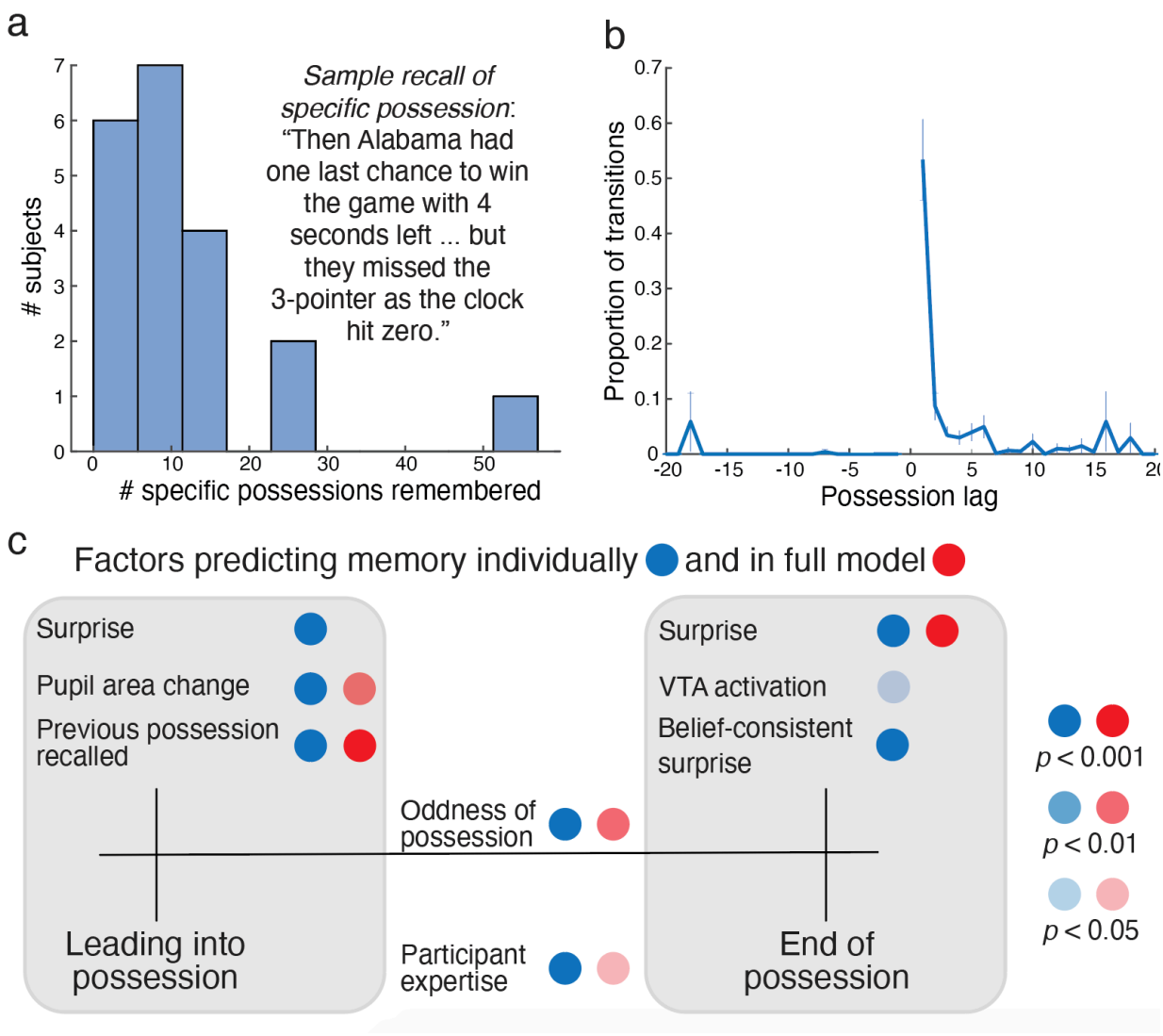

Fig 6: Effects of surprise and physiological and neural factors on memory. a) Sample recall of a specific possession and the distribution of the number of possessions recalled across subjects characterize memory performance. b) Transitions between successively recalled memories show a strong forward asymmetry from possession-to-possession. Error bars represent SEM. c) Schematic of factors that predict memory individually (blue circle) and that survive in a full mixed-effects logistic regression model (red circle) shows that a wide range of variables predict memory for possessions, including pupil area increases leading into a possession, participant expertise, (independently rated) oddness of a possession, and surprise at the end of a possession. Note that all of these significant relationships were in the positive direction.

Theoretically, surprise and other factors occurring at possession boundaries could affect memory of the upcoming possession, or the possession just completed – that is, a proactive or retroactive enhancement, respectively. Our first prediction was that surprise at the end of the possession would influence memory; this prediction is based on prior studies that found a relationship between neural activity at the end of an event and subsequent memory for the just-completed event[18,64]. Indeed, we found significant correlations between the memorability of each possession across subjects and end-of-possession surprise ($r = 0.36$, $p = 0.005$, permutation test shuffling possessions within each game). Additionally, the VTA has been linked to storage of long-term episodic memories[65] and peak VTA activation could predict memory on a trial-by-trial basis for the just-ending possession (see Fig S5). However, surprise and/or other factors leading into the possession could also influence memory for subsequent events. For instance, surprise re-orients attention[66] and increases subsequent learning rates[67,68]. Similarly, pupil dilations also predict higher learning rates[39,69,70]. As such, many factors could influence long-term memory of events in the game.

We thus submitted the above factors to individual mixed-effects logistic regression modeling for predicting possession memory, and each significant individual factor was included in a full model (Fig 6C). These factors were modeled as predictors for memory of both the prior possession and the just-completed possession. These results are shown schematically in Fig 6D and are detailed in Table S2.

Critically, pupil area changes across the prior possession boundary predicted subsequent memory, in line with attentional reorienting and/or an increased learning rate. In contrast, peak VTA activation predicted memory for the just-ending possession, though this did

not survive in the full model. Surprise at both the end and the start of the possession predicted memory for the possession when submitted individually; surprise at the end of the possession survived as a significant predictor of memory in the full model, but surprise at the start of the possession did not. Therefore, our analysis suggests that the memory-related effects of surprise leading into a possession may be mediated by its influence on these other factors. Belief-consistent (but not belief-inconsistent) surprise at the end of the possession also predicted memory, and in a separate model, it did so significantly better than belief-inconsistent surprise ($p < 0.001$); however, belief-consistent surprise did not survive in the full model. Additionally, we included the following factors, all of which predicted memory individually and in the full model: subject expertise, based on the number of games watched across one's lifetime, highlighting the importance of domain expertise on memory[71]; the "oddness" of a possession, based on an independently-rated assessment of unusual basketball plays (e.g., a lane violation during a critical free throw), which could index "surprise" in a manner not captured by our win probability metric; and, following the temporal contiguity effects depicted in Fig 6B, whether or not the previous possession was recalled. Altogether, these results capture complex, multifaceted aspects of real-world memory and highlight the importance of surprise, among other associated factors, in shaping those memories.

Discussion

Our findings reveal that the popular activity of competitive sports viewing is a valuable model of naturalistic surprise, and our analyses of this task led to multiple novel behavioral and physiological discoveries in support of the tenets of event segmentation theory (EST)[9,16].

Namely, surprises appear to strongly drive the segmentation of internal event representations, indexed by increased subjective perception of event boundaries, increased pupil dilation, an increased likelihood of significant neural representational shifts (as measured using a Hidden Markov Model), and increased subsequent memory for events.

Further results provide novel evidence for reinforcement learning models of reward prediction errors (RPEs) in a naturalistic setting. In the nucleus accumbens, we observed classic RPE effects reflecting dynamic changes to continuous predictions of the winner of a basketball match[46]. In the ventral tegmental area, we found that activity generally correlated with a valence-independent updating of beliefs, extending previous work to a naturalistic setting[31,48,49,58,72]. Finally, in showing that surprise correlates with subjective enjoyment, we provide support for the intriguing idea that, when information is not instrumental for survival, humans may actually prefer unpredictable scenarios[11,41].

Importantly, our approach illustrates a marked dissociation of surprise effects based on whether they bolstered or contradicted the current belief. Belief-inconsistent surprise better predicted subjective event boundaries and transitions between neural states in regions such as V1 and vmPFC, while belief-consistent surprise better predicted long-term memory. However, measures such as pupil dilations and signaling in dopamine-related regions showed no such distinction, responding in a graded fashion with increasing surprise in a manner that was invariant to the current predominant belief. Ultimately, the fact that different flavors of surprise have different behavioral, physiological, and neural outcomes demonstrates that individuals' predictions may have both a binary aspect (i.e., which team will win?) and a

probabilistic one (i.e., how likely is it?)[24]. These discrepancies raise the question of how and where these two putative aspects of predictions diverge.

Lastly, one noteworthy aspect of our investigation is that the probabilistic, latent predictions derived from our model can be validated behaviorally (as shown in our prediction test, Fig 1C). We speculate that similar latent belief states underlie people's responses to real-world events in other domains, including news, fiction, and film – people are elated and dismayed in proportion to their surprise at breaking news stories or narrative swings, are more likely to enjoy exciting stories that have multiple twists and turns, and their most long-lasting memories are formed in precisely those moments where their beliefs substantially shift. Future studies should continue to leverage naturalistic stimuli with quantifiable latent variables to investigate how humans respond to an ever-changing world.

References

1. O'Reilly, J. X. *et al.* Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proc. Natl. Acad. Sci.* **110**, E3660-9 (2013).
2. Rouhani, N., Norman, K. A., Niv, Y. & Bornstein, A. M. Reward prediction errors create event boundaries in memory. *bioRxiv* (2019).
3. Rouhani, N., Norman, K. & Niv, Y. Dissociable effects of surprising rewards on learning and memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **44**, 1430 (2018).
4. Kutas, M. & Hillyard, S. A. Brain potentials during reading reflect word expectancy and semantic association. *Nature* **307**, 161–3 (1984).
5. Jang, A. I., Nassar, M. R., Dillon, D. G. & Frank, M. J. Positive reward prediction errors during decision-making strengthen memory encoding. *Nat. Hum. Behav.* **3**, 719–732 (2019).
6. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. in *Classical Conditioning II: Current Research and Theory.* (eds. Black, A. H. & Prokasy, W. F.) 64–99 (Appleton-Century-Crofts, 1972).
7. Sinclair, A. H. & Barense, M. D. Surprise and destabilize: Prediction error triggers episodic memory updating. *Learn. Mem.* **25**, 369–381 (2018).
8. Sutton, R. S. & Barto, A. G. *Introduction to reinforcement learning.* (Cambridge: MIT press., 1998).
9. Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M. & Gershman, S. J. Structured event memory: a neuro-symbolic model of event cognition. *bioRxiv* (2019).

10. Bromberg-Martin, E. S. & Hikosaka, O. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* **63**, 119–26 (2009).

11. Ely, J., Frankel, A. & Kamenica, E. Suspense and surprise. *J. Polit. Econ.* **123**, 215–60 (2015).

12. Gold, B. P. *et al.* Musical reward prediction errors engage the nucleus accumbens and motivate learning. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 3310–5 (2019).

13. Gan, S., Tuggle, C. A., Mitrook, M. A., Coussement, S. H. & Zillman, D. The thrill of a close game: Who enjoys it and who doesn't? *J. Sport Soc. Issues* **21**, 53–64 (1997).

14. Hutchinson, J. B. & Barrett, L. F. The power of predictions: An emerging paradigm for psychological research. *Curr. Dir. Psychol. Sci.* **28**, 280–91 (2019).

15. Faraji, M., Preuschoff, K. & Gerstner, W. Balancing new against old information: The role of puzzlement surprise in learning. *Neural Comput.* **30**, 34–83 (2018).

16. Zacks, J. M., Kurby, C. A., Eisenberg, M. L. & Haroutunian, N. Prediction error associated with the perceptual segmentation of naturalistic events. *J. Cogn. Neurosci.* **23**, 4057–66 (2011).

17. Zacks, J. M. *et al.* Human brain activity time-locked to perceptual event boundaries. *Nat. Neurosci.* **4**, 651–5 (2001).

18. Baldassano, C. *et al.* Discovering event structure in continuous narrative perception and memory. *Neuron* **95**, 709-721.e5 (2017).

19. Clewett, D., Gasser, C. & Davachi, L. Dynamic arousal signals construct memories of time and events. *bioRxiv* 765214 (2019). doi:10.1101/765214

20. Preuschoff, K., 't Hart, B. M. & Einhäuser, W. Pupil dilation signals surprise: Evidence for

noradrenaline's role in decision making. *Front. Neurosci.* **5**, 1–12 (2011).

21. Braem, S., Coenen, E., Bombeke, K., van Bochove, M. E. & Notebaert, W. Open your eyes for prediction errors. *Cogn. Affect. Behav. Neurosci.* **15**, 374–80 (2015).

22. Filipowicz, A. L. S., Glaze, C. M., Kable, J. W. & Gold, J. I. Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *bioRxiv* 1–39 (2019).

23. Yu, A. J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681–92 (2005).

24. Johnson, S. G. B., Merchant, T. & Keil, F. C. Belief digitization: Do we treat uncertainty as probabilities or as bits? *J. Exp. Psychol. Gen.* (2020).

25. Chang, L. J. *et al.* Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *bioRxiv* 487892 (2019).

26. Lerner, Y., Honey, C. J., Silbert, L. J. & Hasson, U. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* **31**, 2906–15 (2011).

27. Hindy, N. C., Ng, F. Y. & Turk-Browne, N. B. Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nat. Neurosci.* **19**, 665–7 (2016).

28. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron* **81**, 267–79 (2014).

29. Hampton, A. N., Bossaerts, P. & O'Doherty, J. P. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.* **26**, 8360–7 (2006).

30. Takahashi, Y. K. *et al.* Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nat. Neurosci.* **14**, 1590–7 (2011).

31. Starkweather, C. K., Gershman, S. J. & Uchida, N. The medial prefrontal cortex shapes dopamine reward prediction errors under state uncertainty. *Neuron* **98**, 1–14 (2018).

32. Nassar, M. R., McGuire, J. T., Ritz, H. & Kable, J. W. Dissociable forms of uncertainty-driven representational change across the human brain. *J. Neurosci.* **39**, 1688–98 (2019).

33. Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A. & Hasson, U. How we transmit memories to other brains: Constructing shared neural representations via communication. *Cereb. Cortex* **27**, 4988–5000 (2017).

34. Karlsson, M. P., Tervo, D. G. R. & Karpova, A. Y. Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* **338**, 135–139 (2012).

35. Durstewitz, D., Vittoz, N. M., Floresco, S. B. & Seamans, J. K. Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* **66**, 438–48 (2010).

36. D'Acremont, M., Schultz, W. & Bossaerts, P. The human brain encodes event frequencies while forming subjective beliefs. *J. Neurosci.* **33**, 10887–97 (2013).

37. Hayes, T. R. & Petrov, A. A. Mapping and correcting the influence of gaze position on pupil. *Behav. Res. Methods* **48**, 510–27 (2017).

38. Page, H. E. The relation between area of stimulation and intensity of light at various levels of visual expectation as measured by pupil construction. *J. Exp. Psychol.* **29**, 177–200 (1941).

39. Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15**, 1040–6 (2012).

40. Peterson, E. M. & Raney, A. A. Reconceptualizing and reexamining suspense as a

predictor of mediated sports enjoyment. *J. Broadcast. Electron. Media* **52**, 544–62 (2008).

41. Geana, A., Wilson, R. C., Daw, N. & Cohen, J. D. Boredom, information-seeking and exploration. *CogSci* 1751–6 (2015).

42. Wilson, R. C., Shenhav, A., Straccia, M. & Cohen, J. D. The Eighty Five Percent Rule for optimal learning. *Nat. Commun.* **10**, 1–9 (2019).

43. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–9 (1997).

44. D'Ardenne, K., McClure, S., Nystrom, L. & Cohen, J. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* **319**, 1264–1267 (2008).

45. Rutledge, R. B., Dean, M., Caplin, A. & Glimcher, P. W. Testing the reward prediction error hypothesis with an axiomatic model. *J. Neurosci.* **30**, 13525–13536 (2010).

46. Cikara, M., Botvinick, M. M. & Fiske, S. T. Us versus them: Social identity shapes neural responses to intergroup competition and harm. *Psychol. Sci.* **22**, 306–13 (2011).

47. Engelhard, B. *et al.* Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* **570**, 509–13 (2019).

48. Takahashi, Y. K. *et al.* Dopamine neurons respond to errors in the prediction of sensory features of expected rewards. *Neuron* **95**, 1395-1405.e3 (2017).

49. Howard, J. D. & Kahnt, T. Identity prediction errors in the human midbrain update reward-identity expectations in the orbitofrontal cortex. *Nat. Commun.* **9**, 1–11 (2018).

50. Iglesias, S. *et al.* Hierarchical prediction errors in midbrain and basal forebrain during

sensory learning. *Neuron* **80**, 519–30 (2013).

51. Barto, A., Mirolli, M. & Baldassarre, G. Novelty or surprise? *Front. Psychol.* **4**, 1–15 (2013).

52. Kafkas, A. & Montaldi, D. Striatal and midbrain connectivity with the hippocampus selectively boosts memory for contextual novelty. *Hippocampus* **25**, 1262–1273 (2015).

53. Matsumoto, M., Matsumoto, K., Abe, H. & Tanaka, K. Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.* **10**, 647–56 (2007).

54. Nour, M. M. *et al.* Dopaminergic basis for signaling belief updates, but not surprise, and the link to paranoia. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E10167–E10176 (2018).

55. Totah, N. K. B., Kim, Y. & Moghaddam, B. Distinct prestimulus and poststimulus activation of VTA neurons correlates with stimulus detection. *J. Neurophysiol.* **110**, 75–85 (2013).

56. Kim, Y., Simon, N. W., Wood, J. & Moghaddam, B. Reward anticipation is encoded differently by adolescent ventral tegmental area neurons. *Biol. Psychiatry* **79**, 878–86 (2016).

57. Bromberg-Martin, E. S. & Hikosaka, O. Lateral habenula neurons signal errors in the prediction of reward information. *Nat. Neurosci.* **14**, 1209–18 (2011).

58. Sharpe, M. J. *et al.* Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nat. Neurosci.* **20**, 735–42 (2017).

59. Newtson, D. & Engquist, G. The perceptual organization of ongoing behavior. *J. Exp. Soc. Psychol.* **12**, 436–450 (1976).

60. DuBrow, S. & Davachi, L. The influence of context boundaries on memory for the

sequential order of events. *J. Exp. Psychol. Gen.* **142**, 1277–86 (2013).

61. Swallow, M. K., Jeffrey, Z. M. & Richard, A. A. Event boundaries in perception affect memory encoding and updating. *J. Exp. Psychol. Gen.* **138**, 236 (2010).

62. Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y. & Mendelsohn, A. Knowledge acquisition is governed by striatal prediction errors. *Nat. Commun.* **9**, 1–14 (2018).

63. Howard, M. W. & Kahana, M. J. A distributed representation of temporal context. *J. Math. Psychol.* **46**, 269–99 (2002).

64. Ben-Yakov, A., Eshel, N. & Dudai, Y. Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *J. Exp. Psychol. Gen.* **142**, 1255–63 (2013).

65. Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B. & Gabrieli, J. D. E. Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* **50**, 507–17 (2006).

66. Itti, L. & Baldi, P. Bayesian surprise attracts human attention. *Vision Res.* **49**, 1295–306 (2009).

67. Pearce, J. M. & Hall, G. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–52 (1980).

68. Courville, A. C., Daw, N. D. & Touretzky, D. S. Bayesian theories of conditioning in a changing world. *Trends Cogn. Sci.* **10**, 294–300 (2006).

69. Silvetti, M., Seurinck, R., van Bochove, M. E. & Verguts, T. The influence of the noradrenergic system on optimal control of neural plasticity. *Front. Behav. Neurosci.* **7**, 1–6 (2013).

70. Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X. & Bishop, S. J. Anxious individuals

have difficulty learning the causal statistics of aversive environments. *Nat. Neurosci.* **18**, 590–6 (2015).

71. Chi, M. T. Knowledge structures and memory development. in *Children's thinking: What develops?* (ed. Siegler, R.) 75-96. (Erlbaum, 1978).

72. Starkweather, C. K., Babayan, B. M., Uchida, N. & Gershman, S. J. Dopamine reward prediction errors reflect hidden-state inference across time. *Nat. Neurosci.* **20**, 581–9 (2017).

73. Musz, L., Bellana, B., Zuo, X. & Chen, J. Shared and converging brain responses across people during self-guided exploration of complex photographs and paintings. in *Society for Neuroscience conference* (2019).

74. Eldar, E., Cohen, J. D. & Niv, Y. The effects of neural gain on attention and learning. *Nat. Neurosci.* **16**, 1146–53 (2013).

75. Choplin, N. T. & Edwards, R. P. *Visual fields.* (Slack Incorporated., 1998).

76. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–6 (2019).

77. Behzadi Y, Restom K, Liau J & Liu TT. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).

78. Nastase, S. A., Gazzola, V., Hasson, U. & Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* **14**, 669–87 (2019).

79. Simony, E. *et al.* Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* **7**, 1–13 (2016).

80. Murty, V. P. *et al.* Resting state networks distinguish human ventral tegmental area from

substantia nigra. *Neuroimage* **100**, 580–9 (2014).

81. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, 208–19 (2004).

82. Kumar, M. *et al.* BrainIAK tutorials: User-friendly learning materials for advanced fMRI analysis. *PLoS Comput. Biol.* **16**, e1007549 (2020).

83. 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**, 2142–54 (2012).

84. Clasen, P. C., Beevers, C. G., Mumford, J. A. & Schnyer, D. M. Cognitive control network connectivity in adolescent women with and without a parental history of depression. *Dev. Cogn. Neurosci.* **7**, 13–22 (2014).

Methods

fMRI subjects. Twenty subjects (6 female, 18-35 years old) with normal or corrected-to-normal vision and fluent in English were recruited via campus flyers and word-of-mouth. Subjects were given hourly monetary compensation for participating ($20/hr). Written informed consent was obtained in a manner approved by the Princeton University Institutional Review Boards. All subjects self-professed to having seen or played in more than 50 basketball games (across all competitive levels of the sport).

Stimuli. The last five minutes from all 32 Round-of-64 2012 NCAA® tournament games were acquired as audiovisual files for a fee and with permission from Wazee Digital™. These files were down-sampled to visual dimensions of 1280 x 720 and frame rate of 29.75 Hz and audio dimensions of 48,000 Hz for computational efficiency when presented using the MATLAB Psychtoolbox software. These games were additionally edited to reduce their overall length by eliminating breaks in action (other than brief intervals preceding free throws and in-bound passes) in a manner that did not significantly compromise their overall comprehension, resulting in clips between 5:29 and 7:32 in length. Tournament games from the Round-of-64 were used for the following reasons: tournament teams are given seeds that inform subjects about the teams' relative strengths, which should aid subjects' win probability estimations; tournament games have a heightened sense of importance relative to the regular season, enhancing subject engagement; the Round-of-64 is an early round that we intuited subjects would be unlikely to remember (if they had seen or read about the original broadcasts). Nine games were selected for presentation in the scanner using the following criteria: games were

selected to have as wide-ranging amounts of surprise as possible (Fig S1); games were also selected to have as wide-ranging tournament seeds as possible, except that no games between #1 and #16 seeds were selected, as no #1 seed had ever lost before 2012 – each type of matchup between #2 versus #15 and #8 versus #9 seeds was selected at least once; games involving extremely well-known teams (e.g., North Carolina) were given a lower priority to reduce the likelihood that a subject may have remembered the outcome; a yet-lower priority was given to having three games between teams with the same seed (here, #7 versus #10 was chosen) in case seed proved to be a relevant factor over which direct comparisons were required, though none were analyzed here. The selected games/seeds were as follows: #2 Missouri versus #15 Norfolk State; #3 Florida State versus #14 Saint Bonaventure; #4 Indiana versus #13 New Mexico State; #5 Wichita State versus #12 Virginia Commonwealth University; #6 University of Nevada-Las Vegas versus #11 Colorado; #7 Gonzaga versus #10 West Virginia; #7 Saint Mary's versus #10 Purdue; #7 Notre Dame versus #10 Xavier; #8 Creighton versus #9 Alabama.

Design and procedure. The experiment included four phases: game viewing, free recall, prediction test, and anticipated surprise preference test (Fig 1). Subjects arrived in the scanner suite, learned about MRI safety, gave informed consent to participate, listened to the instructions, and then entered the magnet.

The main section of the experiment consisted of three alternations between game viewing and free recall phases. Game viewing phases included three games in succession as part of one scanning run. To keep run length approximately even, the nine videos were pseudo-

randomly shuffled so that one of the longest three, one of the shortest three, and one of the remaining three were shown in each run. Games were presented in the middle 80% of the screen so the subject's view would not be obstructed. Audio was delivered via scanner earbuds and volume levels were tested before the games using an internet video. Audio from the regular broadcast, including crowd noise and commentary, was included to maintain a naturalistic viewing experience. Subjects were asked to simply watch the games, though they were told beforehand that they would later be asked to freely recall the games using as much specific detail as possible. After each game, they were asked to indicate the team for which they were cheering or if they had no preference. We emphasized that they should not feel obligated to prefer one team or another. After this question, they indicated how exciting they found the game from 1 (not exciting) to 7 (exciting).

During recall phases, subjects freely recalled the three previously seen games in the viewing order as part of one scanning run. Subjects were shown text of the two teams involved in the game and were asked to recall the game in as much detail as possible. For instance, they were asked to include the score and approximate amount of time left during any possession they could remember and any parts of that possession as they unfolded (e.g., a screen, a pass, a drive, a defender's move, the outcome of a shot).

After three rounds of viewing and free recall, subjects remained in the scanner to receive anatomical and field map scans while they performed two more behavioral tasks. In the first, subjects were asked to rate the likelihood of teams winning for a series of possessions from five games (Fig 1C). These games were also taken from the 2012 tournament and were chosen to evenly distribute win likelihoods and resemble the likelihood distributions of the

viewed games as much as possible (Fig S2). Subjects were shown the teams, their seeds, their scores, which team had the ball, and the amount of time remaining, and they were asked to slide a joystick from 0-100% to indicate the likelihood that the higher-seeded team would win. These scenarios were updated after every possession, so each game had multiple ratings. The selected games/seeds were as follows: #2 Duke versus #15 Lehigh; #4 Michigan versus #13 Ohio; #5 New Mexico versus #12 Long Beach State; #6 Cincinnati versus #11 Texas; #8 Memphis versus #9 Saint Louis. We correlated these likelihoods against the algorithm to assess prediction abilities. Data were omitted from the few instances in which subjects knew the outcome.

In the second task, subjects rated (from 1-7) how excited they would be to watch 28 games starting with five minutes remaining from the Round-of-64 in the 2013 tournament (Fig 5A). All games were selected, except for games between #1 and #16 seeds were selected. We found the anticipated level of surprise by finding games from the 2012 regular season corpus with similar win probabilities with five minutes remaining and calculating the actual amount of surprise remaining in those games. Then, we correlated these excitement ratings with anticipated surprise for each subject (Fig 5A).

After these tasks, subjects left the scanner and completed a questionnaire. Questions included approximately how much basketball they had watched (i.e. their expertise), their most and least favorite teams (none were involved in the games they viewed except for one subject's third favorite team), whether they knew or suspected the outcomes of any games that they viewed or that were part of the prediction tests (two subjects knew the outcomes for one viewed game each; two subjects knew the outcome of one game in the belief test), their level of overall engagement (6.55±0.25 out of 10), and whether they found any parts of the games

particularly surprising (two subjects mentioned an unusual lane violation during a critical free throw).

Win probability metric. In order to approximate the "ground truth" of win probabilities, we obtained estimates that were originally created for a basketball analytics website using a custom, proprietary algorithm by Ken Pomeroy (K.P.; www.kenpom.com). These estimates were created using the score difference between the teams, the team in possession of the ball, the amount of time left, and the strength of the two teams. We obtained data from the entire 2012 season via personal correspondence, although the data are available on K.P.'s website.

All probabilities are updated in K.P.'s algorithm after every possession. While in reality, a subject may update their win probability estimates for a team within a possession (e.g., if the ball is passed to the team's best player or a pass is made providing for a wide-open shot), under the large majority of situations, we assumed these probabilities to be stable within the possession and therefore these within-possession parameters were not separately modeled. However, we made some exceptions where we believed both a more fine-grained granularity could be obtained at a discrete moment in time by making interpolated inferences on these estimations in systematic ways. For instance, it is highly likely that an ideal subject would increase their belief in a team winning after that team was fouled while shooting; however, the estimations were limited to the time when the other team obtained the ball after the free-throws, when the outcome of the free-throws were known. Therefore, the ideal observer would hold a different belief than the one provided at the beginning of that possession. In the above scenario, if the probability that the home team would win changed in K.P.'s algorithm

from 50% at the beginning of the possession to 58% after the possession and the free-throws were both made, these numbers were averaged after the foul and before the first free-throw (e.g., 54%). This updated probability was again averaged with the post-possession estimation after the first free-throw (e.g., 56%) before assuming the final state of 58% for that possession after the second free-throw was made. The specifics of all other rare, exceptional scenarios are documented and shared openly along with the behavioral and neural data in this study.

Surprise calculation and related metrics. As an agent traverses the world, they proceed through a series of "states." In viewing basketball, one could conceptualize a momentary state as the current score, the team in possession of the ball, the relative strength of the two teams, and the amount of time left in the game. In each state, an ideal observer, via repeated experience (e.g., watching games), could form a refined, probabilistic belief in some outcome (e.g., which team will win) via an iterative process like temporal difference learning[8].

As formulated elsewhere[11], surprise is the change in prediction from the previous to the current state (here, the change in win likelihood). For the purposes of aligning surprise as a psychological phenomenon with physiological and neuroimaging data, surprise was labeled as '0' for all stable time points (e.g., within a possession) and the magnitude of the surprise scaled with changes in belief state (resembling a "stick" function with spikes of varying magnitudes; Fig 1B).

Some analyses separated "belief-consistent" from "belief-inconsistent" surprise. To do this, we first found the team with a higher than 50% chance of winning. Events that increased this probability were classified as belief-consistent, and events that decreased it were classified

as belief-inconsistent. There were no differences between the mean size of surprise for these constructs (belief-consistent: 7.1 ± 0.12; belief-inconsistent: 6.6 ± 0.08, $p = 0.66$).

Note that in our analyses we used a point estimate of prediction to calculate surprise rather than finding prior and posterior distributions of predictions and computing the difference between them to derive Bayesian surprise[66]. For completeness, we also calculated this metric. To obtain the prior, we found, in the 2012 regular season corpus of games, all instances of games that had a similar state (i.e., similar win probability and the amount of time remaining in the game). From this, the prior was simply the distribution of probabilities in the next states (from all of these corpus instances). The posterior distribution was calculated identically except starting from the next actual state in the game. Then Bayesian surprise was calculated as the KL divergence between prior and posterior distributions. The correlation across possessions between the point estimate of surprise and the Bayesian surprise was $r = 0.95$, indicating that these appear to be highly similar constructs in our data set. We decided to use the point estimate for two reasons. First, it would be difficult to test the average subject's distribution of beliefs on the belief test, whereas asking for point estimates of belief proved simple and reliable. Second, because the correspondence between the two estimates was extremely high, we believe the relative simplicity of the point estimate is desirable.

Free recall scoring. Subject recall recordings were sampled at 11,025 Hz using built-in MATLAB software and converted for transcription. The following characteristics were labeled for each TR in which subjects spoke: recall type – (1) veridical recall specific to particular possession, (2) gist-based, summarized recall that is temporarily imprecise but accurate, (3) in accurate detail

that is nonetheless about the game, (4) irrelevant commentary about the game, and (5) words that are not related to recall[73]; which (if any) game was being recalled; if (1) for recall type, which possession is being recalled, or if (2), which is the first of numerous possessions that are being summarized; if (2) for recall type, which is the last of numerous possessions that are being summarized; if (1) for recall type, does the recall relate to the first or the second half of the possession?

Memory metrics. We calculated the temporal contiguity effect (Fig 6B) by finding, for each position after the initial one, the lag (in number of possessions) with respect to the prior recalled possession.

For logistic regression analysis predicting memory across possessions, we initially ran individual models including all of the following factors: surprise, pupil area change, peak ventral tegmental area (VTA) activation and HMM state transition probability in vmPFC for transitions leading into the possession and at the end of the possession, previous possession memory, oddness of possession, and subject expertise. All analyses for factors leading into possessions had no data (NaN values) for the first possession of the video clip. Peak VTA activation was calculated as the value at each possession at its peak of -2 s minus a baseline averaged from -12 to -8 s (Fig S6). “Oddness” of possessions was determined independently for unusual basketball plays (e.g., a lane violation during a critical free throw). Only 3 out of 157 possessions were labeled this way. Subject expertise was coded as 1-3 depending on the number of games played in/watched over the lifetime (8 subjects had watched 50-200 games; 9 subjects, 200-1000 games; 3 subjects, over 1000 games). Individual factors that were significant entered a

combined model and insignificant factors were dropped individually in reverse order of their significance until a final full model was determined (Fig 6C and Table S2).

Event segmentation behavioral experiment. 15 subjects (8 female, 18-35 years old) with similar characteristics were recruited as above and were given hourly monetary compensation for participating ($10/hr for behavioral studies). Written informed consent was similarly obtained. Subjects watched the same games as those viewed in the scanner, though in this case they watched all 9 games consecutively. They performed the following task as they watched: "Click the mouse when, in your judgment, one unit of the game ends. Mark off the game you'll be seeing into the largest units that seem natural and meaningful to you." These subjects performed the prediction test, but did not perform recall nor the future surprise preference test. We labeled mouse clicks occurring within 2 seconds after a possession change as an endorsement of it as an event boundary and aligned these responses with other possession-level metrics (Fig 2A). Specifically, for each boundary, we computed a *subjective boundary agreement* score, quantifying the agreement across subjects (ranging from 0 for no agreement to 1 for perfect agreement) that a game unit ended at that possession boundary.

Physiological measurements and analysis. Eye tracking data were acquired in the scanner at 1000 Hz using EyeLink 1000 software (SR Research, Inc., Mississauga, Ontario Canada). Eye tracking data from 6 subjects were lost due to technical difficulties. Extreme outliers ($z > 5$) and time periods when the gaze location was off-screen were removed, and after exclusion missing

pupil area and eye location data were interpolated by averaging around the errant time periods.

Pupil area measurements are influenced by both gaze location and video luminance. Normally, experimenters measuring gaze location only present stimuli centrally or in fixed locations (e.g.,[74]); however, we reasoned that restricting subjects gaze while watching videos would artificially affect their viewing experience. Instead, we accounted for gaze location confounds by binning x- and y-coordinates and z-scoring pupil area within each bin (Fig 2A). We addressed confounds related to video luminance by calculating the mean global gray-scaled luminance for the entire screen and local gray-scaled luminance corresponding to the approximate visual angle of the fovea (2°)[75] surrounding the current eye location for each second of video. We addressed the confound of auditory volume by averaging the auditory envelope of the broadcast for each second of video. These visual and auditory sensory regressors were all calculated within-subject and were included in the general linear model relating surprise to pupil area.

For time course analyses locked to possession boundaries, normed pupil area, blink onsets, and saccade onsets were averaged within each second and then across trials. *T*-tests were then performed across subjects on the following differences between time intervals: for pupil area, -6 to -2 s versus 0 to 2 s; for blinks, -2 to 0 s versus 0 to 2 s; for saccades -2 to -1s versus -1 to 0 s. Blink and saccade analyses were largely exploratory. To relate pupil area measurements across possession boundaries to surprise, we entered changes in pupil area, local luminance, global luminance, and auditory envelope into a mixed-effects, linear regression

model using subject as a random effect. We also created similar models using blink and saccade differences rather than pupil area.

FMRI acquisition and preprocessing. Neuroimaging data were acquired on a 3T full-body Siemens Prisma scanner with a 64-channel head coil, using a T2*-weighted echo planar imaging (EPI) pulse sequence (simultaneous multislice factor 4, no in-plane acceleration, TR 1000 ms, flip angle 59°, TE 30 ms, whole-brain coverage 56 slices of 2.5 mm thickness, in-plane isotropic resolution of 2.5 mm, FOV 195 mm). The first preprocessing steps were performed using FMRIprep (https://github.com/poldracklab/fmriprep[76]), including motion correction, susceptibility distortion correction (using field maps or the 'use-syn-sdc' flag in their absence), brain tissue segmentation, and coregistration and affine transformation of the functional volumes to the 1 mm isotropic T1w anatomical and subsequently to MNI space.

The data were imported, down-sampled to a 3 mm isotropic resolution and three scans from the beginning of each run were discarded. Next, the data were smoothed using SUSAN smoothing (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/SUSAN, Python code adapted from: https://github.com/INCF/BrainImagingPipelines/blob/master/bips/workflows/gablab/wips/scripts/modular_nodes.py) with a 5-mm full width-half maximum spatial kernel. Next, the data were masked using an across-run averaged mask, z-scored, high-pass filtered (128 s cutoff), and confound variables [movement in three directions, rotation in three directions, framewise displacement, and six components used to correct for the influence of physiological noise[77]]

were regressed out at the run level. Data from the first 10 s of each video were removed to avoid confounds related to strong onset responses[78].

Regions of interest (ROIs). Binarized V1, precuneus, and vmPFC ROIs were obtained from a previous dataset[79]. V1 indicates early visual cortex and nearby voxels created using voxels near the calcarine sulcus with strongest ISC. Precuneus and vmPFC were calculated using whole-brain functional connectivity using posterior cingulate cortex as the seed, separating the default mode network into 10 parts with significant correlations and labeling the masks based on cluster location. The NAcc ROI was obtained using an association test in Neurosynth (threshold: z>10) based on the term, "nucleus accumbens". A probabilistic ventral tegmental area (VTA) ROI was obtained from [80] via personal correspondence, and a 75% probability threshold was used to binarize it, resulting in 37 voxels. Additionally, HMM analyses were performed using parcels from the Harvard-Oxford Brain Atlas package from FSL 5.0[81].

Hidden Markov model (HMM) analyses. The start of each video was locked to a pulse from the scanner to align the hemodynamic response for each moment in the movie across subjects. In an additional preprocessing step, the imaging data were shifted 5 s (5 TRs) for later alignment with marked events.

All HMM analyses were performed using the Brainiak toolbox function, brainiak.eventseg.event.EventSegment[82]. Each HMM state was composed of a particular mean activity pattern across all voxels within the region, and each instance of a neural pattern is assumed to be normally distributed around this mean. Following prior work[18], this particular

analysis function imposes the constraint that the HMM cannot re-visit a state once it leaves that state. In other words, each new neural pattern is either assigned to the same state as the previous time step, or a new (not-previously-visited) state. To train the HMMs within each ROI/parcel, we first found the best-fitting number of states for each game. We tested versions of the model with different numbers of states ranging from 1 to 30 with the goal of obtaining the best model to apply to a withheld subject in a cross-validated fashion. We used 14 subjects in our training set and 5 subjects in our validation set, and the best-fitting number of states was determined using the log likelihood of fit to the validation set. Then, the model was fit to the withheld test subject using this number of states. Full details of the basic model are described elsewhere[18].

We were first interested in where the HMM placed state transitions and how this aligned with possession boundaries and surprise at those boundaries. To address the question of whether the HMM state transitions aligned with possession boundaries, we first averaged the HMM state transition time course (=1 when there is a transition, 0 otherwise) across subjects; we then smoothed both the averaged HMM state transition time course and the possession boundary time course (binary 1/0) by taking running averages using a moving window of ± 7 s; finally, we correlated the two smoothed time courses. To assess the strength of these correlations versus chance, we created null distributions scrambling the length of each possession within a game 10,000 times and compared the true versus scrambled distributions (Fig 3B).

Next, we were interested in whether HMM boundaries increase with surprise, which we computed both across games and possessions. Across games, we correlated the best-fitting

number of states per minute for each game with mean surprise in that game, and we compared this with null distributions by scrambling the mean surprise for each game 10,000 times (Fig 3C). Control correlations were also run to verify these correlations could not be explained by the number of possessions or number of camera angle changes (which were enumerated separately by the experimenter), and linear regression analyses including all of these factors were run to investigate whether the surprise effect survived when these factors were considered. Significance was assessed in a similar manner by comparing the true betas to those in permutation tests.

Across possessions, we asked whether HMM state transitions occurred more often at surprising possession boundaries. To answer this question, we first calculated (for each region of interest, and for each possession boundary) HMM *state transition agreement*, operationalized as the proportion of subjects showing at least one HMM state transition in the 15-s window spanning the possession boundary (±7 s around the boundary) (Fig 3D). Next, we correlated HMM state transition agreement with the amount of surprise at each boundary. Finally, we compared this correlation with null distributions whereby we scrambled the surprise assigned to each possession within each game. In other analyses, we repeated this procedure for belief-consistent and belief-inconsistent surprise, all against similar scrambled null distributions. We also correlated HMM state transition agreement with subjective boundary agreement from the other group of subjects. For HMM analyses applied to parcels from the Harvard-Oxford Brain Atlas, we calculated false discovery rate values for each parcel and analysis (q).

General linear model (GLM) analyses. We created parametric regressors by aligning discrete events with each video. These events were convolved using the canonical HRF to create a time course for the regression. For signed analyses, positive values indicated positive events for the preferred team, and vice versa. When no team was preferred in these analyses, the entire game was modeled as a row of zeros. For the positive- and negative-only analysis, negative and positive events were modeled zeros, respectively. For the unscaled analyses, all events were given the same binary positive or negative value. Data scrubbing[83] was performed by finding TRs with a framewise displacement above 0.33 and removing the segment starting 1 TR before and 2 TRs after[84]. The following regressors entered each GLM: the amount of game remaining in seconds of game time, which team was in possession of the ball, one of the various surprise metrics, the auditory envelope, and global luminance. GLMs were performed using the linear_model.LinearRegression function in Python's scikitlearn toolbox. For each analysis, betas were calculated for each subject and effects were measured using *t*-tests against zero. Scaled versus unscaled models were compared using within-subject *t*-tests.

Additional information on statistical analyses. For permutation analyses, we assessed significance by whether true values fell below the 2.5th or above the 97.5th percentile of the null distribution. For mixed-effects linear and logistic regression models, we used subject as a random effect and other variables as fixed effects. Linear and logistic models were created using 'lmer' and 'glmer' functions in R, respectively, with the latter specified as binomial distributions.

Data availability. The datasets and analysis code from this study will be made available at the time of publication on the Open Science Framework website. Neuroimaging data are shared in the brain imaging data structure (BIDS) format amenable to meta-analyses and reproducible neuroscience.

Acknowledgments: This work was supported by a CV Starr fellowship to JWA and the ONR MURI grant N00014-17-1-2961 to KAN and UH. We thank Wazee Digital and the NCAA for game footage, Lisa Musz for the free recall scoring rubric, James Howard and Jeff Zacks for helpful comments on drafts of this manuscript, and Chris Baldassano, Kelly Bennion, Silvy Collin, Nick Depinto, Robert Hawkins, Manoj Kumar, Qihong Lu, Rolando Masis-Obando, Lizzie McDevitt, Anne Mennen, Sebastian Michelmann, Ida Momennejad, Mark Pinsk, Victoria Ritvo, Nina Rouhani, Monika Schönauer, and Jamal Williams for assisting with data collection and/or various aspects of this project.

Author contributions: J.W.A. conceived the experiment. S.D.M., U.H., and K.A.N. contributed to experimental design early on and numerous analysis ideas. J.W.A. programmed the experiment, collected most of the data, and performed the bulk of the analyses. T.H. scored the recall data. K.P. computed and provided all win probability metrics. J.W.A., S.D.M., & K.A.N. wrote the manuscript. All authors discussed the results and revised the paper.

Declaration of interests: K.P. runs a profitable sports analytics website from which the win probability metrics were obtained. However, his role in this project was limited to sharing and discussing these metrics. Furthermore, although he may benefit from larger exposure, there is no foreseeable commercial benefit he would obtain from the results of this publication.

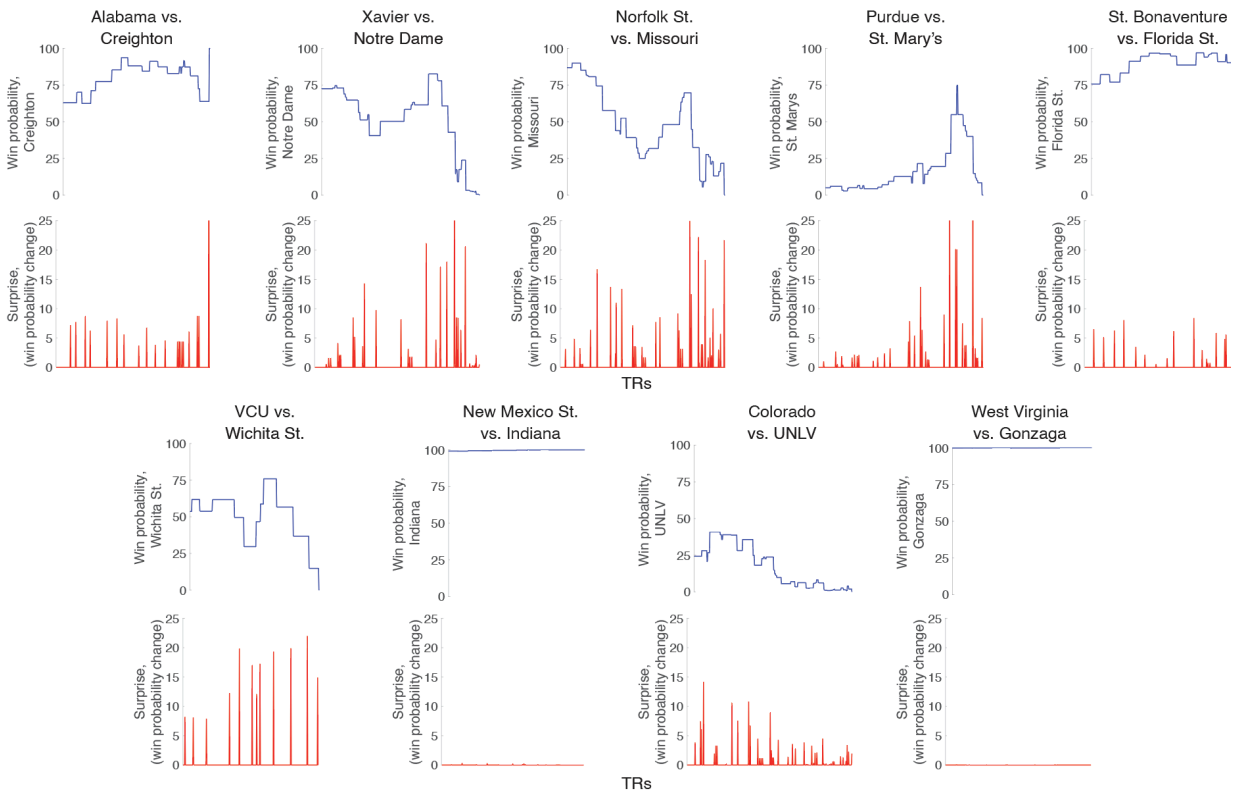

Fig S1. Win probability and surprise values for all nine games. Similar to Fig 1B.

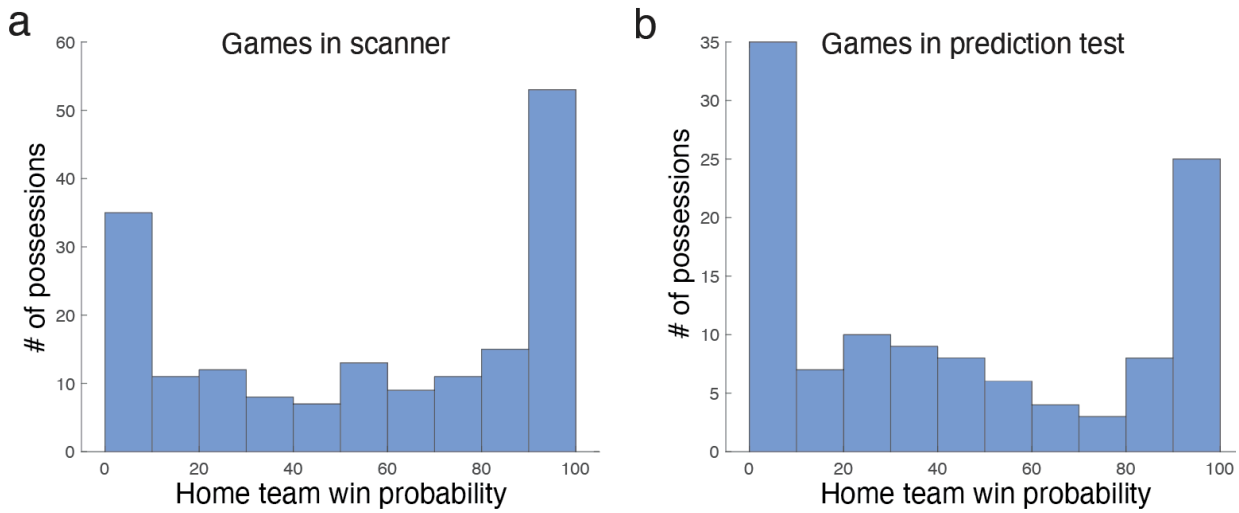

Fig S2. Distribution of win probabilities. Plotted are the number of possessions with win probability values binned every 0.1 for games used in the scanner (a) and in the prediction test (b).

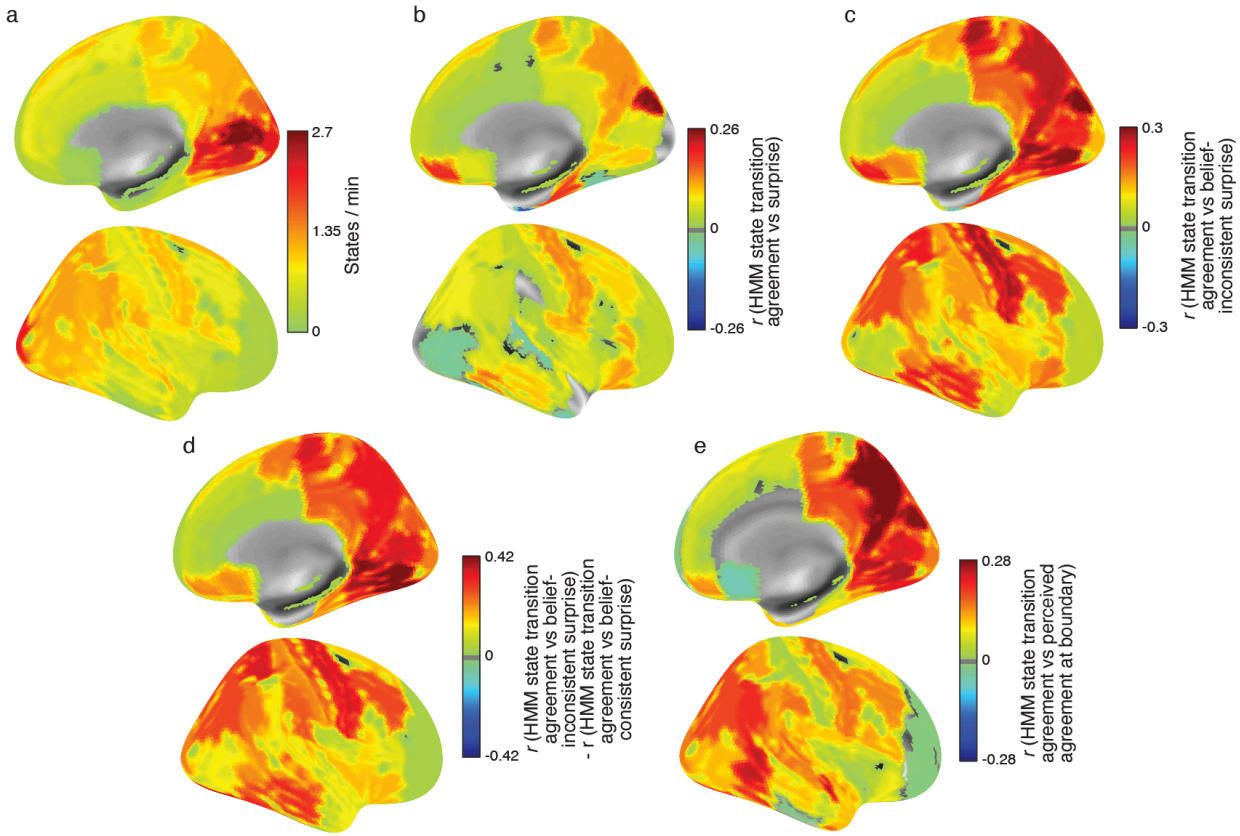

Fig S3. Exploratory analyses applying Hidden Markov models (HMMs) to 48 cortical parcels. a) Number of states per minute in each parcel reveals a cortical hierarchy, with the highest number of states in early visual regions and fewer states in higher-level regions. b-c) Correlation between HMM state transition agreement and surprise (b) / belief-inconsistent surprise (c) at possession boundaries. d) Difference in the correlation between HMM state transition agreement and belief-inconsistent surprise, and the correlation between HMM state transition agreement and belief-consistent surprise. e) Correlation between HMM state transition agreement and subjective boundary agreement at possession boundaries. Data are shown for every parcel and these correlations resemble those done for a priori ROIs in Fig 3D. All values, including statistical significance, are in table S2.

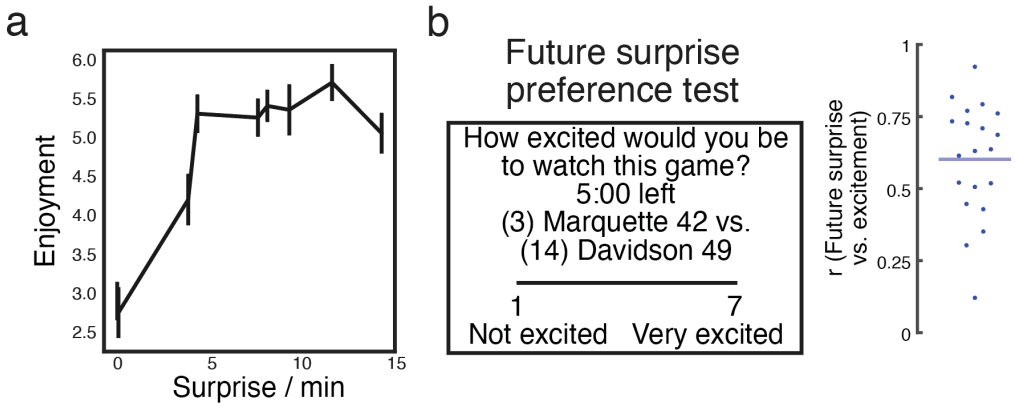

Fig S4: Surprise correlates with enjoyment and neural activity in reward-related regions. a) Mean surprise positively predicted game-level enjoyment. b) A survey given at the end of the experiment (left) showed that subjects' excitement for watching a new set of games strongly correlated with the amount of likely future surprise in those games (as determined by finding similar games in corpus data).

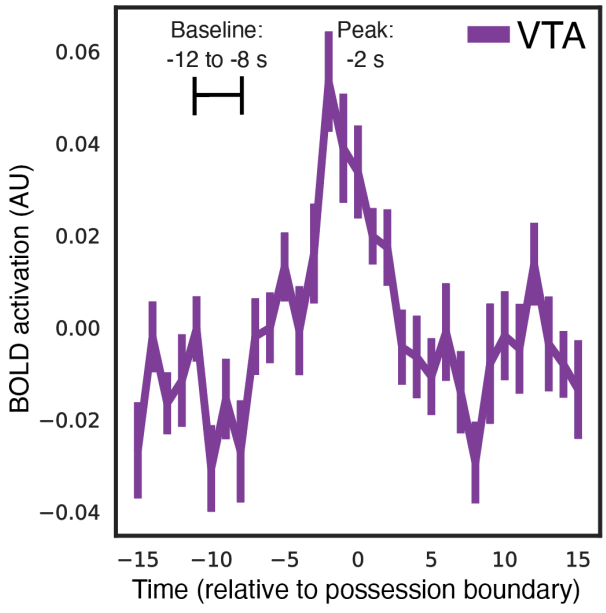

Fig S5. BOLD VTA activation relative to possession boundaries for time-shifted (5 s) data. We used peak activity at -2 s (subtracting a -12 to -8 s) baseline to predict memory on a trial-by-trial basis.

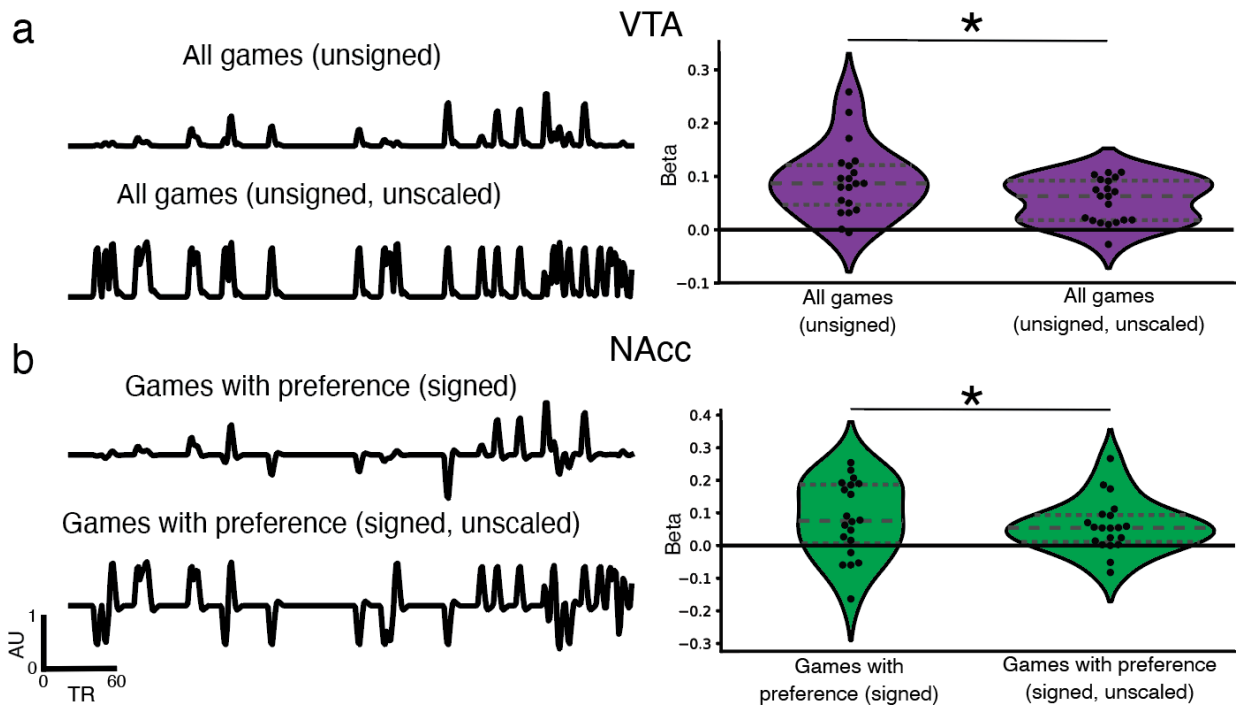

Fig S6. Unsigned VTA and signed NAcc responses fit scaled better than unscaled models. a) (left) Depictions for one game of scaled and unscaled models of unsigned surprise included in GLMs of VTA. (right) Betas were higher for scaled than unscaled models in VTA (p = 0.002). b) (left) Similar to a), except for signed surprise as scaled and unscaled. (right) Betas were higher for scaled than unscaled models in NAcc (p = 0.009).

	States / min	Surprise			BeliefIncons			BeliefIncons - BeliefCons			Subjective event boundaries		
Parcel		*r*	*p*	*q*	*r*	*p*	*q*	*r* (diff)	*p*	*q*	*r*	*p*	*q*
Frontal Pole	0.27	0.05	0.424	0.767	0.05	0.438	0.528	0.04	0.742	0.775	-0.01	0.622	0.664
Insular Cortex	0.56	0.06	0.619	0.808	0.13	0.247	0.374	0.18	0.164	0.297	0.01	0.674	0.704
Superior Frontal Gyrus	1.17	0.1	0.133	0.767	0.15	0.11	0.243	0.16	0.319	0.454	0.09	0.155	0.251
Middle Frontal Gyrus	0.8	0.09	0.287	0.767	0.21	0.048	0.194	0.28	**0.04**	0.136	0.15	0.053	0.147
Inferior Frontal Gyrus, pars triangularis	0.94	0.12	0.121	0.767	0.14	0.126	0.243	0.12	0.447	0.584	0.12	0.132	0.239
Inferior Frontal Gyrus, pars opercularis	1.07	0.03	0.56	0.798	0.1	0.361	0.456	0.14	0.488	0.596	0.13	0.105	0.219
Precentral Gyrus	1.15	0.14	0.132	0.767	0.28	**0.003**	0.08	0.36	**0.01**	0.092	0.16	**0.043**	0.147
Temporal Pole	0.48	-0.01	0.954	0.954	0.07	0.572	0.641	0.13	0.369	0.51	0.11	0.18	0.274
Superior Temporal Gyrus, anterior	1.3	0.09	0.337	0.767	0.13	0.272	0.388	0.13	0.478	0.596	0.27	**0.005**	0.05
Superior Temporal Gyrus, posterior	1.12	-0.05	0.603	0.808	0.05	0.756	0.772	0.15	0.38	0.511	0.17	**0.051**	0.147
Middle Temporal Gyrus, anterior	0.88	0.06	0.516	0.798	0.09	0.463	0.531	0.1	0.634	0.704	0.06	0.461	0.541
Middle Temporal Gyrus, posterior	0.8	0.11	0.218	0.767	0.24	**0.02**	0.105	0.32	**0.018**	0.108	0.11	0.181	0.274
Middle Temporal Gyrus, temporooccipital	1.1	0.07	0.382	0.767	0.16	0.132	0.243	0.22	0.204	0.343	0.24	**0.003**	0.05
Inferior Temporal Gyrus, anterior	0.91	-0.05	0.808	0.883	0.03	0.864	0.864	0.11	0.644	0.704	0.17	**0.049**	0.147
Inferior Temporal Gyrus, posterior	0.29	0.1	0.492	0.798	0.22	0.063	0.194	0.29	**0.018**	0.108	-0.02	0.853	0.871
Inferior Temporal Gyrus, temporooccipital	0.95	0.15	0.131	0.767	0.25	**0.031**	0.148	0.29	**0.038**	0.136	0.21	**0.011**	0.065
Postcentral Gyrus	0.74	0.04	0.56	0.798	0.14	0.15	0.247	0.22	0.137	0.292	0.02	0.545	0.622
Superior Parietal Lobule	1.31	0.07	0.344	0.767	0.24	**0.018**	0.105	0.35	**0.022**	0.114	0.14	0.099	0.219
Supramarginal Gyrus, anterior	0.99	0	0.827	0.883	0.14	0.131	0.243	0.26	0.073	0.214	0.16	0.067	0.167
Supramarginal Gyrus, posterior	1.32	0.05	0.422	0.767	0.11	0.352	0.456	0.13	0.515	0.605	0.13	0.093	0.218
Angular Gyrus	1.2	0.06	0.365	0.767	0.16	0.066	0.194	0.23	0.106	0.237	0.21	**0.007**	**0.05**
Lateral Occipital Cortex, superior	1.29	0.08	0.497	0.798	0.2	0.066	0.194	0.28	**0.034**	0.134	0.19	**0.02**	0.093
Lateral Occipital Cortex, inferior	1.21	-0.03	0.827	0.883	0.05	0.681	0.727	0.13	0.532	0.61	0.12	0.14	0.243
Intracalcarine Cortex	2.72	0.06	0.41	0.767	0.23	**0.013**	0.105	0.36	**0.034**	0.134	0.23	**0.007**	0.05
Frontal Medial Cortex	0.4	0.2	0.137	0.767	0.24	0.054	0.194	0.22	0.091	0.224	0.14	0.119	0.232
Juxtapositional Lobule Cortex	0.84	0.02	0.785	0.883	0.14	0.153	0.247	0.24	0.069	0.214	0.06	0.304	0.388
Subcallosal Cortex	0.17	0.07	0.302	0.767	0.16	0.077	0.202	0.22	0.087	0.224	-0.06	0.556	0.622
Paracingulate Gyrus	0.81	0.04	0.553	0.798	0.07	0.463	0.531	0.08	0.659	0.704	0.06	0.309	0.388
Cingulate Gyrus, anterior	0.58	0.02	0.683	0.867	0.03	0.715	0.747	0.03	0.887	0.887	0	0.89	0.89
Cingulate Gyrus, posterior	1.04	0.06	0.445	0.775	0.18	0.115	0.243	0.26	0.088	0.224	0.19	**0.024**	0.103
Precuneous Cortex	1.21	0.13	0.214	0.767	0.26	**0.016**	0.105	0.34	**0.014**	0.108	0.28	**<0.001**	**<0.001**
Cuneal Cortex	1.68	0.26	**0.008**	0.357	0.3	**0.006**	0.08	0.26	0.164	0.297	0.27	**0.004**	0.05
Frontal Orbital Cortex	0.44	0.13	0.284	0.767	0.18	0.14	0.243	0.18	0.156	0.297	0.08	0.314	0.388
Parahippocampal Gyrus, anterior	0.16	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
Parahippocampal Gyrus, posterior	0.53	0.18	0.241	0.77	0.26	0.063	0.19	0.28	**0.027**	0.13	0.03	0.608	0.664
Lingual Gyrus	2.37	0.1	0.282	0.767	0.29	**0.005**	0.08	0.42	**0.004**	0.089	0.25	**0.007**	0.05
Temporal Fusiform Cortex, anterior	0.23	-0.17	0.16	0.767	-0.06	0.602	0.658	0.07	0.495	0.596	0.1	0.274	0.379
Temporal Fusiform Cortex, posterior	0.49	0.19	0.135	0.767	0.22	0.075	0.202	0.2	0.143	0.292	0.1	0.312	0.388
Temporal Occipital Fusiform Cortex	1.12	-0.05	0.768	0.883	0.15	0.161	0.253	0.34	**0.01**	0.092	0.13	0.107	0.219
Occipital Fusiform Gyrus	1.54	0.02	0.762	0.883	0.24	**0.019**	0.105	0.41	**0.002**	0.089	0.23	**0.019**	0.093
Frontal Operculum Cortex	0.66	0.01	0.852	0.89	0.09	0.369	0.456	0.15	0.318	0.454	0.07	0.375	0.452
Central Opercular Cortex	0.93	0.12	0.175	0.767	0.1	0.33	0.456	0.06	0.826	0.844	0.11	0.123	0.232
Parietal Operculum Cortex	1.26	0.03	0.603	0.808	0.16	0.123	0.243	0.27	0.106	0.237	0.11	0.213	0.306
Planum Polare	1.02	0.09	0.258	0.767	0.14	0.126	0.243	0.16	0.265	0.415	0.15	0.056	0.147
Heschl's Gyrus (includes H1 and H2)	1	0.01	0.808	0.883	0.1	0.351	0.456	0.17	0.288	0.437	0.11	0.215	0.306
Planum Temporale	0.99	0.1	0.329	0.767	0.14	0.136	0.243	0.15	0.237	0.384	0.16	**0.035**	0.139
Supracalcarine Cortex	2	0.08	0.378	0.767	0.23	**0.007**	0.08	0.34	**0.01**	0.092	0.17	**0.048**	0.147
Occipital Pole	2.23	0	0.951	0.954	0.12	0.272	0.388	0.23	0.189	0.329	0.16	0.151	0.251

Table S1. Number of HMM states per minute and correlations between HMM transitions at boundaries with various measures of surprise and subjective event boundaries. Uncorrected *p* values and FDR-corrected *q* values are provided for each correlation based on permutation tests that shuffled possessions within each game. The anterior parahippocampal gyrus did not transition enough between states to warrant correlations. BeliefCons = belief-consistent surprise, BeliefIncons = belief-inconsistent surprise. These data are plotted in Fig S3.

	Solo model				Final model
Factor	**Estimate**	**SEM**	***z* value**	**Pr(>|*z*|)**	**Pr(>|*z*|)**
Surprise_pre	0.046	0.009	4.7	< 0.001	ns, 0.85
PAC_pre	0.31	0.13	2.4	0.02	0.02
PrevMem	1.6	0.19	8.6	< 0.001	< 0.001
vmPFC_HMM_pre	3.65	3.2	1.1	0.26	NA
VTA_pre	0.09	0.13	0.7	0.49	NA
Odd	1.26	0.4	3.0	0.003	0.003
Expert	0.85	0.22	3.8	< 0.001	0.01
Surprise_post	0.06	0.008	7.2	< 0.001	< 0.001
PAC_post	0.13	0.12	1.0	0.29	NA
vmPFC_HMM_post	3.25	3.2	1.0	0.31	NA
VTA_post	0.29	0.12	2.4	0.02	ns, 0.09
BeliefCons	0.03	0.004	7.5	< 0.001	ns, 0.06
BeliefIncons	0.003	0.006	0.4	0.67	NA

Table S2. Factors predicting memory in solo models and a full model. Note: PAC = pupil area change, PrevMem = previous possession recalled, Odd = oddness of possession, Expert = subject expertise, BeliefCons = belief-consistent surprise, BeliefIncons = belief-inconsistent surprise, VTA = ventral tegmental area activation, vmPFC_HMM = HMM transition agreement in vmPFC. Suffixes: _pre = leading into possession, _post = end of possession. All significant solo variables were entered into a full model and were dropped in the order of the highest value until the final model was reached. ns = not significant in final model.