Long-term, multi-event surprise enhances autobiographical memory

James W. Antony[1,*], Jacob Van Dam[1], Alexander J. Barnett[2], Kelly A. Bennion[1]
[1]Department of Psychology & Child Development,
California Polytechnic State University, San Luis Obispo
[2]Department of Psychology, University of Toronto
*corresponding author: jwantony@calpoly.edu

December 10, 2022

Abstract:

Neurobiological models of learning emphasize the importance of prediction errors (surprises) for memory formation. These effects have focused on memory for information surrounding a momentary surprising event; however, it is less clear whether surprise that unfolds across multiple events and timescales impacts memory. We asked basketball fans about their most positive and negative autobiographical memories of individual plays, games, and seasons, allowing surprise measurements spanning seconds, hours, and months. We used advanced analytics on National Basketball Association play-by-play data and betting odds spanning 17 seasons, >22K games, >5.6M plays to compute and align the estimated surprise values of each memory. We found that surprising events biased positive memories on the scale of seconds and months and negative memories across all three timescales. Critically, game and season memories could not be explained by surprise at shorter timescales, suggesting that long-term, multi-event surprise influences memory. These results expand the role of surprise in models of learning and reinforce its relevance in real-world domains.

Introduction

Predictions about upcoming events – and their eventual outcomes – have a substantial effect on learning and memory[1–6]. A wealth of research has demonstrated that momentary prediction errors, or surprises, are particularly well-remembered[1,2,7–12] and that an array of physiological and mental processes seem to underpin these memory benefits. For instance, surprise engages the dopaminergic and serotonergic midbrain systems[8,10,13–17], alters and enhances hippocampal activity[16,18–24], increases pupil dilation[8,25–29], and enhances perceptual[13,16,30,31] and attentional processing[7,32–39] to the surprising stimulus. Surprises can be further differentiated into events that are better than expected for the agent because they signal reward (+ reward prediction error, or + signed surprise) and others that are worse than expected (- signed surprise)[40–45]. The physiological response to these events is critical for learning and updating values of stimuli in the service of the optimal behavior to approach rewards and avoid punishments[41]. With regard to memory, outcomes with + signed surprise are better remembered in some paradigms[3,40,46,47], whereas in others, the magnitude, rather than the direction of surprise (or unsigned surprise) drives memory[1,2,8,11,40].

The prevailing focus in these studies is how surprise affects learning and memory for momentary events on the scale of seconds; however, it is clear that humans can predict outcomes well beyond the upcoming moment[48]. A very non-exhaustive list of things we can make predictions about beyond the next event include a series of upcoming stimuli[49,50], the linguistic content of not just upcoming words[51–53] but paragraphs[54,55], attributes about oneself months and years into the future[56–59], political elections months and years into the future[60], and the winners of upcoming sports games[61,62] and championships[63]. Given that

The authors declare no competing interests.

we commonly make long-term predictions and that surprise drives memory, it is worth asking whether long-term predictions resulting in errors (long-term surprises) also drive memory.

The problem of long-term predictions is intriguing in domains like politics or sports that feature probabilistic updates about a single outcome over long stretches of time. As humans can represent information at multiple timescales of granularity, it is unclear which timescale should be operative at any given time, whether errors accumulate over time, or whether each of these variables impact memory. Consider the following example: Imagine you predicted in late January 2020 that Bernie Sanders would become the Democratic nominee for US president, which was ∼20% likely per 538. In this hypothetical scenario, imagine that throughout February and March 2020, Sanders won state after state, and therefore Sanders' likelihood inched upwards every few days until it reached 98% by April 1. Had Sanders won one more state on April 7 and secured the nomination (100%), would it be surprising? The answer depends on the timescale. If judging from April 1, no: with only a 2% surprise, the outcome was essentially a foregone conclusion. However, if judging from late January, at 80% surprise, the answer is undoubtedly yes. Importantly, these kinds of events are both long-term (the interval between an initial prediction and final outcome spans more than a single moment) and composed of multiple sub-events (there are numerous updates to predictions before that outcome).

Here our primary focus was to determine whether surprise occurring across multiple timescales and events affects memory. To do so, we framed our questions around events spanning multiple moments – where the events could still be surprising as a collective even if they were not individually. For this, we asked basketball fans about their most positive and negative memories of entire basketball games and seasons. To link these results with the literature on momentary surprise, we also asked fans about individual plays, allowing us to replicate findings in this literature and extend them to real-world memories [see also[64,65]]. We chose the sports domain for three main reasons. First, recent explosions in publicly available sports data allowed us to precisely quantify predictions about the likelihoods that a given team would win a game or championship. Using such predictions, we can form estimates of surprise following the outcome of a play, game, or season. Second, sports contain a natural hierarchy of timescales. In the National Basketball Association (NBA), which is the focus here, the broadest scale over which teams compete is the season, and seasons consist of collections of games, which consist of collections of plays. Therefore, we could find and situate the context of subject memories within a multi-scale predictive framework with respect to game and championship likelihoods. Third, affiliating with sports teams often provides fans a strong sense of identity[66], and as a result, sports memories can be highly arousing and vivid[67,68]. Emotional events are recollected more vividly[69–71], more often[72–74], and for longer intervals than neutral events[72,75], and they also often serve as the basis for experiments surrounding highly vivid, "flashbulb" memories of public events[76–79] . We therefore anticipated that subjects would be able to readily access these memories. We predicted that surprise would play an outsize role in memory across all three timescales and, critically, that the contributions of surprise at the longer timescales could not be attributed to surprise at shorter scales.

Results

Characterizing subject responses

Subjects (N=122, 34 female) took a survey asking them about their most positive and negative memories of individual plays, games, and seasons as basketball fans (Figure 1A). We asked them to include as much detail as possible, especially emphasizing that we would like to specifically find each memory via later internet search. Our goal was to precisely identify the play, game, or season from which subject memories occurred so we could quantify their surprise-related attributes. We intentionally left the question open-ended and avoided implying that they should use surprise as a heuristic.

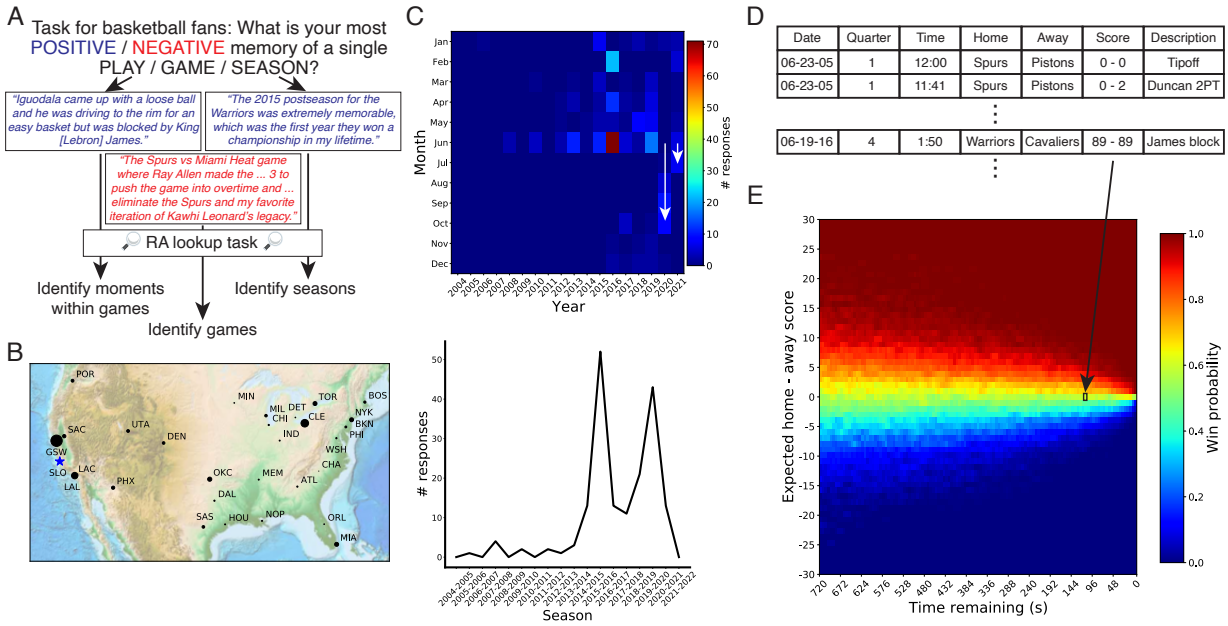

Figure 1: Memory task, basic response characteristics, and analytical approach. (A) Subjects were asked to freely respond to prompts about their most positive and negative memories of plays, games, and seasons as basketball fans. We then linked these responses to specific teams and times across multiple time scales (moments in games, games within seasons, seasons). (B-C) We characterized the response frequencies by the primary NBA team's geographical location for plays, games, and seasons (B) as well as their temporal distributions by month and year of play and game responses (C, top) and starting year of season responses (C, bottom). Due to the Covid-19 pandemic, the end of the NBA season was shifted four months in 2020 and one month in 2021 (white arrows in C). (D-E) We imported play-by-play data for all games across 17 seasons (2004-05 to 2020-2021) (D) to find the context for all subject responses and to train a win probability model (E). The win probability model computed the likelihood of winning the game given each game context, including the strength of the two teams and the current score (which were combined into an adjusted score), the time remaining in the game, and the team in possession of the ball (only 'Home' shown in E). For instance, the play from (A) in which Lebron James blocked a shot corresponds to the win probability in the black square in (F).

Twenty-nine of the thirty NBA teams served as the primary team of interest in subject responses (Figure 1B; Figure S1). We conducted the study in a location that was >320 km (200 miles) from the nearest NBA team stadium in order to reduce the bias towards a single fanbase, though there remained a bias towards the nearest teams (responses across all types, the San Francisco-area-based Golden State Warriors: 261; Los Angeles Lakers: 83) (Figure 1B). Memories for plays and games were biased towards recent years (after 2013) and towards the end of the postseason, which typically occurred in June but was moved to October in 2020 and July in 2021 for Covid-19 pandemic rescheduling (Figure 1C, top). Memories for seasons were similarly biased towards recent years (Fig 1C, bottom). As data were collected from February until November 2021, we also plotted the age of the memories between the study and the date of their reported memories, which showed a similar bias (Figure S2).

Momentary surprise drove memory for positive and negative plays

We began our investigations by examining surprise for positive and negative plays, which resembles the momentary surprise studied in many human memory paradigms [e.g.,[11]] and in traditional reinforcement learning paradigms [e.g.,[43]]. Our main analytical approach for play and game memories relied on scraping the NBA API for play-by-play data (Figure 1D). These data included >5.6M plays from >22K games between 2004-2021. Our derivation of momentary surprise was based on the primary outcome variable of relevance to sports fans: which team wins a given game. One can conceptualize watching a basketball

game as navigating a state space of predictions about the eventual winner that becomes updated with each change in game context (Figure 1E; Figure S3). Four factors influenced win probability in our model: the score difference between the two teams (oriented as positive for the home team), the relative strength of the teams, the amount of game time remaining, and the team with possession of the ball. Context changes included scores, timeouts, and three types of plays that potentially change possession (turnovers, rebounds, and tip-offs). We then computed surprise as the derivative in the win probability time course across these changes in game context (Figure 2A). This construct was later split into an unsigned value using the absolute value of this derivative and a signed value that was oriented positively for each subject's preferred team and negatively for their non-preferred team. Importantly, we validated that the surprise metrics from all positive and negative plays based on our algorithm corresponded tightly with those from an expert website (Figure 2B).

Before characterizing surprise, we investigated when these plays occurred within games and within seasons. Both positive and negative plays tended to come from the very end of the 4th quarter of games and into overtime periods, and positive plays tended to have less time remaining in the game than negative plays [positive: median = 5 s, N = 70; negative: median = 53 s, N = 74; Mann-Whitney U = 1904.5, p = 0.006] (Figure 2C). Interestingly, examining the mean and standard deviation of surprise across all games revealed that mean surprise went down modestly throughout the average game, accounting for the large number of games in which the outcome is all but certain by the end ("blowouts") (Figure S4A). However, standard deviation of surprise increased exponentially towards the end of the game, suggesting the final moments offer the largest potential for surprise (Figure S4B). We also calculated the percentile of the game from which the play occurred relative to the chosen team's season. For example, the first game of the season was in the 0th percentile and the last game was in the 100th percentile. Similar to how responses tended to come from the final months (Figure 1C), these responses tended to come from the very end of the primary team's season. In this case, negative plays tended to come closer to the end of a team's season than positive plays [median percentile, positive: 89.3th; negative: 99.0th; Mann-Whitney U = 1922, p = 0.006] (Figure 2D).

Based on prior results showing that momentary surprise influences memory[1,2,8,11,40], we predicted that memories for positive and negative plays would preferentially be selected from a subset of the most surprising plays. We first tested this by examining whether unsigned surprises from the chosen play memories were higher than a null distribution of all plays from all games. It should be noted that the null distribution was non-normal – most plays had little relevance on the eventual outcome: 32.4% and 78.6% of plays had $<$0.01 and $<$0.05 surprise, respectively, whereas only 0.1% had $>$0.25 surprise. Subject play memories were significantly greater than this null distribution overall [median, positive: 0.31; negative: 0.11, null: 0.019; positive vs. null Mann-Whitney $U = 3.4*10^8$, $p < 0.001$; negative vs. null Mann-Whitney $U = 3.2*10^8$, $p < 0.001$; positive vs. negative Mann-Whitney U = 3372, p = 0.002], with 78.5% and 51.4% of positive plays and 71.6% and 17.6% of negative plays having $>$0.05 and $>$0.25 surprise (Figure 2E, top). We next reasoned that sports fans have limited time and likely do not watch every game or part of every game. Instead, in order to maximize their experience of surprise[80], which acts as a reward in itself in low-stakes contexts like sports games[8,62,81–85], they may prioritize their fandom by watching select games or parts of games. If this were the case, a null distribution of plays that is more representative of the plays actually experienced by subjects would be those from the games the subjects reported, or even just plays from the 4th (final) quarter of those games. Positive and negative play surprise was higher than the null distribution from only the games that subjects reported (null median: 0.025; Mann-Whitney $U > 1*10^6$, $p < 0.001$, both) and also higher than plays from the 4th quarters (null median: 0.036; Mann-Whitney $U > 2.5*10^5$, $p < 0.001$, both) (Figure 2E, bottom).

We next asked whether these memories represented particularly positive or negative signed surprises with respect to the subject's preferred team[8,43,86]. For our null distribution, we used a double distribution of the same plays, as if assuming a hypothetical fan might cheer for each team in each game. We again used

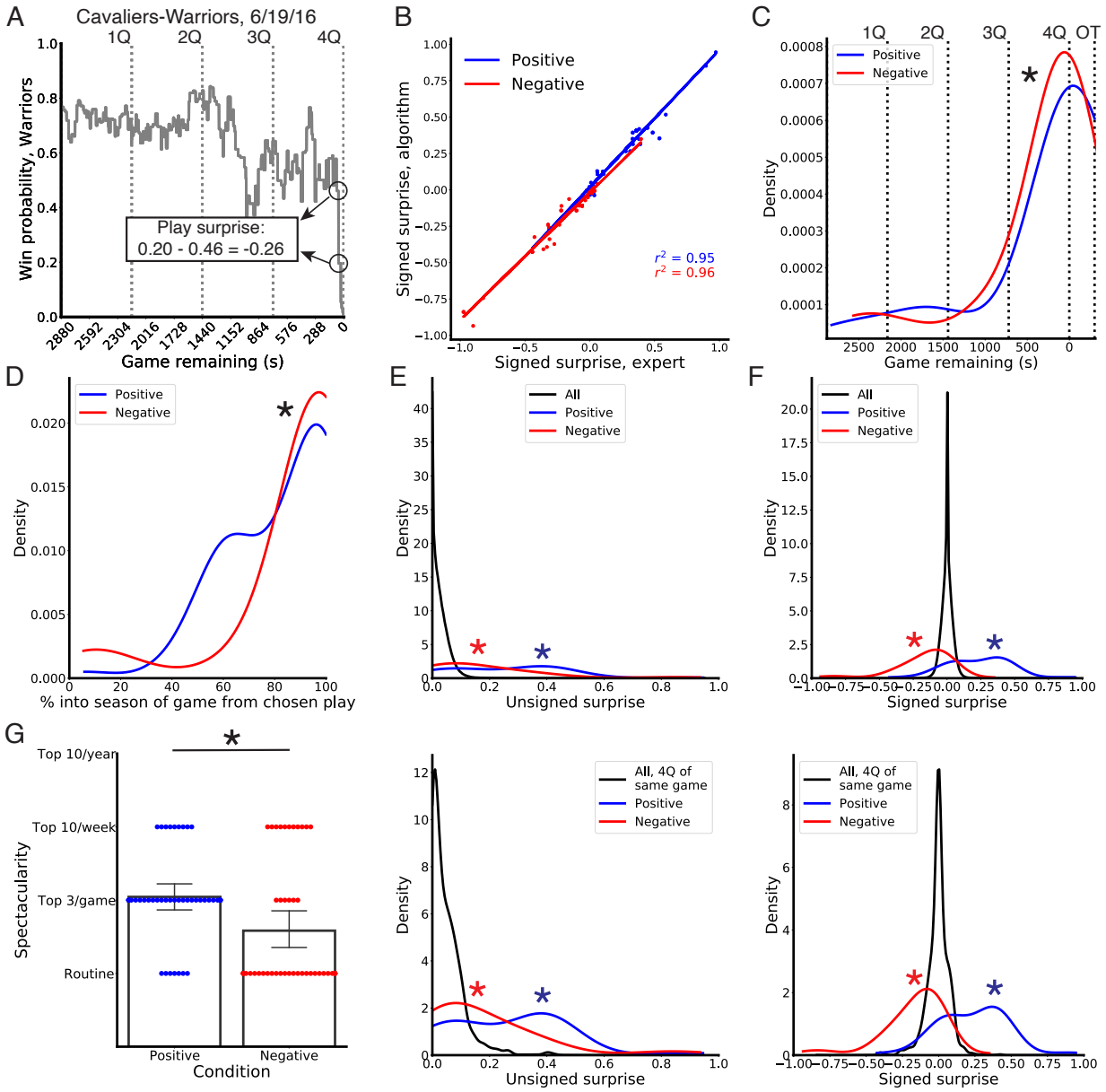

Figure 2: Surprise predicts memory for positive and negative plays. (A) We computed play surprise as the difference in win probabilities across successive moments, shown here using Game 7 from the 2016 NBA Finals between the Cleveland Cavaliers and Golden State Warriors. We computed an unsigned version of this metric using the absolute value and a signed version of this metric by considering this value with respect to the subject's preferred team (e.g., positive corresponds to a preferred outcome). (B) Surprise values from our algorithm correlated strongly with those from an expert. (C) Positive and negative responses were strongly biased towards the end of games. 1-4Q: 1st-4th quarter; OT: 1st overtime. (D) Positive and negative responses were also strongly biased towards the end of the corresponding team's season. (E) Positive and negative responses corresponded with moments of higher unsigned surprise than a null distribution of all basketball plays (top) and a more focused null distribution of the 4th quarter of the games chosen by subjects (bottom). (F) Positive and negative responses corresponded to higher and lower signed surprise, respectively, than null distributions of all basketball plays (top) and the 4th quarter of the games chosen by subjects (bottom). (G) The spectacularity of a given play (as assessed by raters), which assesses an alternative way in which a basketball play can be surprising that is unrelated to win probability, was higher for positive than negative play responses. *: $p < 0.01$

all plays from all games, all plays from the chosen games, and all plays from the 4th quarter of the chosen games. In all cases, positive memories were more positive and negative memories were more negative than the null distributions (median, positive: 0.29; negative: -0.11, null in all cases: 0.0; Mann-Whitney $U > 6.5*10^5$, $p < 0.001$) (Figure 2F). Therefore, subjects tended to report plays with exceptionally good or bad win probability swings for their preferred teams. In sum, the influence of unsigned and signed surprises on memory for plays replicates prior findings on the influence of momentary surprise on memory[8,40] and extends them into real-world domains[64,65].

Finally, we also considered that our operationalization of surprise as win probability change is only one dimension along which basketball can be surprising[8]. Rather, one might consider a surprising play in terms of the way it unfolds against a lifetime of viewing other plays, such as a spectacular or improbable shot. Since viewing feats of athleticism may be one of the primary reasons one watches sports, we hypothesized that such plays would be more likely to be categorized as positive, though we acknowledge the alternative possibility that an improbable shot from an opposing team may also "sting" more deeply. We operationalized this type of surprise as the 'spectacularity' of a given play, which we separately rated without regard for the context of the game. We rated these plays as (1) a routine play like a layup or baseline jump shot, (2) a tough play like a long shot or a fadeaway jumpshot that might be considered in the top 3 in a given game, (3) a particularly athletic play that might be featured on a "Top 10 plays of the day" news segment across all sports, such as a half-court shot or behind-the-basket, changing-hands-in-mid-air layup, or (4) a truly sensational play that might be considered on a year-end "Top 10 plays of the year" segment like a full-court shot or a famous dunk by Michael Jordan off of a missed free throw attempt. We found that positive plays had higher spectacularity than negative plays [positive: median = 2, mean = 2.0, N = 45; negative: median = 1, mean = 1.6, N = 48; Mann-Whitney U = 1477, p = 0.001], indicating that spectacularity was another dimension influencing subject memories and that it was higher for positive than negative memories (Figure 2G).

Long-term surprise influenced negative memories for games

Next, we aimed to determine whether surprise on a longer timescale – lasting beyond a single moment – would also influence memory. For this analysis, we considered two long-term surprise metrics calculated over the course of games. The first was full-game surprise, or the difference between the pre-game win probability and the final result. The second metric was within-game surprise, or the difference between the lowest probability for the eventual winner and the final result (Figure 3A). This metric captured the familiar concept of a "comeback" that looks at how maximally wrong one's predictions might have been during the course of a game. We validated that both of these metrics corresponded strongly with those from an expert sports analyst for both positive and negative memories of games (Figure 3B).

Next, similar to our characterizations for play memories, we showed that both positive and negative game memories were biased towards the end of a team's season; similar to the temporal distributions of play memories, negative games were more likely to be recalled closer to the end of a team's season [median percentile, positive: 98.8th, based on N = 77; negative: 100th, based on N = 74; Mann-Whitney U = 2225, p = 0.01] (Figure 3C). We also found the time that the winning team had its lowest win probability. We compared these times to those from a null distribution of all games, which tended to peak early with a skewed distribution towards the end of the games (Figure 3D), again reflecting how many game outcomes were nearly certain from an early time point. For both positive and negative game memories, these times occurred later in the game than in the null distribution [median game remaining, positive: 1364 s, based on N = 77; negative: 1364 s, based on N = 74; null: 2081, based on N = 22539; positive vs. null Mann-Whitney $U = 9.9*10^5$, $p < 0.001$; negative vs. null Mann-Whitney $U = 1.1*10^6$, $p < 0.001$], showing that the preferred team had a better chance of winning than is typical at some point into the game. However, these low win probability time points came later in the game for negative than positive games [positive vs.

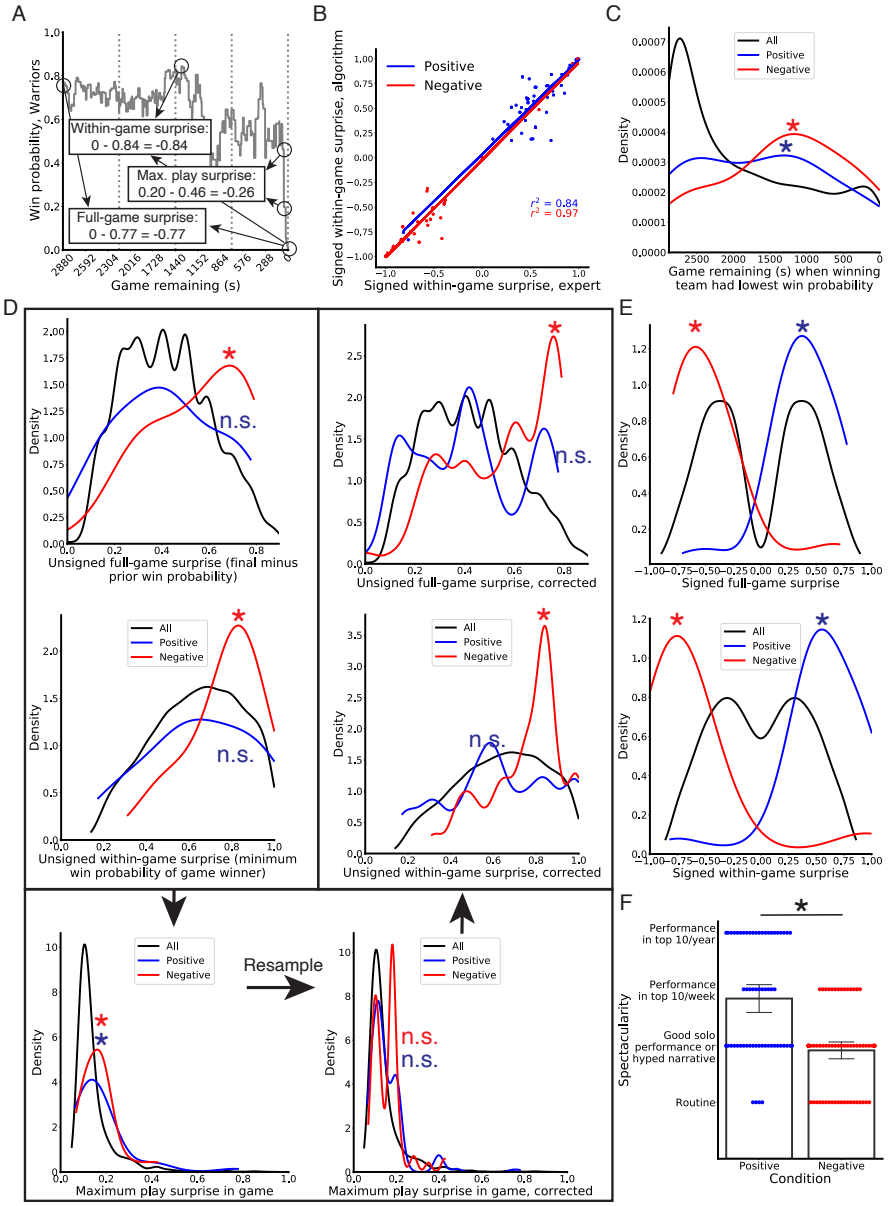

Figure 3: Full- and within-game surprises predicted memory for positive and negative games. (A) We computed full-game surprise as the difference between the pre-game and final win probability and within-game surprise ("comeback") as the difference between the lowest and final win probability for the game winner. Additionally, we found the maximum play surprise for each game. (B) Signed within-game surprise values correlated strongly with those from an expert site (see Figure S5 for full-game surprise). (C) The game time from when the winning team had their lowest win probability, marking the onset of a comeback, was later in the game for negative, but not positive games. These times were not heavily distributed towards the end of the game, which suggests there is more than a single play that drove the comeback and therefore it occurred over a longer timescale. (D) Absolute unsigned full-game surprise (top left) and within-game (middle left) were significantly greater than the null distribution (all games) for negative, but not positive, game responses. Maximum unsigned play surprise within a game was greater for the positive and negative games chosen by subjects than a null distribution of all games (bottom left). We next selected subsets of the positive and negative games that did not differ from the null null distribution (bottom right). Using these modified distributions, negative games still had greater unsigned full-game (top right) and within-game surprise (middle right) values than the null distribution. (E) Signed full-game surprise (top) and within-game (bottom) values were significantly higher and lower for positive and negative games, respectively, than the null distribution using all games. (F) The spectacularity of a given game (as assessed by raters) was higher for positive than negative game responses. *: $p < 0.02$

negative Mann-Whitney $U = 3331$, $p < 0.001$]. Notably, these times for negative games were not heavily clustered around the final moments of the game, which suggests that negative game memories were not solely driven by games that were lost in the final seconds. Rather, the events in the game leading to the loss were spread across many accumulated – albeit small – surprises over time.

We again asked whether game memories were influenced by unsigned surprise – in this case, surprise aggregated over multiple events spanning a timescale of minutes to hours. Similar to our play analyses, we compared full-game and within-game unsigned surprise values for positive and negative game memories against a null distribution of these values in all games. The null distribution for full-game surprise was uneven, as teams tended to fall into clusters of relative strengths, producing local peaks (Figure 3D, top left). Nevertheless, the distribution was biased towards lesser surprise, as one should expect if pre-game probabilities hold predictive power (median = 0.41). Intriguingly, full-game surprises from negative game memories were significantly greater than those from the null distribution and the positive game memory distribution, whereas positive memories were not significantly different from the null [median full-game surprise, positive: 0.41; negative: 0.60; positive vs. null Mann-Whitney $U = 8.7*10^5$, $p = 0.998$; negative vs. null Mann-Whitney $U = 5.5*10^5$, $p < 0.001$; positive vs. negative Mann-Whitney $U = 1980$, $p = 0.001$]. These results suggest that negative game memories preferentially came from instances in which the subject's preferred team was expected to win. Next, we compared within-game surprise for all games against positively and negatively remembered games. In the null distribution, comeback sizes were above 0.50 (mean = 0.64, median = 0.65) (Figure 3E, middle left). To explain this, consider that scoring in a basketball game is a stochastic process, and simulating a basketball game between two equally strong teams resembles examining whether heads occur more than tails at a coin flip game over 100 flips: The eventual winner between heads and tails is likely to be losing at some point during the coin flip game, leading to > 0.50 within-game surprise on average. Positively remembered games were no different than the null distribution of all games, whereas negatively remembered games had significantly higher within-game surprise than the null distribution and the positive memory distribution [median within-game surprise, positive: 0.62; negative: 0.84; positive vs. null Mann-Whitney $U = 8.3*10^5$, $p = 0.556$; negative vs. null Mann-Whitney $U = 5.4*10^5$, $p < 0.001$; positive vs. negative Mann-Whitney $U = 2055$, $p = 0.003$]. These results suggest that, similar to results for full-game surprise, the preferred team was winning by a larger-than-average value at some point during negatively remembered games (i.e., they blew the lead).

We next examined an alternative hypothesis: since games are composed of plays, negative game memories could simply be driven by single plays rather than the more long-term surprise of multiple accumulated events that are by themselves not particularly surprising. The distribution of times from which the lowest win probability occurred was not clustered near the end of the game (when the largest surprises tend to occur) (Figure 3C), which argues against the case for single plays driving game memories. Nevertheless, to further address this concern, we found the most surprising play in the positive and negative games (Figure 3A), and we compared this value to a distribution of maximum play surprise values for all games. As the distribution of maximum surprise values for each game rather than all values, this null distribution is higher than the play surprise distribution (median: 0.118). Importantly, both positive and negative maximum play surprise distributions were higher than the null distribution and did not differ [median maximum play surprise, positive: 0.134; negative: 0.179; positive vs. null Mann-Whitney $U = 6.9*10^5$, $p = 0.002$; negative vs. null Mann-Whitney $U = 5.8*10^5$, $p < 0.001$; positive vs. negative Mann-Whitney $U = 2660$, $p = 0.48$] (Figure 3E, bottom left), potentially leaving this as an alternative account for the full- and within-game surprise results. To rule out this account, we first resampled our data with replacement until we found 100 instances in which each maximum play surprise distribution did not differ from the null (Figure 3E, bottom right) ($p > 0.10$). We then tested whether these subsets of games still showed differential levels of full-game and within-game unsigned surprise. Negative game memories within this subset had significantly higher full- (Figure 3E, top right) and within-game unsigned surprise (Figure 3E, middle right) than the null

distribution in 99% and 100% of the samples, respectively (see Figure S6 for distributions averaged across samples). Positive game memories did not differ from the null, resulting in significant differences in 10% and 7% of the samples. Therefore, while large momentary surprises may partially contribute to memories for events aggregated across numerous moments like games, they cannot fully explain them. This indicates that subjects preferentially remembered negative games with surprise aggregating over multiple events.

We next examined whether positive and negative game memories were drawn from games in which a fan's team overperformed or underperformed. We again created a null distribution by considering full- and within-game surprise values if a hypothetical fan cheered once for each team per game. Full-game signed surprise for positive and negative game memories was significantly higher and lower than this null distribution, respectively (Figure 3F, top) [median, positive: 0.41, negative: -0.54, null: 0.0; positive vs. null Mann-Whitney $U = 9.9*10^5$, $p < 0.001$; negative vs. null Mann-Whitney $U = 2.6*10^6$, $p < 0.001$; positive vs. null Mann-Whitney $U = 5245$, $p < 0.001$]. We found similar results for within-game surprise (Figure 3F, bottom) [median, positive: 0.55, negative: -0.74, null: 0.0; positive vs. null Mann-Whitney $U = 5.3*10^5$, $p < 0.001$; negative vs. null Mann-Whitney $U = 2.9*10^6$, $p < 0.001$; positive vs. null Mann-Whitney $U = 5188$, $p < 0.001$]. As positive game memories did not have higher unsigned full- and within-game surprise values than the null distribution, this analysis supports the idea that these games were good for their preferred team but not especially surprising overall. Conversely, negative game memories had higher unsigned surprise than expected, and these surprises also came with games in which the preferred team underperformed expectations.

Finally, similar to how one might consider a spectacular play to be surprising in a manner orthogonal to our win probability-derived metric, we also independently rated the spectacularity of games against a hypothetical corpus of other games. We operationalized game spectacularity as (1) a routine game where nothing especially notable occurred, (2) a good solo performance or especially strong extra-game narrative leading up to the game (e.g., the final game between two historic players), (3) a top-10 performance by a particular player in a given week, or (4) a top-10 performance by a player in a given year. One common positive game memory along these lines reported by 6 subjects was for January 23, 2015, when Golden State Warriors player, Klay Thompson, broke a record by scoring 37 points in a single quarter. We found that spectacularity for games, like for plays, was significantly higher for positive than negative memories [positive: median = 3, mean = 2.8, $N = 61$; negative: median = 2, mean = 1.9, $N = 74$; Mann-Whitney $U = 3365$, $p = 0.001$] (Figure 3G).

Long-term surprise influenced memories for positive and negative seasons

To assess whether memories could be influenced by prediction errors across far longer intervals, such as months, we assessed positive and negative subject memories of entire seasons. This analysis involved a different approach, using sports betting as a model of prediction rather than derivations from play-by-play data. Sports betting odds are released at multiple time points throughout a season, which bettors can use to place money on which team will win the championship. We used these odds to predict final season outcomes for all NBA teams and examined how positive and negative memories for seasons unfolded relative to these predictions (Figure 4A). To assess final season outcomes, we devised a novel system that ordinally ranked teams according to how long they lasted into the season and postseason, and we broke ties using their regular season records (see Methods for more details). Under this system, the championship winner was assigned #1 and the team with the worst record was assigned #30 (Figure 4B).

First, we asked whether final outcomes influenced memories apart from predictions. In this case, we used Wilcoxon signed-rank tests with the average final season ranking as the population mean (15.5) rather than comparing these results against the null distribution of a ranking system, which is a uniform distribution. We found that positive season memories had significantly better rankings for the subject's preferred team than the average season [median = 1, mean = 2.8, $N = 96$; Wilcoxon signed-rank $Z = 8.6$, $p = 0.001$], whereas

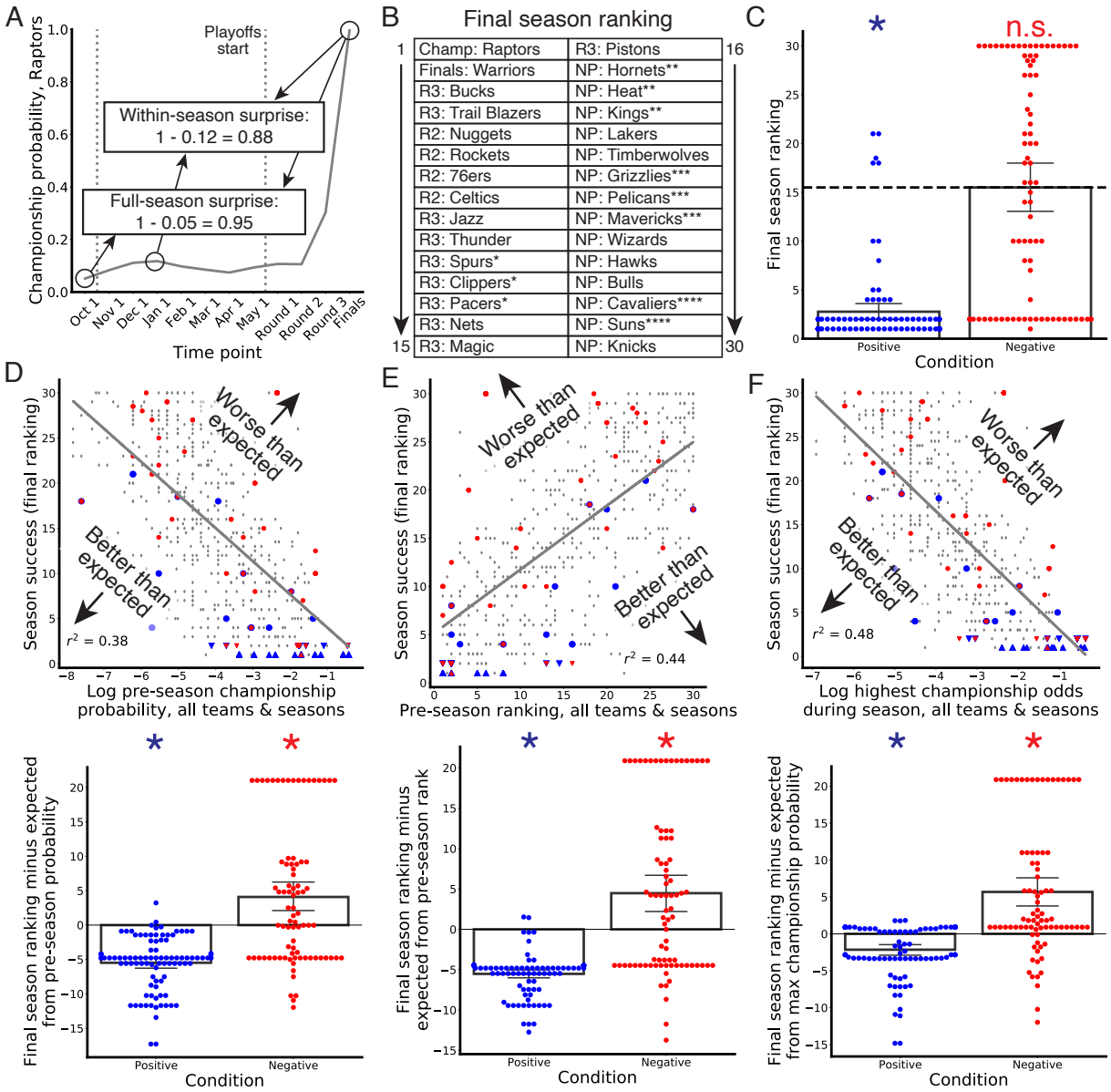

Figure 4: Surprise within and across full seasons predicted positive and negative season memories. (A-B) To calculate surprise across seasons, we first found Vegas championship betting odds updated monthly and before and after each playoff round, shown in (A) for the 2018-19 NBA champion Toronto Raptors. We then computed a novel metric which ranks teams by their success in the playoffs and regular season record. This metric was shown in (B) for all teams at the end of the 2018-19 season. (C) Positively recalled seasons were strongly influenced by this final ranking, as subjects chose seasons in which their preferred team was higher ranked than the average ranking (dotted line), whereas negative seasons did not differ from average. (D-F) To assess how season memories may have differed based on how outcomes differed from predictions, we devised linear regression models predicting final season rankings based on the following (top graphs): pre-season championship probability (D), pre-season ranking (E), and highest within-regular season championship probability (E). In each top plot, all seasons from all teams were shown as gray diamonds, positive season memories blue circles, and negative season memories red circles, with championship winners and runner-ups marked instead as upward and downward arrows, respectively. Regression lines were shown in gray. In each bottom plot, we plotted the residual between final outcomes and expectations based on each regression. Positive and negative season memories were better and worse than these expectations, respectively, for each residual metric. *: $p < 0.01$

negative seasons did not differ from the average [median = 16, mean = 15.5, N = 83; Wilcoxon signed-rank Z = 0.89, p = 0.38] (Figure 4C). To assess the role of prediction in season memories, we first measured the linear relationship for the final season outcomes from two variables constituting full-season surprise and one variable constituting within-season surprise. These variables were the logs of pre-season championship probabilities, pre-season rankings (which ranked teams ordinally according to pre-season probabilities), and the logs of the highest within-season championship probabilities (excluding the playoffs). All three metrics predicted final season outcome (Figure 4D-F, top) [r^2 in predicting final rankings, log pre-season odds: 0.38; pre-season rankings: 0.44; log highest within-season probability: 0.48, all $p < 0.001$]. Next, we used the residuals from these linear relationships to assess how positive and negative season memories measured up in comparison to what was predicted before and during the season. For all three metrics, positive season memories were better than predicted [log pre-season odds: median = -4.8, mean = -5.5, Wilcoxon signed-rank Z = 8.4, $p < 0.001$; pre-season ranking: median = -4.8, mean = -5.5, Wilcoxon signed-rank Z = 8.5, $p < 0.001$; log highest within-season odds: median = -1.7, mean = -2.1, Wilcoxon signed-rank Z = 4.9, $p < 0.001$] and negative memories were worse than predicted [log pre-season odds: median = 1.4, mean = 4.0, Wilcoxon signed-rank Z = 2.8, p = 0.005; pre-season ranking: median = 2.6, mean = 4.5, Wilcoxon signed-rank Z = 2.8, $p < 0.004$; log highest within-season odds: median = 1.8, mean = 5.7, Wilcoxon signed-rank Z = 5.1, $p < 0.001$] (Figure 4D-F, bottom).

We next examined whether our results were driven by subjects reporting won or lost championships for their favorite team, which may play a special role for sports fans beyond simply winning games[63]. Therefore, we re-ran these analyses excluding both finals teams in the championship game. In all cases, positive and negative seasons were better and worse than expected, respectively (all $p < 0.003$).

Finally, just as numerous plays compose a game, numerous plays and games compose a season. We therefore asked whether these data could be alternatively explained by especially surprising plays or games. For this analysis, we found the largest play surprise, full-game surprise, and within-game surprise values pooled across the entire season for each team in the corpus and for the teams represented by positive and negative season memories. Maximum play unsigned surprise across the season did not differ between the null distribution of all team seasons and those from positive and negative season memories [median, positive: 0.6, negative: 0.59, null: 0.58; positive vs. null Mann-Whitney $U = 3.5*10^4$, p = 0.72; negative vs. null Mann-Whitney $U = 2.2*10^4$, p = 0.33]. Maximum full-game unsigned surprise values were higher for both positive and negative season memories than the null distribution [median, positive: 0.87, negative: 0.87, null: 0.84; positive vs. null Mann-Whitney $U = 1.6*10^4$, $p < 0.001$; negative vs. null Mann-Whitney $U = 1.5*10^4$, $p < 0.001$], and maximum within-game unsigned surprise values were lower for both positive and negative season memories than the null distribution [median, positive: 0.98, negative: 0.96, null: 0.98; positive vs. null Mann-Whitney $U = 3.0*10^4$, $p < 0.001$; negative vs. null Mann-Whitney $U = 2.7*10^4$, $p < 0.001$]. This may have arisen because participants generally chose seasons from teams who performed at the extremes of team strength (Figure 4C), and therefore the potential for larger surprise values was higher. Nevertheless, after resampling 100 subsets of these distributions such that they no longer differed from the null distribution, we still found that positive and negative season memories were better and worse than predicted, respectively (percentage of samples in which p values were significant: positive, adjusting for full-game surprise = 100%; negative, adjusting for full-game surprise = 92%; positive, adjusting for within-game surprise = 100%; negative, adjusting for within-game surprise = 93%) (Figure S7). Therefore, surprise aggregated across a full season could not be explained by a single play or game.

Altogether, subjects reported positive and negative season memories – which involved numerous accumulated games and plays over the course of months – that substantially exceeded and fell below long-term predictions, respectively. Importantly, these findings cannot simply be explained by considering championship teams or surprises on a shorter timescale, such as especially surprising plays or games. These results were captured pointedly by one subject annotating their own memory of a relatively, though not objectively,

positive season for their favorite team: "The [Sacramento] Kings have been a bad team ever since I can remember, so my emotions correlating to my favorite team have more to do with my experience than with the performance of the team. I don't think the Kings have even made playoffs since 2002."

Discussion

We found that subjects' reported memories were biased towards events in which surprise unfolded on the scale of seconds, hours, and months. Given that fans enjoy good outcomes and dislike bad outcomes for their teams relative to expectations, as humans show pleasure or displeasure relative to expectations in a range of domains[87–90], one might expect that subjects would recall disproportionately high positive and negative *signed* surprises. However, reported memories also had disproportionately high *unsigned* surprises, suggesting that reported memories were not simply due to fan preference. Individually surprising moments (in this case, for individual plays) were well-remembered, which replicates prior results on momentary surprise[1–3,8,11,91] and extends them into real-world domains[64,65]. The latter aspect is important, because extending ideas from laboratory-based experiments in reinforcement learning (which consist of numerous repetitions of nearly identical trials) into real-world contexts (in which situations may not be precisely repeated) has been a central challenge[92]. More intriguingly, surprising games and seasons – which involved surprise accumulated over multiple sub-events and longer timescales – were also well-remembered. Critically, surprises at these longer scales could not be explained by particularly large surprises at shorter scales (plays in the case of games; plays and games in the case of seasons).

We predicted that subjects would preferentially remember games with high full-game or within-game surprise, which relate to the common terms of "upsets" and "comebacks", respectively. However, this prediction was only supported for negative game memories. We offer three possible reasons for this. First, when responding for positive games compared to negative games, subjects relied more heavily on spectacularity, an alternative form of surprise. This may have led them to de-emphasize win probability-based forms of surprise as a heuristic in reporting their memories. Though this may mask a positive result, future studies with more leading questions may show the importance of this metric (e.g., "What was your most positive and surprising sports memory for an individual game?"). Second, subjects often chose games in which their team won a championship (35 / 77 positive games), which is a primary goal of sports[63]. Often when teams win the championship, their team is already favored to win the final game owing to their superiority, which would reduce both full-game and within-game surprise due to being expected to outscore the other team throughout the game. From a surprise perspective, this result may be remain puzzling, as it is unclear why such an incremental increase in championship likelihood – even to 100% – should be so positive. We speculate that the jump from a near-certainty to a certainty – and all of the surrounding accolades, bragging rights, and victory parades that follow, perhaps after years or decades of unfulfilled fandom – is positive enough and possibly reflects a strong enough *reward magnitude* that fans value these games in spite of the previously high expectations. Factoring in the subjective reward magnitude of winning a championship into an expected value of magnitude times probability could account for this preference[93].

A final possibility relates to the phenomenon of "wishful thinking" on the part of sports fans: sports fans of a preferred team often provide biased reports of games[94] and overestimate their likelihood of winning [[61,95]; for this effect in other contexts, see[56,60,96–98]]. We estimated prediction using our objective win probability model, but wishful thinking could bias this prediction: if the subjective prediction were higher than the objective one for the preferred team, the positive surprise for preferred outcomes would be dampened and negative surprise would be exacerbated. This would lead to greater disappointment for sports fans of the losing team[99–101] and could be partially responsible for the asymmetries we observed in game memories.

In spite of the overall findings consistent with surprise driving long-term memories, many negative season memories exceeded pre-season predictions but corresponded to seasons in which the preferred team lost

the championship (season rank #2 in Figure 4C). These "near-misses"[102,103] have been reported as aversive[104–106] and increase with a higher perceived likelihood of the preferred outcome[107]. Yet, it is notable that many championship losses were also regarded as positive season memories. We speculate that this may come down to the perspective (or timescale) one takes on such a season, how one realistically values alternatives[108], and the specific value one places on championships versus other outcomes[63]. In other words, it may rely on whether it was enjoyable to see a team do so well from the vantage point of the beginning of the season or agonizing to see them fail after being so close to being champions.

To explain the long-term surprise effects in this paper, sports fans must hold some lingering trace of their predictions from prior time points that become compared against much later outcomes. This presents an interesting puzzle for how these memory biases might arise computationally. Consider that an experienced sports fan watching a game might resemble a reinforcement learning agent trained via temporal difference learning (hereafter, TD-RL agent). That is, when encountered with a successive series of states before some final reward outcome, a TD-RL agent learns to predict the value of each state based on its history of reward probabilities and magnitudes[41,109]. Similarly, experienced sports fans can learn to associate game states with win probabilities[8]. Next, consider the following scenarios. In one scenario, the agent's preferred team has a 98% win probability with 10 minutes remaining in a game, and in another scenario, they have a 10% win probability. After this, both cases converge, in that the win probability gradually decreases or increases to 30% with 1 minute remaining. Then, the preferred team loses by the same score in both scenarios. In either scenario, the experienced sports fan, like the TD-RL agent, will have updated their state probabilities (to 30%) by the 1-minute time point, and therefore the final negative signed surprise is equal in both cases. Additionally, with the same final score, the difference between the scenarios cannot be explained by the overall accumulation of positive and negative events. Therefore, in order for long-term surprise to affect memory, the agent must not simply update its predictions in a memoryless[110,111] and rapid manner but instead must keep traces of predictions beyond the immediate moment – such as from the 10-minute mark in these scenarios.

It is intriguing to consider which neural processes produce long-term surprise effects. In sporting events, performance differences accrue over time to decide a winner. This arrangement resembles many lab-based evidence accumulation tasks featuring a series of stimuli (e.g., clicks, visual flashes, visual motion) that unfold over time and create a series of states that later become mapped non-linearly onto some outcome variable (e.g., win or lose, left or right reward trial)[112,113]. Recent findings suggest evidence accumulates and affects behavior across multiple trials and at multiple timescales in regions across the cortex[114–119]. Therefore, a future study could, for example, examine whether midbrain or cortical areas respond differentially to well-controlled trials that manipulate the long-term, multi-event surprise, such as by manipulating the overall architecture of the trial (e.g., more right clicks at the beginning, creating a 98% probability of a right reward trial, versus more left clicks, creating a 10% probability) while controlling for the overall amount of evidence for each possibility.

Another relevant neural effect follows from findings that neurons and neuronal areas integrate evidence at different time scales[120–125]. In the above 98% vs. 10% scenario, a long timescale representation could be formed with 10 minutes remaining and, because it is not immediately updated due to its long timescale, surprise is higher in the 98% than the 10% case when the final result unfolds. However, whether neural integration – which has generally been studied on the scale of seconds to minutes – operates on scales of months (as in the season data here) remains unclear.

The dopamine system is also relevant to consider. Dopamine neurons respond with respect to deviations from predictions, such that better outcomes are associated with higher responses and worse outcomes with lower responses[43,126]. Additionally, recent studies have shown that dopaminergic regions also respond to unsigned prediction errors[8,14,15] and changes in beliefs[127,128]. Dopamine neuron responses are modulated by the reward histories of multiple past trials, indicating that they can integrate information beyond the

immediately preceding moment to create predictions[129]. However, this integration is distinct from how a momentary response might differ based on multiple accumulated events in the past. Additionally, it has been shown that dopamine neuron responses reflect updated predictions on a nearly moment-by-moment basis, in line with TD learning[130]. As noted above, an overtrained TD-RL agent would make optimal predictions but would not show differential responses based on different history in leading to the same momentary state. Therefore, it is unclear whether dopamine would respond differently in the 98% vs. 10% cases in the final moments and whether it plays a role in these memory effects.

The ability to form long-term memories for arbitrary, non-repeated events is most intimately linked with the function of the hippocampus[131,132]. Regarding the computational approach to this problem, some recent artificial intelligence models have included episodic memory modules that simulate the hippocampus by saving contextual attributes of the current state in an ongoing fashion for later reference[92,133–136]. Here, a reasonable solution to how long-term surprise could influence memory is by maintaining prior predictions by continuously saving them and comparing them against later outcomes. A possible future test of this idea is whether a hippocampal amnesiac patient would differ from controls in having a smaller bias towards long-term surprising information but relatively intact memory for momentary surprising information.

By asking for subjects to report a only single positive and negative memory at each timescale, we showed that surprise affects memory accessibility but not necessarily its availability[73,137]. Probing all or a larger subset of basketball memories could determine whether surprise affects availability as well. Antony et al.[8] asked basketball fans to freely recall basketball game clips and found that surprise also drove memory for individual plays when subjects were encouraged to respond with everything they could remember. Therefore, this concern has been partially addressed for momentary surprise in laboratory settings, but future studies could examine whether surprise plays a role by probing a broader swath of memories[64,65]. Another promising future direction follows laboratory findings that surprise boosts surrounding memories in some paradigms[24,138,139], whereas in others it has no effect[3,91]. Therefore, future studies could examine memory effects for information before or after surprising events, such as plays, games, or seasons before or after the chosen memories.

The greatest factor in determining whether an event is surprising is the high-level context[140,141] or schema[142,143]. This factor guides predictions of what should be expected in a particular situation, and therefore surprise varies across domains[144]. Indeed, the broader literature on surprise includes domains as wide-ranging as music[145], language[51], politics[64,65], and narratives[55,142,146,147]. In this paper, we used the term, "surprise", to refer to changes in game or championship probability. While our data suggest that memories were also influenced by other variables, such as seeing spectacular feats of athleticism (which we argue is another form of surprise), who wins a game or championship may be the most relevant variable for sports fans. If the primary task or context differed, such that they were asked to instead count the number of passes or listen carefully to the commentary, we would expect other factors to drive memory. Just as a routine play (e.g., a layup) occurring at a critical moment and corresponding to high surprise would be considered highly surprising in our model, an everyday word would be surprising in a particular context (e.g., "Tornado wind speeds reach over 100 dogs per hour.").

The outcomes of sports contests can be instrumental to one's sense of identity[66] and mood[67], influencing neural substrates heavily linked with emotion[8,86,148] and physiological factors like testosterone[149] and creating lasting memories[68]. In fact, our subjects often noted that both positive and negative outcomes were among the best or worst things that happened to them in a given month or year (and in a few cases, even a lifetime) (Fig S8B). Dovetailing with a prior report showing that surprise influences a host of factors related to memory and emotion[8], these findings underscore the importance of initial predictions in determining how events are remembered in the real world. Our novel approach of probing memories over multiple timescales and events shows that long-term surprises can influence memory in a manner similar to momentary surprise, widening the door for more behavioral and neural questions investigating their underpinnings and

highlighting an overlooked influence in studies of prediction and memory.

Methods

Subjects

Undergraduate subjects (N = 122, 34 female, 18-28 years old, mean: 19.5) were recruited at California Polytechnic State University, San Luis Obispo (Cal Poly). This school is located along the central coast of California and therefore conveniently pools students who grew up as fans from at least four major NBA fan bases (Los Angeles Lakers, Los Angeles Clippers, Sacramento Kings, and the San Francisco Area-based Golden State Warriors). Subjects took the study for course credit among a number of alternative studies. All subjects considered themselves basketball fans and attested to having seen or played in more than 50 basketball games. All subjects gave consent according with Cal Poly IRB #2020-068.

Questionnaire

Before the study, subjects were screened as basketball fans and were told they would be responding with autobiographical positive and negative memories of plays, games, and seasons. We specifically asked them to recall everything from memory and not to look up any information. Each prompt began with an open-ended question about the memory, e.g., "As a FAN, what is your MOST POSITIVE memory of a SINGLE PLAY in a college or professional basketball game?" We asked them to describe in as much specific detail as possible what happened on the court (the play, game, or season itself), any contextual information about when it occurred, such as plays, games, or seasons leading up to the chosen memory, and then information about their experience, such as where they were, how they watched it or tended to watch it, and who they were with or tended to watch it with. To ensure we had as many identifying details as possible for plays and games and to provide for as rich an account as possible given the multidimensional nature of memories[150], we separately asked numerous clarifying follow-up questions. We asked about the teams involved and when the memories occurred, asking subjects to be as specific as possible without looking anything up. We then asked about fandom, including the team for which they were cheering, how big of a fan they were (1 = almost indifferent, 2 = preferred team in that sport, 3 = favorite team in that sport, 4 = favorite team in any sport), whether they were also specifically cheering against the other team (or teams) and how much they disliked that team or teams (1 = indifferent/almost indifferent, 2 = generally root against the team, 3 = least favorite team in that sport, 4 = least favorite team in any sport). Next, we asked about the emotional impact of the memory, including to rate it with respect to other memories in their life. They rated each memory on an 11-point scale (-5 to +5), as the worst/best thing that happened ever/in a given year/month/week/day or that it was a neutral memory (rating of 0). They also estimated how long the play affected them on a 5-point scale, with 1-4 indicating minutes, hours, days, and months, and 5 indicating they were still affected. Finally, in light of data suggesting that positive sports memories may be revisited more often than negative ones[151,152] and the general fact that rehearsing information makes it more accessible, we also asked how many times they estimated they rewatched the play (for play memory), the full game or parts of the game (for game memory), or parts of the season (for season memory).

After responding to each of these questions, we also collected data on subjects' favorite and least favorite men's and women's NCAA basketball teams, favorite and least favorite NBA and WNBA teams, their favorite sport, their favorite team in any sport, an estimation of the number of basketball games they had watched (0-50, 50-200, 200-1000, 1000+), whether they had played basketball [in an organized fashion, only recreationally (e.g., pickup games), only a little (e.g., gym class), or never], and if so, how long they had played.

Subjects were given the option to respond as fans of any major college or professional American sports

league (men's and women's NCAA, NBA, WNBA). While we originally anticipated a more even breakdown of college and professional responses, they skewed heavily towards the NBA in every major category: Of the 122 responses, we were able to identify 70 as positive NBA plays, 77 positive games, 96 positive seasons, 74 negative plays, 74 negative games, and 83 negative seasons. Therefore, we chose to focus our analyses on the NBA for simplicity and to avoid under-powering other analyses, though surprise drives memory in men's college basketball (Antony et al., 2021) and likely in other leagues.

NBA play-by-play data

Play-by-play data from 17 seasons (2004-05 to 2020-21) were first retrieved from an NBA API. These data included information about each game, including the date of the game, the teams involved, quarter and game time updates, score changes, and descriptions of each play. We pared down the data into game events involving a change in context, which included scores, rebounds, turnovers, and tips from jump balls. In total, these data included >5.6M plays from >22K games. We focused on these 17 seasons because nearly all responses were from this time period and they occurred after the NBA expansion from 29 to 30 teams, simplifying the analyses. All geographical moves by teams (e.g., the Seattle Supersonics moving to become the Oklahoma City Thunder) were re-configured to use the current team location and NBA 3-letter code.

Win probability model

For the main play and game data analyses, we operationalized predictions by creating a win probability metric based on four factors from the play-by-play data: the score difference between the two teams [oriented as positive (negative) when the home team was winning (losing)], the relative strength of the teams, the amount of game time remaining, and the team in possession of the ball[8]. In practice, the first two metrics were combined into one as the 'expected score difference' by the end of the game. Using publicly available data, we found for each team the average number of points they scored minus points scored against them during that season; the difference between these subtractions then indicated the relative difference in team strength for an entire game. Note that this measurement does not capture fluctuations in team strength within a season, whether due to injuries, new player arrivals and departures, or other factors. We then divided this by the total number of seconds per game (to estimate how much the stronger team should outscore the weaker team, in units of points per second), multiplied it by the number of seconds left in the game, and subtracted this number from the current score difference. For instance, if for one season the Milwaukee Bucks scored an average of 106 points and had an average of 101 points scored against them (+5 net) and the Boston Celtics scored an average of 98 points and had an average of 97 points scored against them (+1 net), the Bucks would be expected to outscore the Celtics by 4 points over an entire game. Therefore, the expected score difference at the beginning of the game would be 4 points in the Bucks' favor. If the Celtics led by 1 at halftime, the Bucks would nonetheless be expected to outscore them by 2 points over the course of the second half (4 points per game advantage * 50% of the game remaining), so the expected score difference would be 1 point in the Bucks' favor.

To create the model, we divided possible game states into a grid in three dimensions: the expected home minus away team score difference from -30 to 30; the number of seconds remaining in the game in increments of 6 (e.g., 2880 – 2774 s, 2774 – 2768 s) until the final 6 seconds, after which we used increments of 2 (e.g., 6 – 4 s, 4 – 2 s, 2 – 0 s); and whether the home team possessed the ball (1/0). We used finer increments near the end of the game because of the rapid swings in context that occur in NBA games with very little time remaining. To compute win probabilities, we found all possible games matching each state and "peeked ahead" to see how often the home team won the game. Probabilities for contexts with a greater difference in score than 30 points were set to 0.998 or 0.002. This model essentially resembles a "look-up table" rather than an algorithm such as logistic regression. We chose this approach due to non-linearities in the state space. For example, win probabilities for the home team in possession of the ball with

a -4 to +4 point lead with 5 seconds remaining are the following: 0.018, 0.097, 0.197, 0.36, 0.652, 0.876, 0.938, 0.983, 0.998. Additionally, the win probability benefit of having possession of the ball is not uniform across all score differences and game times but rather peaks when the score is close towards the end of the game (Figure S3). To validate this approach, training linear and logistic regression models using the same factors yields higher win prediction error rates as defined by mean differences between all predictions and outcomes (mean error for linear regression: 0.355; logistic regression: 0.316; lookup method: 0.304).

Note that our emphasis is not on the process by which subjects learn the value (win probability) of the game states, akin to the goal of a reinforcement learning agent. Rather, we start from the assumption that subjects have predictive mental models of win probabilities that approximate the true probabilities, as previously validated[8]. A rare event – such as a team coming back to win after having only a 1% win probability – does not mean the prediction was wrong if the same game state occurred 99 other times in which the team lost.

Aligning subject responses with NBA play-by-play data

To find win probabilities and (subsequently) surprise values using our algorithm for play memories, we logged subject responses by the game date, teams involved, quarter, and seconds remaining. Then, we searched through the play-by-play database to find the specific win probability values before and after the chosen play by finding the specific number of seconds remaining when the play ended. In many cases, multiple plays occurred within a single second (such as with free throws, wherein multiple plays elapse with no time expiring, or with shots taken within a single second, such as shots taken with only a fraction of a second on the game clock). In these cases, we manually parsed the data to find the correct moment.

To find win probabilities for game memories, we logged subject responses by the game date and teams involved. To find championship probabilities, we logged subject responses by the season and team involved.

Surprise metrics

Numerous ways of quantifying surprise have been devised [for a systematic review, see[144]]. All surprise constructs used here relate point estimates of how outcomes differ from probabilistic predictions. To find play surprise, we computed the difference in win probabilities from before to after the chosen plays. These values were oriented as positive for the home team and were flipped if the subject preferred the away team to correctly align signed surprise. For unsigned play surprise, we instead used the absolute values. To externally validate our surprise metric for plays, we looked up the signed difference in win probability values (signed surprise) given the game context using an expert analyst's site (Fig. 2B). We compared our surprise metrics against the expert using Pearson correlations.

To find full-game surprise, we computed the difference in pre-game (prior) win probabilities and the final probability (1/0). To find within-game surprise, we first found the lowest in-game win probability of the eventual game winner and computed the difference between this and the eventual outcome. For instance, if the home team won and their pre-game win probability was 0.54 and their lowest win probability was 0.22 in the 3rd quarter, full-game surprise would be 1 – 0.54 = 0.46 and within-game surprise would be 1 – 0.22 = 0.78. Similar to play surprise, both of these values were oriented as positive for the home team and flipped if the subject was cheering for the away team. We also computed unsigned values of these metrics using the absolute values. To externally validate our win probability on full- and within-game surprise, we similarly found the pre-game probability and minimum win probability for the winning team on the same expert analyst's site.

To find season surprise, we used a slightly different approach. As an outcome variable, we created a final season ranking metric that used how far teams survived into the season and playoffs to determine ranking, and we split ties using their regular season record. That is, NBA champions were given a #1

ranking, runner-ups were given a #2 ranking, conference championship finalists were given #3-4 rankings (decided by inverse regular season record), conference semifinalists were given #5-8 rankings, etc. If two teams survived the same amount into the season and had the same regular season record, their rankings were averaged (e.g., if #3-4 had the same record, each was given a ranking of 3.5). This metric was more fine-grained than one that is analogous to games, wherein winners are assigned a probability of 1 and losers 0, which would involve giving the champion a 1 and all other teams 0. An alternative metric would be regular season record; however, (1) many fans would consider this unsatisfactory because regular season is not what determines champions, and (2) to prevent injuries in the playoffs, many recent basketball teams have questioned the utility of maximizing their regular season record in favor of simply playing well enough to position themselves for a playoff run while managing workloads[153,154]. Finally, this metric also resembles how major sports leagues determine the order in which teams can draft new players.

We used three different measures to model subject prediction at the beginning of and throughout the season. Each measure relied on the Vegas championship odds for each team and season as they were updated throughout the year. Before 2009, these odds were only updated before the season and before each round of the playoffs. After 2009, the odds were updated monthly before and during the season and sometimes during the all-star break. Odds are posted according to a formula, such that if they are above 0, the championship likelihood is

$$\frac{100}{100 + odds}$$

and if they are below zero, the odds are

$$\frac{|odds|}{100 + |odds|}$$

For our first predictor, we used the final pre-season odds for pre-season prediction. For our second predictor, we used all of the available (non-final) odds to find the best odds throughout the season for a given team. Our third predictor resembles the first except that it ranks teams by their pre-season odds. We used linear regression to find the best fit between these predictors and the final season ranking metric. We used the logarithm of the 1st and 3rd predictors because doing so produced a more linear fit. We then used the residual from this regression line to determine whether teams did better or worse than expected given this predictive relationship.

Statistical analyses

Our main comparisons in this paper were of two main varieties. First, we compared attributes of positive and negative reported memories against null distributions from the NBA play-by-play corpus and betting odds. Second, we directly compared attributes of positive and negative reported memories themselves. For each test, we first conducted skewness and kurtosis tests of normality. In every case, these tests violated assumptions of normality, so we used two-sided Mann-Whitney U tests for independent-sample comparisons and Wilcoxon signed-rank tests for 1-sample comparisons.

Data and code availability

All relevant data and code necessary to reproduce these results – including how to grab NBA API data – are available here.

Author Contributions

J.W.A. conceived, programmed, and analyzed data from the experiment, and drafted the manuscript. J.V.D. contributed to study design and scored a large portion of the recall data. A.J.B. contributed to study design. K.A.B. contributed to study design and coordinated all data collection.

Acknowledgements

The authors thank Jarett R. Massey for his contributing role at the beginning of the project and assistance with data collection and scoring. The following individuals also assisted with data collection and scoring: (and then leave Jay out of this): Jesús V. Figueroa, Trevor Guerra, Jake Henige, Jay R. Massey, Nami Saito, and Michela Smith. The authors also thank Chung Hee Bennion for her assistance and support throughout the entire project.

*

References

[1] Nina Rouhani, Kenneth Norman, and Yael Niv. Dissociable effects of surprising rewards on learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(9):1430, 2018.

[2] Nina Rouhani, Kenneth A Norman, Yael Niv, and Aaron M Bornstein. Reward prediction errors create event boundaries in memory. *Cognition*, 203:104269, 2020.

[3] Anthony I. Jang, Matthew R. Nassar, Daniel G. Dillon, and Michael J. Frank. Positive reward prediction errors during decision-making strengthen memory encoding. *Nature Human Behaviour*, 3(7): 719–732, 2019.

[4] Jiawen Huang, Isabel Velarde, Wei Ji Ma, and Christopher Baldassano. Schema-based predictive eye movements support sequential memory encoding. *PsyArXiv*, 2022.

[5] R.A. Rescorla and A.R. Wagner. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black and W.F. Prokasy, editors, *Classical Conditioning II: Current Research and Theory.*, pages 64–99. Appleton-Century-Crofts, New York, NY, 1972.

[6] Janet Metcalfe. Learning from errors. *Annual Review of Psychology*, 68:465–489, 2017.

[7] Garvin Brod. Predicting as a learning strategy. *Psychonomic Bulletin and Review*, 28(6):1839–1847, 2021.

[8] J W Antony, T H Hartshorne, K Pomeroy, T M Gureckis, U Hasson, S D McDougle, and K A Norman. Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. *Neuron*, 109(2):377–390, 2021.

[9] Narinder Kapur. Syndromes of retrograde amnesia : A conceptual and empirical synthesis. *Psychological Bulletin*, 125(6):800–825, 1999.

[10] John E Lisman and Anthony A Grace. The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron*, 46(5):703–13, 2005.

[11] Andrea Greve, Elisa Cooper, Alexander Kaula, Michael C Anderson, and Richard Henson. Does prediction error drive one-shot declarative learning ? *Journal of Memory and Language*, 94:149–165, 2017.

[12] Alex Pine, Noa Sadeh, Aya Ben-Yakov, Yadin Dudai, and Avi Mendelsohn. Knowledge acquisition is governed by striatal prediction errors. *Nature Communications*, 9(1):1673, 2018.

[13] Alexandros Kafkas and Daniela Montaldi. Striatal and midbrain connectivity with the hippocampus selectively boosts memory for contextual novelty. *Hippocampus*, 25(11):1262–1273, 2015.

[14] Yuji K. Takahashi, Hannah M. Batchelor, Bing Liu, Akash Khanna, Marisela Morales, and Geoffrey Schoenbaum. Dopamine neurons respond to errors in the prediction of sensory features of expected rewards. *Neuron*, 95:1395–1405.e3, 2017.

[15] James D. Howard and Thorsten Kahnt. Identity prediction errors in the human midbrain update reward-identity expectations in the orbitofrontal cortex. *Nature Communications*, 9:1611, 2018.

[16] Darya Frank, Alex Kafkas, and Daniela Montaldi. Experiencing surprise: The temporal dynamics of its impact on memory. *The Journal of Neuroscience*, 42(33):6435–6444, 2022.

[17] Vishnu P. Murty, Kevin S. LaBar, and R. Alison Adcock. Distinct medial temporal networks encode surprise during motivation by reward versus punishment. *Neurobiology of Learning and Memory*, 134:55–64, oct 2016.

[18] Katherine Duncan, Nicholas Ketz, Souheil J. Inati, and Lila Davachi. Evidence for area CA1 as a match/mismatch detector: A high-resolution fMRI study of the human hippocampus. *Hippocampus*, 22(3):389–398, 2012.

[19] Oded Bein, Katherine Duncan, and Lila Davachi. Mnemonic prediction errors bias hippocampal states. *Nature Communications*, 11:3451, 2020.

[20] Marta I Garrido, Gareth R Barnes, Dharshan Kumaran, Eleanor A Maguire, and Raymond J Dolan. Ventromedial prefrontal cortex drives hippocampal theta oscillations induced by mismatch computations. *NeuroImage*, 120:362–370, 2015.

[21] Nicole M Long, Hongmi Lee, and Brice A Kuhl. Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *Journal of Neuroscience*, 36(50): 12677–87, 2016.

[22] Matthias J Gruber, Liang-tien Hsieh, Bernhard P Staresina, Christian E Elger, Juergen Fell, Nikolai Axmacher, and Charan Ranganath. Theta phase synchronization between the human hippocampus and prefrontal cortex increases during encoding of unexpected information: A case study. *Journal of Cognitive Neuroscience*, 30(11):1646–56, 2017.

[23] Darya Frank, Marcelo A Montemurro, and Daniela Montaldi. Pattern separation underpins expectation-modulated memory. *Journal of Neuroscience*, 40(17):3455–3464, 2020.

[24] Alyssa H Sinclair, Grace M Manalili, Iva K Brunec, R Alison Adcock, and Morgan D Barense. Prediction errors disrupt hippocampal representations and update episodic memories. *PNAS*, 118 (51):1–12, 2021.

[25] Kerstin Preuschoff, Bernard Marius 't Hart, and Wolfgang Einhäuser. Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, 5:1–12, 2011.

[26] Senne Braem, Ena Coenen, Klaas Bombeke, Marlies E. van Bochove, and Wim Notebaert. Open your eyes for prediction errors. *Cognitive, Affective and Behavioral Neuroscience*, 15:374–80, 2015.

[27] A L S Filipowicz, C M Glaze, J W Kable, and J I Gold. Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *eLife*, 9:e57872, 2020.

[28] David Clewett, Camille Gasser, and Lila Davachi. Dynamic arousal signals construct memories of time and events. *Nature Communications*, 11:4007, 2020.

[29] Maria Theobald, Elena Galeano-keiner, Garvin Brod, Maria Theobald, Elena Galeano-keiner, Garvin Brod, and Maria Theobald. Predicting vs guessing: The role of confidence for pupillometric markers of curiosity and surprise. *Cognition and Emotion*, 36(4):731–740, 2022.

[30] Dirk Ostwald, Bernhard Spitzer, Matthias Guggenmos, Timo T Schmidt, Stefan J Kiebel, and Felix Blankenburg. Evidence for neural encoding of Bayesian surprise in human somatosensation. *NeuroImage*, 62(1):177–188, 2012.

[31] Richárd Reichardt, Bertalan Polner, and Péter Simor. Novelty manipulations, memory performance, and predictive coding: The role of unexpectedness. *Frontiers in Human Neuroscience*, 14(April): 1–11, 2020.

[32] John M. Pearce and Geoffrey Hall. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87(6):532–52, 1980.

[33] Peter Redgrave, Tony J Prescott, Kevin Gurney, Peter Redgrave, and Tony J Prescott. Is the short-latency dopamine response too short to signal reward error? *Trends in Neurosciences*, 2236(98): 146–151, 1999.

[34] Maurizio Corbetta and Gordon L. Shulman. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3):201–215, mar 2002.

[35] Charan Ranganath and Gregor Rainer. Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, 4(March):193–202, 2003.

[36] Margaret M Bradley. Natural selective attention: Orienting and emotion. *Psychophysiology*, 46:1–11, 2009.

[37] Freya Sukalla, Heather Shoenberger, and Paul D Bolls. Surprise! An investigation of orienting responses to test assumptions of narrative processing. *Communication Research*, 43(6):844–862, 2016.

[38] Elise Demeter, Brittany Glassberg, Marissa L Gamble, and Marty G Woldorff. Reward magnitude enhances early attentional processing of auditory stimuli. *Cognitive, Affective, & Behavioral Neuroscience*, 22:268–280, 2022.

[39] Laurent Itti and Pierre Baldi. Bayesian surprise attracts human attention. *Vision Research*, 49(10): 1295–306, 2009.

[40] Nina Rouhani and Yael Niv. Signed and unsigned reward prediction errors dynamically enhance learning and memory. *eLife*, 10:e61077, 2021.

[41] Yael Niv and Geoffrey Schoenbaum. Dialogues on prediction errors. *TICS*, 12(7):265–272, 2008.

[42] N. J. Mackintosh. A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82(4):276–298, jul 1975.

[43] Wolfram Schultz, Peter Dayan, and P Read Montague. A neural substrate of prediction and reward. *Science*, 275:1593–9, 1997.

[44] K D'Ardenne, SM McClure, LE Nystrom, and JD Cohen. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, 319:1264–1267, 2008.

[45] Hannah M Bayer and Paul W Glimcher. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47(1):129–41, jul 2005.

[46] Cristian B. Calderon, Esther de Loof, Kate Ergo, Anna Snoeck, Carsten N. Boehler, and Tom Verguts. Signed reward prediction errors in the ventral striatum drive episodic memory. *Journal of Neuroscience*, 41(8):1716–1726, 2021.

[47] Esther De Loof, Lien Naert, Filip Van Opstal, and Tom Verguts. Signed reward prediction errors drive episodic learning. *PLoS ONE*, 13(1):e0189212, 2018.

[48] Ida Momennejad and Marc W Howard. Predicting the future with multi-scale successor representations. *bioRxiv*, pages 1–27, 2018.

[49] G Elliott Wimmer and Christian Büchel. Learning of distant state predictions by the orbitofrontal cortex in humans. *Nature Communications*, 10:2554, 2019.

[50] Zoran Tiganj, Wei Tang, and Marc W Howard. A computational model for simulating the future using a memory timeline. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43: 1173–9, 2021.

[51] Milena Rabovsky, Steven S Hansen, and James L Mcclelland. Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, 2:693–705, 2018.

[52] Ariel Goldstein, Zaid Zada, Eliav Buchnik, Mariano Schain, Amy Price, Bobbi Aubrey, Samuel A Nastase, Amir Feder, Dotan Emanuel, Alon Cohen, Aren Jansen, Harshvardhan Gazula, Gina Choe, Aditi Rao, Catherine Kim, Colton Casto, Lora Fanda, Werner Doyle, Daniel Friedman, Patricia Dugan, Lucia Melloni, Roi Reichart, Sasha Devore, Adeen Flinker, Liat Hasenfratz, Omer Levy, Avinatan Hassidim, Michael Brenner, Yossi Matias, Kenneth A Norman, Orrin Devinsky, and Uri Hasson. Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, 25(3):369–380, mar 2022.

[53] Manoj Kumar, Ariel Goldstein, Sebastian Michelmann, Jeffrey M Zacks, and Uri Hasson. Bayesian surprise predicts human event segmentation in story listening. *PsyArXiv*, pages 1–38, 2022.

[54] Mante S Nieuwland and Jos J A Van Berkum. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7):1098–1111, 2006.

[55] Hsiang-yun Sherry Chien and Christopher J Honey. Constructing and Forgetting Temporal Context in the Human Cerebral Cortex. *Neuron*, 106:675–686, 2020.

[56] Neil D Weinstein. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5):806–820, 1980.

[57] Hal Ersner-Hershfield, G Elliott Wimmer, and Brian Knutson. Saving for the future self : Neural measures of future self-continuity predict temporal discounting. *SCAN*, 4:85–92, 2009.

[58] Sasha Brietzke and Meghan L. Meyer. Temporal self-compression: Behavioral and neural evidence that past and future selves are compressed as they move away from the present. *Proceedings of the National Academy of Sciences of the United States of America*, 118(49), 2021.

[59] George Loewenstein and David Schkade. Wouldn't it be nice? Predicting future feelings. In Daniel Kahneman, Ed Diener, and Norbert Schwarz, editors, *Well-Being: Foundations of Hedonic Psychology*, pages 85–105. Russell Sage Foundation, 1997.

[60] Donald Granberg and Edward Brent. When prophecy bends: the preference-expectation link in U.S. presidential elections, 1952-1980. *Journal of Personality and Social Psychology*, 45(3):477–491, 1983.

[61] Elisha Babad and Yosl Katz. Wishful thinking–against all odds. *Journal of Applied Social Psychology*, 21(23):1921–1938, 1991.

[62] Patrick J Ferguson and Karim R Lakhani. Consuming contests: Outcome uncertainty and spectator demand for contest-based entertainment. *Harvard Business School*, 2021.

[63] James D Whitney. Winning games versus winning championships: The economics of fan interest and team performance. *Economic Inquiry*, XXVI(October):703–724, 1988.

[64] Adam R Congleton and Dorthe Berntsen. How suspense and surprise enhance subsequent memory: The case of the 2016 United States Presidential Election. *Memory*, 30(3):317–329, 2022.

[65] Kimberly S. Chiew, Bailey B. Harris, and R. Alison Adcock. Remembering election night 2016: Subjective but not objective metrics of autobiographical memory vary with political affiliation, affective valence, and surprise. *Journal of Experimental Psychology: General*, 151(2):390–409, 2021.

[66] N R Branscombe and D L Wann. The positive social and self concept consequences of sports team identification. *Journal of Sport and Social Issues*, 15(2):115–127, 1991.

[67] L Wann and R Branscombe. Die-hard and fair-weather fans: Effects of identification on BIRGing and CORFing tendencies. *Journal of Sport and Social Issues*, 14(2):103–117, 1990.

[68] Elizabeth A Kensinger. When the Red Sox shocked the Yankees: Comparing negative and positive memories. *Psychonomic Bulletin & Review*, 13(5):757–763, 2006.

[69] Elizabeth A Kensinger and Suzanne Corkin. Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words ? *Memory & Cognition*, 31(8):1169–1180, 2003.

[70] Jingyi Wang, Arielle Tambini, and Regina C. Lapate. The tie that binds: temporal coding and adaptive emotion. *Trends in Cognitive Sciences*, 26(12):1103–1118, 2022.

[71] R Brown and J Kulik. Flashbulb memories. *Cognition*, 5:73–99, 1977.

[72] Florin Dolcos, Kevin S Labar, and Roberto Cabeza. Remembering one year later: Role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. *PNAS*, 102(7): 2626–2631, 2005.

[73] Samantha E Williams, Jaclyn H Ford, and Elizabeth A Kensinger. The power of negative and positive episodic memories. *Cognitive, Affective, & Behavioral Neuroscience*, pages 1–35, 2022.

[74] Tali Sharot, Mauricio R Delgado, and Elizabeth A Phelps. How emotion enhances the feeling of remembering. *Nature Neuroscience*, 7(12):1376–1380, 2004.

[75] Eleanor A. Maguire and Catherine J. Mummery. Differential modulation of a common memory retrieval network revealed by positron emission tomography. *Hippocampus*, 9(1):54–61, 1999.

[76] David B. Pillemer. Flashbulb memories of the assassination attempt on President Reagan. *Cognition*, 16(1):63–80, 1984.

[77] John Neil Bohannon. Flashbulb memories for the space shuttle disaster: A tale of two theories. *Cognition*, 29(2):179–196, 1988.

[78] Daniel B. Wright. Recall of the Hillsborough disaster over time: Systematic biases of 'flashbulb' memories. *Applied Cognitive Psychology*, 7(2):129–138, 1993.

[79] William Hirst, Elizabeth A Phelps, Randy L Buckner, Marcia K Johnson, Keith B Lyle, Mara Mather, and Karen J Mitchell. Long-term memory for the terrorist attack of September 11: Flashbulb memories, event memories, and the factors that influence their retention William. *Journal of Experimental Psychology: General*, 138(2):161–176, 2009.

[80] Ho Ka Chan and Taro Toyoizumi. An economic decision-making model of anticipated surprise with dynamic expectation. *arXiv*, 2021.

[81] Andy Clark. A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenology and the Cognitive Sciences*, 17(3):521–534, 2018.

[82] Abigail Hsiung, Jia-hou Poh, Scott A Huettel, and R Alison Adcock. Spoiler alert! Curiosity encourages patience and joy in the presence of uncertainty. *PsyArXiv*, pages 1–34, 2022.

[83] Gan Su-lin, Charles A Tuggle, Michael A Mitrook, Sylvere H Coussement, and Dolf Zillman. The thrill of a close game: Who enjoys it and who doesn't? *Journal of Sport & Social Issues*, 21(1): 53–64, 1997.

[84] Jeffrey Ely, Alexander Frankel, and Emir Kamenica. Suspense and surprise. *Journal of Political Economy*, 123:215–60, 2015.

[85] Andra Geana, Robert C Wilson, Nathaniel Daw, and Jonathan D Cohen. Boredom, information-seeking and exploration. *CogSci*, pages 1751–6, 2016.

[86] Mina Cikara, Matthew M. Botvinick, and Susan T. Fiske. Us versus them: Social identity shapes neural responses to intergroup competition and harm. *Psychological Science*, 22(3):306–13, 2011.

[87] William J. Villano, A. Ross Otto, C. E.Chiemeka Ezie, Roderick Gillis, and Aaron S. Heller. Temporal dynamics of real-world emotion are more strongly linked to prediction error than outcome. *Journal of Experimental Psychology: General*, 149(9):1755–1766, 2020.

[88] Sudeep Bhatia, Barbara Mellers, and Lukasz Walasek. Affective responses to uncertain real-world outcomes: Sentiment change on Twitter. *PLoS ONE*, 14(2):1–13, 2019.

[89] N. T. Feather. Valence of outcome and expectation of success in relation to task difficulty and perceived locus of control. *Journal of Personality and Social Psychology*, 7:372–386, 1967.

[90] J S Verinis, Brandsma J M, and C N Cofer. Discrepancy from expectation in relation to affect and motivation: Tests of Mcclelland'S hypothesis. *Journal of Personality and Social Psychology*, 9(1): 47–58, 1968.

[91] Aya Ben-Yakov, Verity Smith, and Richard Henson. The limited reach of surprise: Evidence against effects of surprise on memory for preceding elements of an event. *Psychonomic Bulletin & Review*, 29(3):1053–1064, 2021.

[92] Samuel J. Gershman and Nathaniel D. Daw. Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, 68:101–128, 2017.

[93] Brian Knutson, Jonathan Taylor, Matthew Kaufman, Richard Peterson, and Gary Glover. Distributed neural representation of expected value. *Journal of Neuroscience*, 25(19):4806–4812, 2005.

[94] Albert H Hastorf and Hadley Cantril. Case reports they saw a game: a case study. *The Journal of Abnormal and Social Psychology*, 49(1):129–134, 1951.

[95] Paul C Price. Wishful thinking in the prediction of competitive outcomes. *Thinking & Reasoning*, 6 (2):161–172, 2000.

[96] Rose W. Marks. The effect of probability, desirability, and ”privilege” on the stated expectations of children. *Journal of Personality*, 19(3):332–351, mar 1951.

[97] Francis W. Irwin. Stated Expectations as Functions of Probability and Desirability of Outcomes. *Journal of Personality*, 21(3):329–335, mar 1953.

[98] Francis W. Irwin and Joan G. Snodgrass. Effects of independent and dependent outcome values upon bets. *Journal of Experimental Psychology*, 71(2):282–285, 1966.

[99] David E Bell. Disappointment in decision making under uncertainty. *Operations research*, 33(1), 1985.

[100] David W Rainey, Janet Larsen, and John H Yost. Disappointment theory and disappointment among baseball fans. *Journal of Sport Behavior,*, 32(3), 2009.

[101] Wilco Van Dijk. The impact of probability and magnitude of outcome on disappointment and elation. *Organizational Behavior and Human Decision Processes*, 69(3):277–284, 1997.

[102] Daniel Kahneman and Daniel Friedman. The Psychology of Preferences. *Scientific American*, pages 160–173, 1981.

[103] R. L. Reid. The psychology of the near miss. *Journal of Gambling Behavior*, 2(1):32–39, 1986.

[104] Luke Clark, Andrew J Lawrence, Frances Astley-Jones, and Nicola Gray. Gambling near-misses enhance motivation to gamble and recruit win-related brain circuitry. *Neuron*, 61:481–490, 2009.

[105] Desmond C Ong, Noah D Goodman, and Jamil Zaki. Near-misses sting even when they are uncontrollable. *CogSci*, 2015.

[106] V H Medvec, S F Madey, and T Gilovich. When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology*, 69(4):603–610, 1995.

[107] Luke Clark, Ben Crooks, Robert Clarke, Michael R F Aitken, and Barnaby D Dunn. Physiological responses to near-miss outcomes and personal control during simulated gambling. *Journal of Gambling Studies*, 28:123–137, 2012.

[108] B. A. Mellers, A. Schwartz, K. Ho, and I. Ritov. Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science*, 8(6):423–429, 1997.

[109] R. S. Sutton and A. G. Barto. *Introduction to reinforcement learning.* Cambridge: MIT press., vol 2, no. edition, 1998.

[110] Samuel J Gershman, Christopher D Moore, Michael T Todd, and Kenneth A Norman. The successor representation and temporal context. *Neural Computation*, 24:1553–68, 2012.

[111] Peter Dayan. Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, 5(4):613–624, 1993.

[112] WT Newsome, KH Britten, and JA Movshon. Neuronal correlates of a perceptual decision. *Nature*, 341:52–54, 1989.

[113] Bingni W Brunton, Matthew M Botvinick, and Carlos D Brody. Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340(6128):95–8, apr 2013.

[114] Lucas Pinto, David W. Tank, and Carlos D. Brody. Multiple timescales of sensory-evidence accumulation across the dorsal cortex. *eLife*, 11:1–21, 2022.

[115] Ari S Morcos and Christopher D Harvey. History-dependent variability in population dynamics during evidence accumulation in cortex. *Nature Neuroscience*, 19(12):1672–1681, 2017.

[116] Athena Akrami, Charles D. Kopec, Mathew E. Diamond, and Carlos D. Brody. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, 554(7692):368–372, 2018.

[117] Zenas C. Chao, Kana Takaura, Liping Wang, Naotaka Fujii, and Stanislas Dehaene. Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. *Neuron*, 100 (5):1252–1266.e3, 2018.

[118] Catherine Wacongne, Etienne Labyt, Virginie Van Wassenhove, Tristan Bekinschtein, Lionel Naccache, and Stanislas Dehaene. Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51): 20754–20759, 2011.

[119] Eun Jung Hwang, Jeffrey E. Dahlen, Madan Mukundan, and Takaki Komiyama. History-based action selection bias in posterior parietal cortex. *Nature Communications*, 8(1):1–14, 2017.

[120] Christopher Baldassano, Janice Chen, Asieh Zadbood, Jonathan W. Pillow, Uri Hasson, and Kenneth A. Norman. Discovering event structure in continuous narrative perception and memory. *Neuron*, 95(3):709–721.e5, 2017.

[121] Sarah Folkerts, Ueli Rutishauser, and Marc W. Howard. Human episodic memory retrieval is accompanied by a neural contiguity effect. *Journal of Neuroscience*, 38(17):4200–4211, 2018.

[122] Albert Tsao, Jørgen Sugar, Li Lu, Cheng Wang, James J. Knierim, May Britt Moser, and Edvard I. Moser. Integrating time from experience in the lateral entorhinal cortex. *Nature*, 561(7721):57–62, 2018.

[123] Ian M. Bright, Miriam L.R. Meister, Nathanael A. Cruzado, Zoran Tiganj, Elizabeth A. Buffalo, and Marc W. Howard. A temporal record of the past with a spectrum of time constants in the monkey entorhinal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 117(33):20274–20283, 2020.

[124] Uri Hasson, Orit Furman, Dav Clark, Yadin Dudai, and Lila Davachi. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron*, 57(3):452–462, feb 2008.

[125] Christopher J. Honey, Thomas Thesen, Tobias H. Donner, Lauren J. Silbert, Chad E. Carlson, Orrin Devinsky, Werner K. Doyle, Nava Rubin, David J. Heeger, and Uri Hasson. Slow cortical dynamics and the accumulation of information over long timescales. *Neuron*, 76(2):423–434, 2012.

[126] Roy A Wise. Dopamine, learning and motivation. *Nature Reviews Neuroscience*, 5(June):1–12, 2004.

[127] Matthew M. Nour, Tarik Dahoun, Philipp Schwartenbeck, Rick A. Adams, Thomas H.B. FitzGerald, Christopher Coello, Matthew B. Wall, Raymond J. Dolan, and Oliver D. Howes. Dopaminergic basis for signaling belief updates, but not surprise, and the link to paranoia. *Proceedings of the National Academy of Sciences of the United States of America*, 115(43):E10167–E10176, 2018.

[128] Clara Kwon Starkweather, Benedicte M. Babayan, Naoshige Uchida, and Samuel J. Gershman. Dopamine reward prediction errors reflect hidden-state inference across time. *Nature Neuroscience*, 20(4):581–9, 2017.

[129] Ethan S Bromberg-Martin, Masayuki Matsumoto, Hiroyuki Nakahara, and Okihide Hikosaka. Multiple timescales of memory in lateral habenula and dopamine neurons. *Neuron*, 67(3):499–510, 2010.

[130] Hyunggoo R Kim, Athar N Malik, John G Mikhael, Mitsuko Watabe-uchida, Samuel J Gershman, and Naoshige Uchida. A unified framework for dopamine signals across timescales. *Cell*, 183:1600–1616, 2020.

[131] Lila Davachi. Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, 16(6):693–700, 2006.

[132] H. Eichenbaum, A. P. Yonelinas, and C. Ranganath. The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, 30(1):123–152, 2007.

[133] Alexander Pritzel, Benigno Uria, Sriram Srinivasan, Demis Hassabis, Daan Wierstra, and Charles Blundell. Neural episodic control. *Proceedings of the 34th International Conference on Machine Learning*, 2016.

[134] Samuel Ritter, Jane X Wang, Zeb Kurth-nelson Siddhant, Charles Blundell, Razvan Pascanu, and Matthew Botvinick. Been there, done that: Meta-learning with episodic recall. *Proceedings of the 35th International Conference on Machine Learning*, 2018.

[135] Qihong Lu, Uri Hasson, and Kenneth A Norman. Learning to use episodic memory for event prediction. *bioRxiv*, 2020.

[136] Máté Lengyel and Peter Dayan. Hippocampal contributions to control: The third way. *Advances In Neural Information Processing Systems*, 20:889–896, 2008.

[137] Endel Tulving and Zena Pearlstone. Availability versus accessibility of information in memory for words. *Journal Of Verbal Learning And Verbal Behavior*, 5:381–391, 1966.

[138] Adam R Congleton and Dorthe Berntsen. The devil is in the details: Investigating the influence of emotion on event memory using a simulated event. *Psychological Research*, 84(8):2339–2353, 2020.

[139] Erin Kendall Braun, G Elliott Wimmer, and Daphna Shohamy. Retroactive and graded prioritization of memory by reward. *Nature Communications*, 9:4886, 2018.

[140] Richard N Henson and Pierre Gagnepain. Predictive, interactive multiple memory systems. *Hippocampus*, 1326:1315–1326, 2010.

[141] Jörn Alexander Quent, Richard N. Henson, and Andrea Greve. A predictive account of how novelty influences declarative memory. *Neurobiology of Learning and Memory*, 179(July 2020), 2021.

[142] Walter Kintsch. Learning from text, levels of comprehension, or: Why anyone would read a story anyway. *Poetics*, 9:87–98, 1980.

[143] Rainer Reisenzein, Gernot Horstmann, and Achim Schützwohl. The cognitive-evolutionary model of surprise: A review of the evidence. *Topics in Cognitive Science*, 11:50–74, 2019.

[144] Alireza Modirshanechi, Johanni Brea, and Wulfram Gerstner. Surprise: a unified theory and experimental predictions. *bioRxiv*, 2021.

[145] Benjamin P. Gold, Ernest Mas-Herrero, Yashar Zeighami, Mitchel Benovoy, Alain Dagher, and Robert J. Zatorre. Musical reward prediction errors engage the nucleus accumbens and motivate learning. *Proceedings of the National Academy of Sciences of the United States of America*, 116(8): 3310–5, 2019.

[146] V. Tobin. *Elements of surprise: Our mental limits and the satisfactions of plot.* Harvard University Press., 2018.

[147] Jesús Bermejo-Berros, Jaime Lopez-Diez, and Miguel Angel Gil Martínez. Inducing narrative tension in the viewer through suspense, surprise, and curiosity. *Poetics*, 93(PB):101664, 2022.

[148] Anne Botzung, David C Rubin, Amanda Miles, Roberto Cabeza, and Kevin S Labar. Mental hoop diaries: Emotional memories of a college basketball game in rival fans. *Journal of Neuroscience*, 30 (6):2130–2137, 2010.

[149] Paul C Bernhardt, James M Dabbs, Julie A Fielden, and Candice D Lutter. Testosterone changes during vicarious experiences of winning and losing among fans at sporting events. *Physiology & Behavior*, 65(1):59–62, 1998.

[150] Tyler M Tomita, Morgan D Barense, and Christopher J Honey. The similarity structure of real-world memories. *bioRxiv*, pages 1–38, 2021.

[151] Carolyn W Breslin and Martin A Safer. Effects of event valence on long-term memory for two baseball championship games. *Psychological Science*, 22(11):1408–1412, 2011.

[152] Jennifer M Talarico and Kira M Moore. Memories of 'The Rivalry': Differences in how fans of the winning and losing teams remember the same game. *Applied Cognitive Psychology*, 756:746–756, 2012.

[153] Torbjørn Soligard, Martin Schwellnus, Juan-manuel Alonso, Roald Bahr, Ben Clarsen, H Paul Dijkstra, Tim Gabbett, Michael Gleeson, Martin Hägglund, Mark R Hutchinson, Christa Janse Van Rensburg, Karim M Khan, Romain Meeusen, John W Orchard, Babette M Pluim, Martin Raftery, Richard Budgett, and Lars Engebretsen. How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *British Journal of Sports Medicine*, 50: 1030–1041, 2016.

[154] Masaru Teramoto, Chad L Cross, Daniel M Cushman, Travis G Maak, David J Petron, and Stuart E Willick. Game injuries in relation to game schedules in the National Basketball Association. *Journal of Science and Medicine in Sport*, 20(3):230–235, 2017.

Supplementary figures

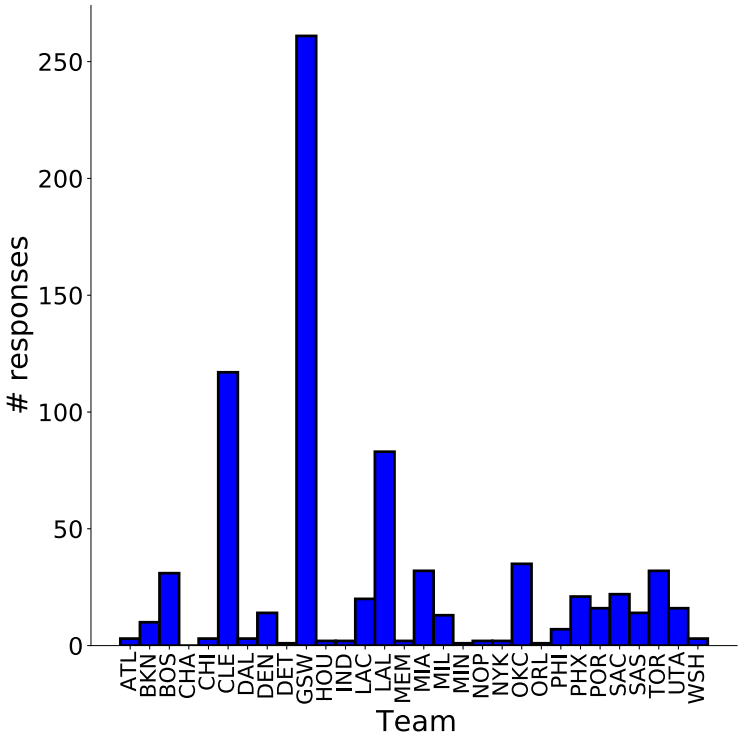

Figure S1: Distribution of responses for each team combined for plays, games, and seasons.

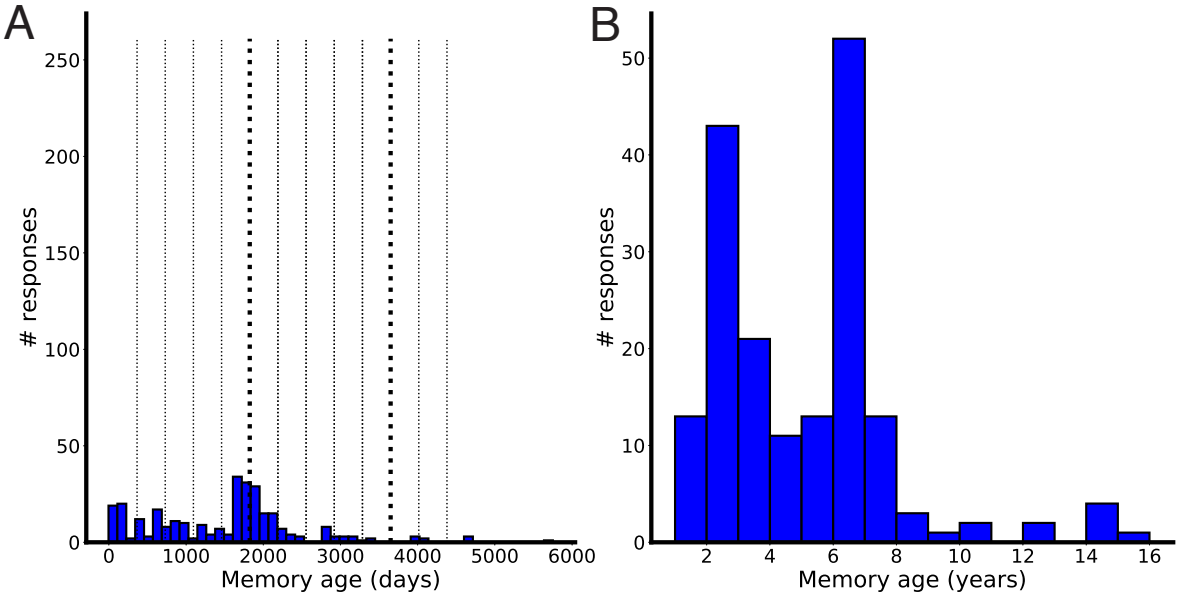

Figure S2: Memory ages. Data were collected over the course of 9 months, so as an alternative to Fig 1C-D, we plotted the time in days between the game date and study date for plays and games (A) or the time in years between the season start and study date (B).

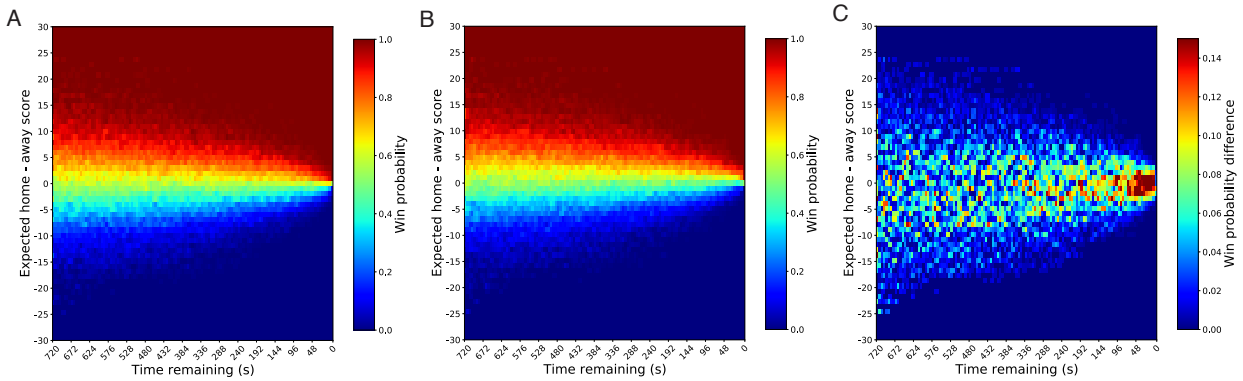

Figure S3: Full win probability model. (A-B) We plotted win probability by expected home minus away score difference and game time remaining when the home (A) and away teams had possession of the ball. To illustrate the importance of team possession, we also plotted the contrast of home possession minus away possession (C). All graphs show data only from the 4th quarter of games.

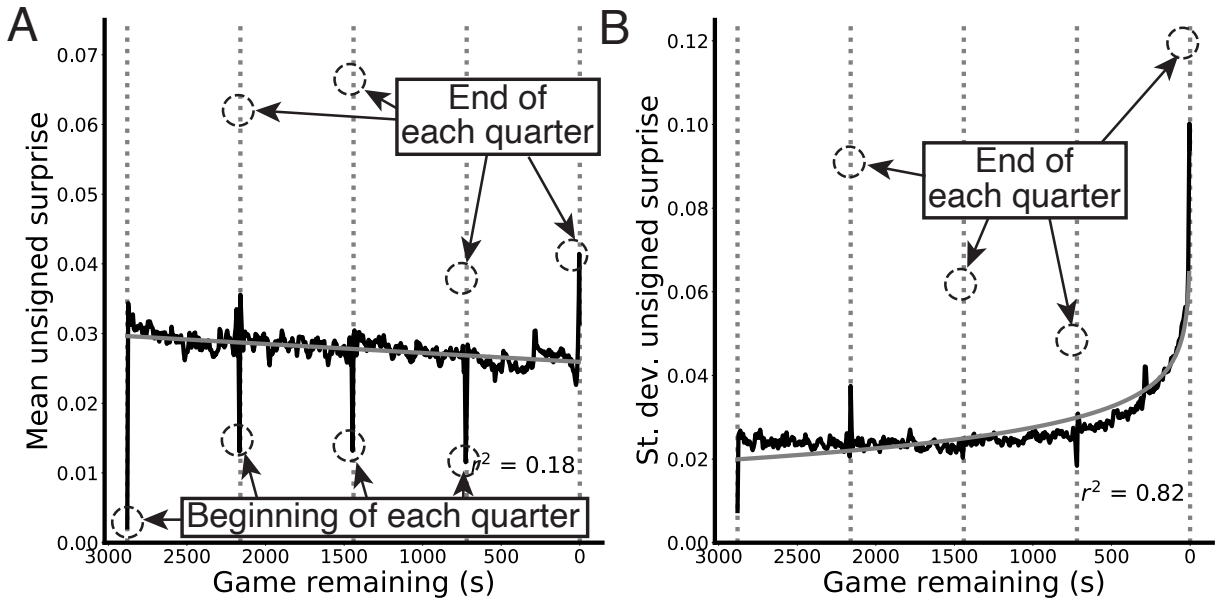

Figure S4: Unsigned surprise over time across a game. (A-B) We plotted the mean (A) and standard deviation of surprise (B) against game time using our metric. Mean surprise had momentary troughs and peaks at the beginning and end of each game, likely due to the momentary decrease and increase in shots taken within these short time frames. Interestingly, mean surprise tended to decrease throughout the game, owing to the probability that the outcome of many basketball games would be nearly decided at that point. However, standard deviation increased exponentially until the end of games, due to the combination of games with a near-certain outcome, which have almost no surprise, and close games, which involve increasingly rapid swings in win probability space as one nears the finish.

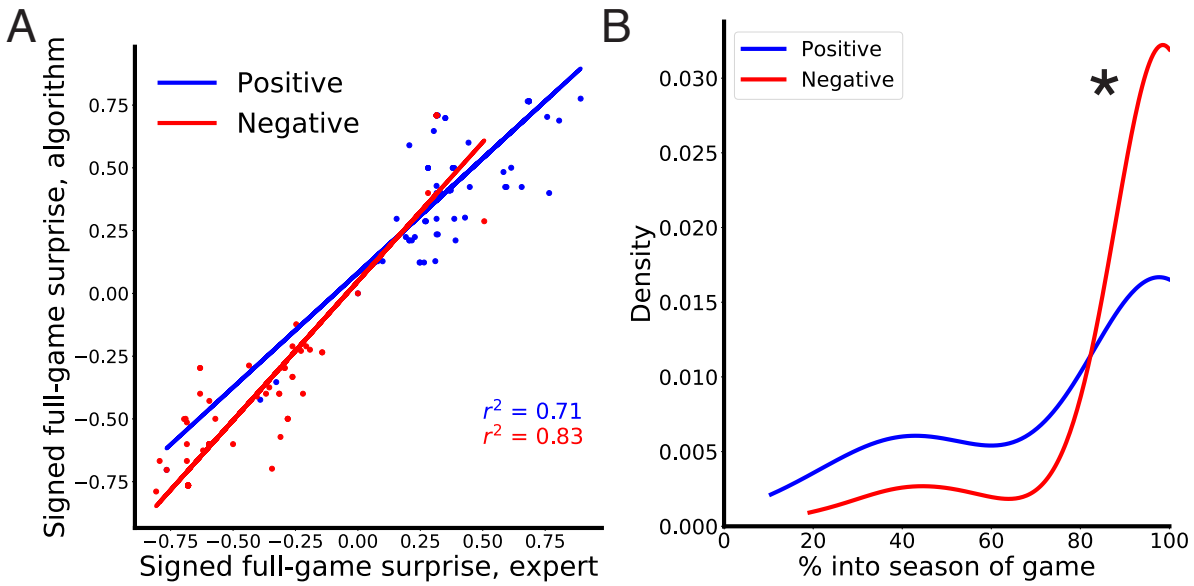

Figure S5: Supplemental figures for game memories. (A) Signed full-game surprise values correlated strongly with those from an expert site. (B) Both positive and negative game responses were strongly biased towards the end of the corresponding team's season, and negative responses were significantly more biased. *: $p = 0.01$

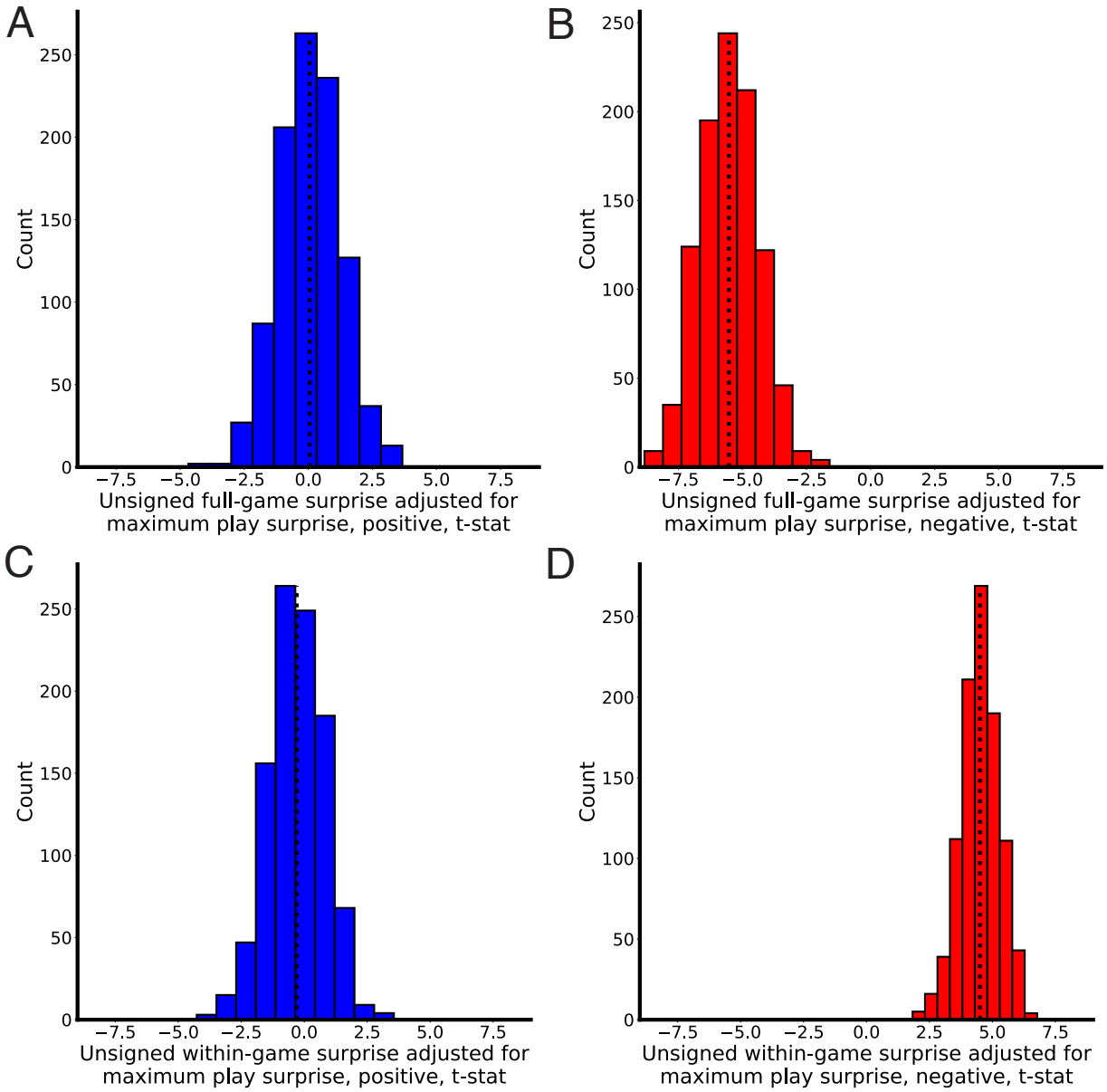

Figure S6: Unsigned full- and within-game surprise after adjusting for differences in play surprise. (A-D) We plotted the t-statistics of subsets of the game data that did not differ in maximum play surprise. (A-B) T-statistics for positive (A) and negative (B) full-game surprise. (C-D) T-statistics for positive (C) and negative (D) within-game surprise. Both negative distributions remained significant, whereas both positive ones did not.

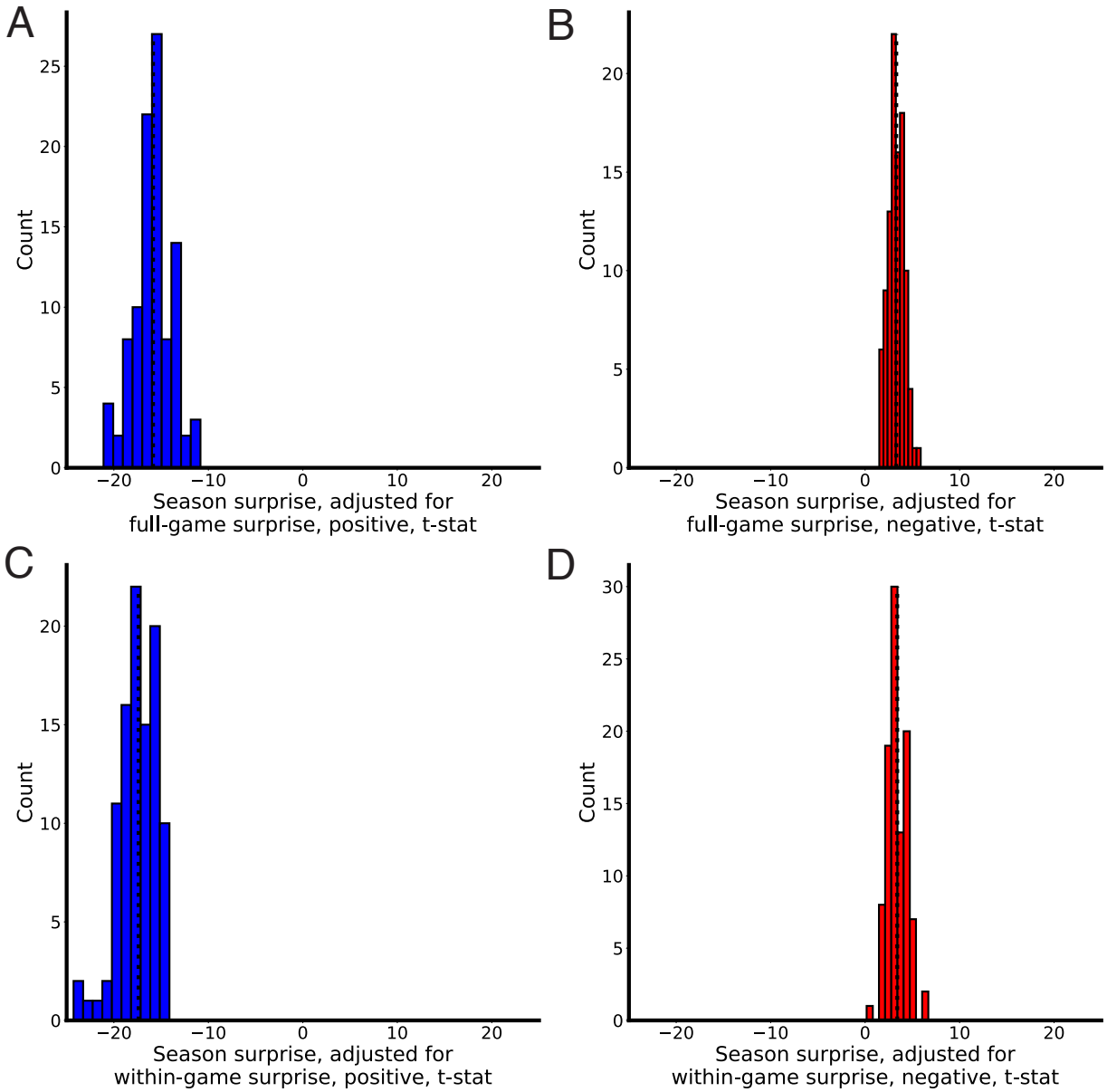

Figure S7: Season surprise after adjusting for differences in maximum full- and within-game surprise. (A-D) We plotted the t-statistics of subsets of the season data that did not differ from the null distribution in maximum game surprise. (A-B) T-statistics for positive (A) and negative (B) seasons, adjusting for maximum full-game surprise. (C-D) T-statistics for positive (C) and negative (D) seasons, adjusting for within-game maximum surprise. All distributions remained significant after adjusting for game surprise deviations from the null.

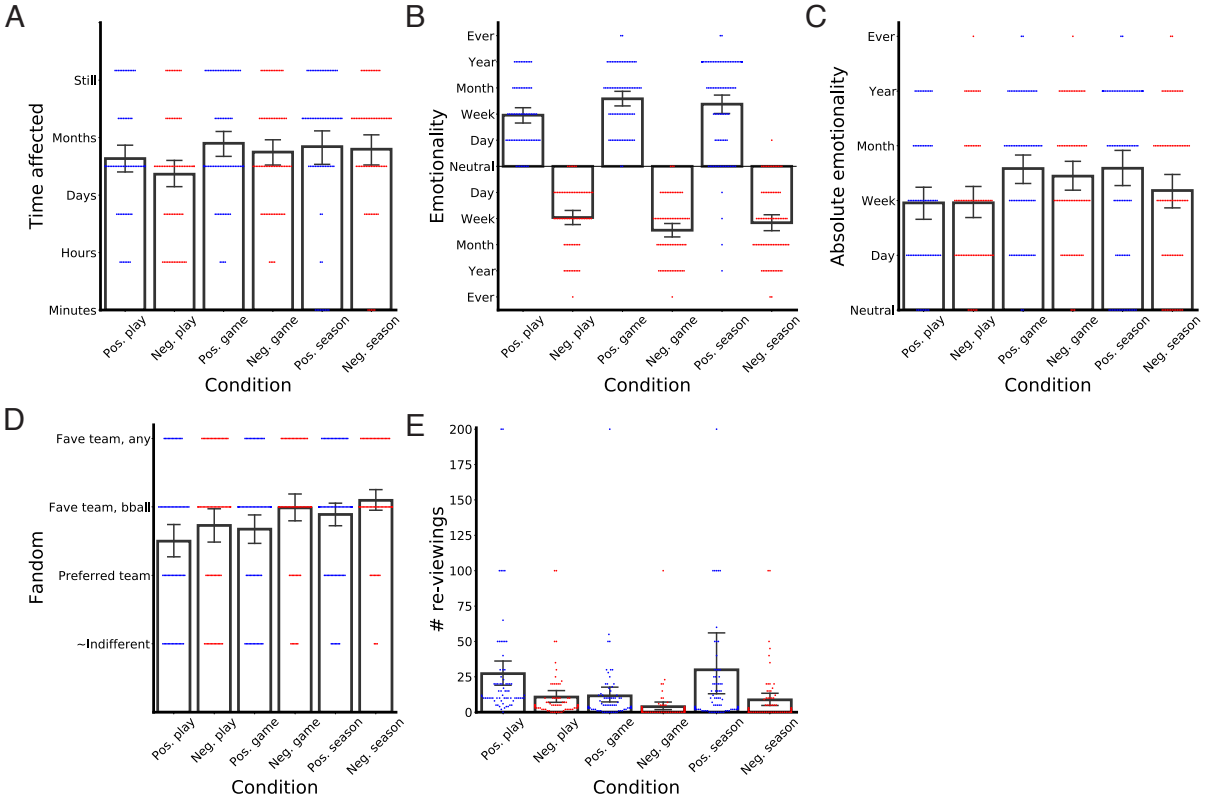

Figure S8: Characteristics of fan memories. We plotted metrics for each type of memory on (A) how long fans were affected by the outcome (1=minutes, 2=hours, 3=days, 4=months, 5=still affected), (B) emotionality with respect to everyday life (-5 to 5; worst to best thing that happened on a given day, week, month, year, ever), (C) unsigned emotionality, (D) their level of fandom (1=almost indifferent, 2=preferred team in that sport, 3=favorite team in that sport, 4=favorite team in any sport), and (E) approximate number of times they re-watched the play, parts of the game, or parts of the season.