Passage-of-Time Dysphoria: A Highly Replicable Decline in Mood During Rest and Simple Tasks that is Moderated by Depression

David C. Jangraw[1,2,+], Hanna Keren[1], Haorui Sun[2], Rachel L. Bedder[3,4], Robb B. Rutledge[3,4,5], Francisco Pereira[1], Adam G. Thomas[1], Daniel S. Pine[1], Charles Zheng[1], Dylan M. Nielson[1*], and Argyris Stringaris[1,6*]

[1] *National Institute of Mental Health, Bethesda, MD*
[2] *Department of Electrical and Biomedical Engineering, University of Vermont, Burlington, VT*
[3] *Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, UK*
[4] *Wellcome Centre for Human Neuroimaging, University College London, London, UK*
[5] *Departments of Psychology and Psychiatry, Yale University, New Haven, CT*
[6] *Division of Psychiatry, University College London, London, UK*
[*] *These authors contributed equally to this work*
[+] *Corresponding author: djangraw@uvm.edu*

Abstract

Does our mood change as time passes, and is this change different in people with depression? These questions are central to affective neuroscience theory and methodology, yet they remain largely unexamined. Here we demonstrate that rest periods lowered participants' mood, an effect we call "passage-of-time dysphoria." This finding was replicated in 19 cohorts totaling 28,482 adult and adolescent participants. The dysphoria was (1) relatively large (13.8% after 7.3 minutes, Cohen's $d = 0.574$), (2) variable across and within individuals but consistent across cohorts, and (3) present during simple visuomotor and gambling tasks. Rest also impacted behaviour: participants were less likely to gamble at the beginning of a task if it was preceded by rest. The dysphoria was inversely related to depression risk and a computationally estimated reward sensitivity parameter. Our results have theoretical implications for the nature of mood and its aberrations, and methodological consequences for the design and interpretation of experiments.

Introduction

Is human mood sensitive to the passage of time? Does our mood drift when we are resting or performing a simple task? In this paper, we answer these questions through a series of experiments across a large range of cohorts and introduce the term "passage-of-time dysphoria" to describe a generalizable and robust decline in mood during rest and simple tasks. This finding has both theoretical and computational implications for mood and psychopathology, and it represents a potentially major methodological confound in affective neuroscience experiments.

An important but typically implicit notion amongst affective neuroscientists is that each participant has a baseline mood or emotional state that will remain constant during an experiment or only vary with emotionally salient events (Penny et al., 2011). This notion of a constant affective background without a spontaneous mood drift – and, critically, without inter-individual variability in mood drift – has serious implications for both mood theory and the analysis and interpretation experiments that are central to affective neuroscience.

In current theoretical and computational accounts, mood is thought to reflect the integration of events –

typically as a discounted sum of rewards and punishments (Keren et al., 2021) or prediction errors (Rutledge et al., 2014) – over time. These models imply the constant affective background assumption in that they hold that the time scale over which these events unfold is irrelevant and that the passage of time itself has no effect on mood. Whilst convenient, this assumption seems naïve in the face of evidence about boredom, mind-wandering, and related affective phenomena (Miner and Glomb, 2010; van Hooff and van Hooft, 2014; Killingsworth and Gilbert, 2010; Robison et al., 2020; Agrawal et al., 2020) that arise when humans engage in tasks or rest. Critical to such phenomena is the passage of time: tasks have to be performed for "sufficiently long periods of time" in order to generate negative affect. Perhaps more generally, the passage of time is fundamental to the explore/exploit question that has been preoccupying neuroscience in recent years (Addicott et al., 2017; Cohen et al., 2007) and concerns the timing at which an agent will switch from their current environment (exploit) to a different task (explore). It is currently thought that negative affective states (such as boredom) building over time provide the subjective motivation to switch from a current to a different task (Geana et al., 2016; Agrawal et al., 2020). Critically, individuals vary in their propensity for such negative affect (Vodanovich et al., 1991) and the likelihood of switching tasks. Given these findings, it is surprising that no attempt has been made to quantify and model the effects of the passage of time on mood.

Beyond its potential importance in understanding mood, taking into account the passage of time has profound methodological implications for two kinds of experimental approaches that neuroscientists conduct. The first class of experiments is "resting state," exemplified by a functional brain scan in which a participant is asked to stare at a fixation cross. Based on the constant affective background assumption, neuroscientists pool resting-state data across individuals and treat these as indicative of underlying traits (as opposed to reflections of variability in response to experimentally imposed rest periods). For example, comparisons of resting-state neuroimaging data between depressed and non-depressed participants are thought to reveal differences in their task-general traits, rather than reflect their state-dependent response to rest periods. The second class of affective experiments is the event-related design, such as a gambling or face recognition task, during which participants experience stimuli (wins or losses) that elicit emotional reactions. Owing to the constant affective background assumption, neuroscientists can conduct event-related analyses, where responses to task stimuli are thought to occur on top of (and are often contrasted to) the affective baseline, which is presumed to be time-invariant.

These methodological assumptions are particularly questionable in light of evidence that spontaneous affective changes vary systematically between the individuals and groups being compared in affective neuroscience. For example, spontaneous negative thoughts are known to occur and vary substantially between humans, as highlighted by extensive work in mind-wandering (Robison et al., 2020; Killingsworth and Gilbert, 2010). Similarly, it is well known from occupational psychology that periods of low or relatively constant stimulation (as occurs in rest or more repetitive experimental tasks) can induce varying levels of boredom (van Hooff and van Hooft, 2014; Miner and Glomb, 2010). These insights from cognitive neuroscience raise the possibility that mood states will follow a similar pattern of inter-individual variability. But the size, stability, and clinical correlates of this variability remain unexplored.

In order to answer these fundamental questions in the understanding of mood and the methodology of affective neuroscience, we examine how the passage of time affects mood in a variety of experiments across studies, participants, and settings. Central to our study is self-reported mood, assessed in keeping with clinical (Costello and Angold, 1988), epidemiological (Pavot and Diener, 1993), and psychological research (notably in ecological momentary assessment (Ebner-Priemer and Trull, 2009)). Our study was initially motivated by our serendipitous finding that participants' self-reported momentary mood worsened considerably during a period of rest (i.e., when participants were asked to passively view a fixation cross for several minutes). We call this effect "passage-of-time dysphoria" (using *dysphoria* as a term for general unease rather than implying any clinical condition). Prompted by this finding, we examined the effects of resting state and simple tasks on self-reported momentary mood in numerous task variations completed by large and varied cohorts of participants on their home computers, including over 100 healthy and depressed adolescents recruited in person and over 1900 adults recruited online from across the United States. Across experimental setups, cohorts, and age groups, we found a robust negative relationship between time spent on task and self-reported mood. We then replicated these findings in a gambling task dataset with over 20,000 participants. Moreover, we conducted a series of longitudinal studies to characterise the effect's consistency within cohorts and its variability within individuals. Crucially, we extend our studies to include clinical populations: we show that passage-of-time dysphoria is reduced in depressed participants and that this effect may be related to reduced reward sensitivity. Finally, we demonstrate that passage-of-time dysphoria is not accounted for by existing terms, such as boredom or mind-wandering.

Overall, our findings indicate that passage-of-time dysphoria is a highly replicable, common phenomenon of relatively large magnitude that seems reduced in depressed individuals. This can have profound implications for experimental design and interpretation in affective neuroscience.

Results

Characterising the Effect

The results to follow characterise the average person's gradual decline in mood during rest and simple tasks, a phenomenon we call "passage-of-time dysphoria". This effect was initially observed in a task where participants were periodically asked to rate their mood. Between these mood ratings, the initial cohort was first asked to stare at a central fixation cross. They were told that the rest period would last up to 7 minutes and that they would be asked to rate their mood "every once in a while". The mood ratings observed during this rest period inspired a number of slightly modified tasks to better characterise the effect and eliminate methodological confounds. Each modification was presented to a new cohort of naïve participants so that memory and expectations would not affect their mood ratings. Each cohort also played a gambling game at some point in the task, which was included to observe the effects of rest on rational behaviour, to maintain links with previous studies of mood and reward (Keren et al., 2021), and to enable related analyses on a large cohort of participants (n=26,896) playing a similar game on their smartphones (Bedder et al., 2020). A full list of the cohorts examined in this study is shown in Supplementary Table 1. The motivation for, and results of, various cohorts and combinations of cohorts are explained in the following sections.

In order to quantify the effects of rest from the full set of mood ratings, we created a linear mixed effects (LME) model with terms for initial mood and mood slope (i.e., change in mood per unit time) as random effects that were fitted to each subject's data. The factors of interest described in the following sections were included in the model as fixed effects. This model is described in detail in the Methods section. One factor of particular interest is a depression risk score for each participant, a continuous value defined as their score on the Mood and Feelings Questionnaire (MFQ, for adolescents) or the Center for Epidemiologic Studies Depression Scale (CES-D, for adults) divided by a clinical cutoff, i.e., MFQ/12 or CES-D/16. The model was fitted to the cohort of all participants who experienced an opening period of rest, visuomotor task, or random gambling. The slope parameter learned for each participant was used to quantify that participant's passage-of-time dysphoria.

In a related analysis, a large cohort of participants playing a similar gambling game on their smartphones were also evaluated for the presence of a mood slope related to time on task. The LME model described above was also fitted to this cohort. And because this was a cohort large enough to fit hyperparameters in a held-out set of participants, this cohort's mood ratings were also fitted to a computational model that estimates each participant's initial mood and their sensitivity to rewards, reward prediction, and time (See Methods section for a full description). Regularization hyperparameters were selected to make a model that was fitted on the first 10 mood ratings best match the final two mood ratings in an exploratory cohort. The model's time sensitivity parameter for each participant was used to quantify their passage-of-time dysphoria.

Passage-of-Time Dysphoria Is Sizeable

In our initial cohort (called 15sRestBetween in Supplementary Table 1) of 40 adults recruited on Amazon Mechanical Turk (MTurk), we asked whether mood would change consistently during a rest period that preceded a gambling game. We observed a gradual decline in mood over time (Figure 1A, blue line). After 9.7 minutes of rest, the change in mood was of considerable size ($Mean \pm SE = 22.4\% \pm 4.15\%$ of the mood scale). This result was later replicated in 5 other adult MTurk cohorts that received shorter opening rest periods (Figure 1A, other lines).

Passage-of-Time Dysphoria Is Robust to Methodological Choices

Because this finding is new, we wanted to examine the impact of possible methodological confounds. We therefore created slightly modified versions of the task to see whether the observed decline in mood ratings might be due to:

1. The aversive nature of rating one's mood
2. The method of rating mood and its susceptibility to fatigue
3. The expected duration of the rest period
4. The aversive nature of multitasking or task switching
5. Regression to the mean

The impact of ratings on mood was investigated by systematically varying the frequency of mood ratings. More frequent ratings did not lead to a more rapidly declining mood. The impact of fatigue on mood ratings was investigated by making every mood rating require an equally easy single keypress. This did not change the decline in mood. The results of these and other control experiments and analyses suggested that the observed dysphoria cannot be explained by the above list of factors (see Supplementary section titled "Eliminating Methodological Confounds").

Passage-of-Time Dysphoria Also Occurs During Tasks

To see whether this decline was specific to rest or more generally linked to time on task, we administered two variants of the task. The first variant (cohort Visuomotor-Feedback, $n = 30$) was designed to mimic rest very closely while requiring the participant to respond regularly and giving feedback on their performance. Specifically, a fixation cross moved back and forth periodically across the screen, the participant was asked to press a button whenever it crossed the centerline, and each response would make the cross turn green if the response was accurate or red if it was too early or late (see Methods Section). In the second variant (cohort Daily-Random-01, $n = 66$), the subject played a random gambling game in which gambling outcomes and reward prediction errors (RPEs) were both random with mean zero. Both of these tasks produced similar mood timecourses, and LME slope parameters were not significantly different from those of the original cohort ($t_{68} = 0.437, p = 0.663$ for visuomotor task, $t_{104} = 1.07, p = 0.287$ for random gambling) (Figure 1B).

Passage-of-Time Dysphoria Is Generalizable

Having replicated the finding, we investigated the generalizability of this result across age groups and recruitment methods. MTurk participants have an expectation of earning money quickly, and any time spent at rest could have been spent doing other jobs. They might therefore feel more acutely the opportunity cost that comes with time spent at rest (Kahneman, 1973). To test the generalizability of the effect, we collected similar rest + gambling data via an online task from adolescent participants recruited in person at the National Institute of Mental Health in Bethesda, MD and asked to complete the task online via their home computers (see Methods Section for details). This group (Adolescent-01, n=116) showed a pattern of declining mood not significantly different from that observed in the MTurk cohort (combining across all MTurk participants that received opening rest or visuomotor task periods, n=637) (Figure 1C) ($Adolescent\ Mean \pm SE = 8.1\% \pm 1.32\%, MTurk\ Mean \pm SE = 10.58\% \pm 0.83\%, t_{751} = 1.22, p = 0.221$).

To more precisely characterise the effect, we fitted a large LME model to the complete cohort of online participants (both adults and adolescents) completing rest or simple tasks (including random gambling) in the first block. These results can be seen in Supplementary Table 2. The slope parameter (rate of mood decline with time) for these participants was $Mean \pm SE = -1.89 \pm 0.185\ \%mood/min$, which was significantly

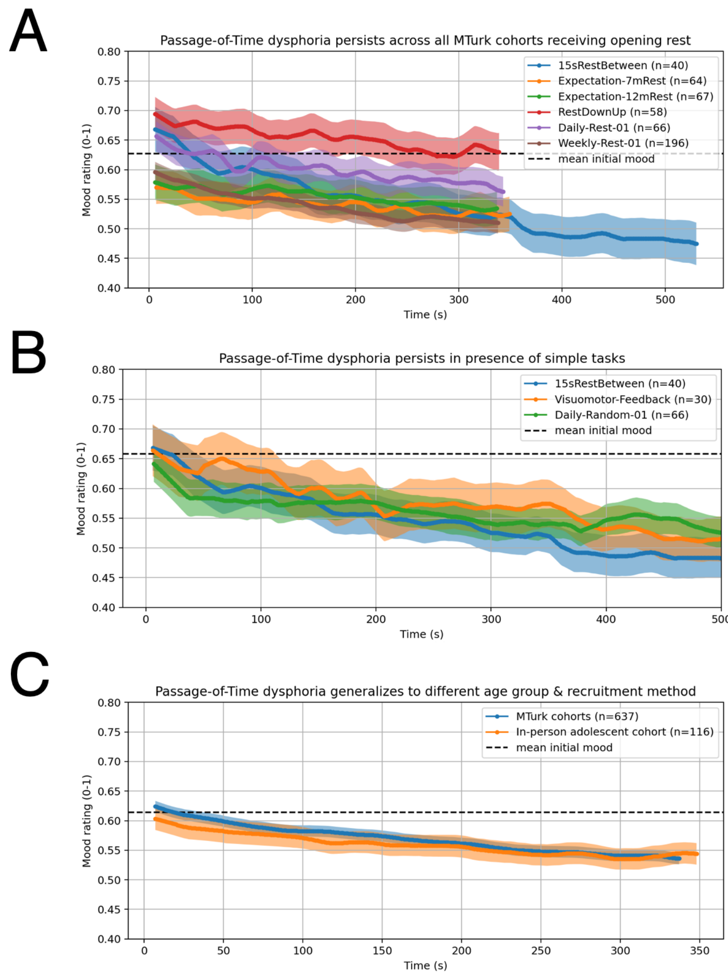

Figure 1: The timecourse of passage-of-time dysphoria is consistently present across many cohorts and task modulations. These plots each show the mean timecourse of mood across participants in various online cohorts for the first block of the task. Each participants mood between ratings was linearly interpolated before averaging across participants. The shading around each line represents the standard error of the mean. Each name in the legend corresponds to a cohort completing a slightly different task (refer to Supplementary Table 1 for details). Mean initial mood refers to the mean of cohort means, not the mean of subject means. (A) Mean timecourse of mood ratings during an opening rest period in all Amazon Mechanical Turk (MTurk) cohorts that received it. Passage-of-time dysphoria was discovered in one cohort (blue line) and replicated in five independent naïve cohorts. (B) The dysphoria was observed not only in rest periods (blue), but also in a simple task requiring action and giving feedback (orange), and in a random gambling task with 0-mean reward prediction errors and winnings (green). (C) The dysphoric effect of time was observed both in adults recruited on MTurk (blue) and in adolescents recruited in person (orange).

less than 0 ($t_{864} = -10.3, p < 10^{-6}$. After 7.3 minutes (the mean duration of the first block of trials), the mean decrease in mood estimated by this LME model was 13.8% of the mood scale. This corresponds with a Cohen's $d = 0.574$, with a 95% CI = (0.464, 0.684), calculated as described in (Feingold, 2015).

Passage-of-Time Dysphoria Is Present but Diminished in a Mobile App Gambling Game

The previous line of analysis suggested that time spent on conventional neuropsychological tasks may also produce dysphoric effects. We tested whether this could be observed in a large dataset ($n = 26,896$) of mood ratings during a similar gambling task played on a mobile app. All analyses were applied to an exploratory cohort of 5,000 of these participants, then re-applied to the confirmatory cohort of all remaining participants after preregistration. We applied the LME modeling procedure to this confirmatory cohort (4.65% of participants were excluded for being outliers in their average response time, leaving $n = 20,877$) and again found a slope parameter that was significantly below zero at the group level ($Mean \pm SE = -0.881 \pm 0.0613$ $\%mood/min, t_{22804} = -14.4, p < 10^{-6}$).

It is notable that even in this relatively engaging game (in which tens of thousands of participants completed the task despite not being paid for participating or penalised for failing to finish), mood tended to decrease with time spent on task.

We note, however, that the dysphoric effect of time-on-task was significantly smaller in this cohort than in the combined cohort of online participants (2-sided Wilcoxon rank-sum test, $W_{20876} = -14.5, p = 8.10 * 10^{-45}$). 87.5% of online participants had negative slopes in the LME analysis, but only 70.2% of mobile app participants did. A histogram of the LME slope parameters for online and mobile app participants is plotted in Figure 2. This shows that, as one might expect, passage-of-time dysphoria is sensitive to task context.

Next, to disentangle the effects of mood from the effects of reward and reward prediction error in this dataset, we fitted the computational model described in the Methods Section to the mobile app data. Including the mood slope parameter in the model decreased the mean squared error on testing data (the last two mood ratings of the task) from 0.336% to 0.325% of the mood scale for the median subject across regularizations, a significant improvement (2-sided Wilcoxon signed-rank test, $W_{499} = 0, p = 1.55 * 10^{-23}$). This suggests that time on task affected a participant's mood beyond the impacts of reward and expectation, and did so in a way that was stable within individuals because improved fits were observed in held-out data. Fits and parameter distributions can be seen in Supplementary Figures 7 and 8. The distribution of participants' time sensitivity parameters β_T (which can be interpreted as the mood slope independent of reward effects) was again centered significantly below zero ($Mean \pm SE = -0.128 \pm 0.00668$ $\%mood/min$, 2-sided Wilcoxon signed-rank test $W_{21895} = 1.00 * 10^8, p = 6.30 * 10^{-90}$).

Passage-of-Time Dysphoria Is Not Present in Freely Chosen Activities

After the surprising finding that passage-of-time dysphoria appeared during an engaging mobile app game, we wondered whether this phenomenon would be observed in daily life, outside the context of a psychological task. We therefore designed and preregistered (https://osf.io/gt7a8) a task in which the initial rest period was replaced with 7 minutes of free time, during which the participant could pursue activities of their choice. Participants completing this task (cohort Activities, n=450) were asked to rate their mood just before and just after the break period. They were then asked to report what they did during that period by rating 27 activities on a 5-point scale from "Not at all" (scored at 0%) to "The Whole Time" (scored at 100%). The most frequent activities reported were thinking (mean 50.2%), reading the news (28.2%), and standing up (26.2%). The rest were performed for less than a quarter of the average break period (see Supplementary Materials Table 3). Those who reported thinking also reported other activities; most participants apparently used this response to indicate not exclusively sitting and thinking, but rather thinking about the things they were doing.

This group was the first to not exhibit passage-of-time dysphoria. The mood ratings just after the free period were not statistically different from the mood ratings before the free period (Mean pre-break mood:

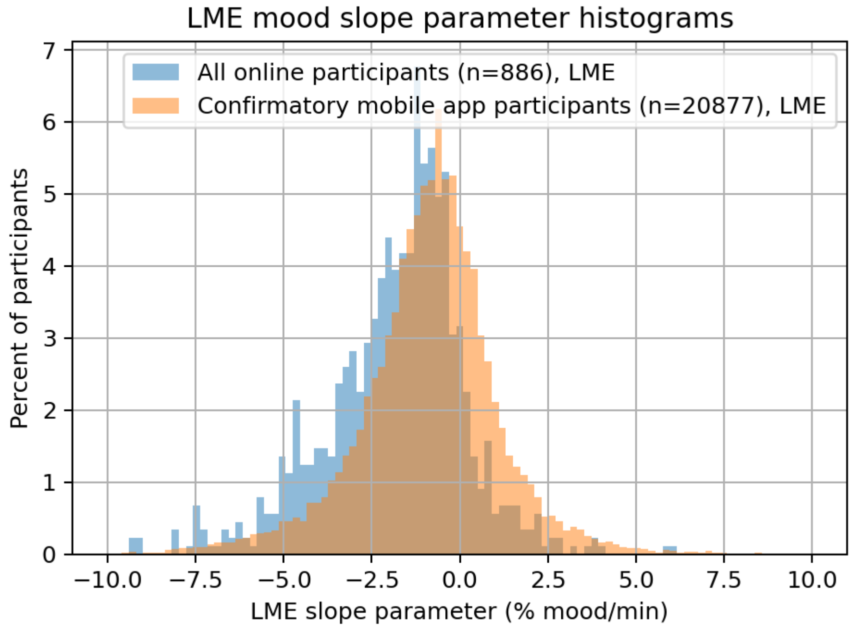

Figure 2: Individual subject LME slope parameters for online participants (blue) and mobile app participants (orange).

65.7%, post-break mood: 66.6%, change in mood: $0.13\%/min$, difference $t_{449} = -1.33$, $p_{H0:decrease} = 0.0918$, $p_{H0:increase} = 0.908$). The change in mood was significantly lower (i.e., more negative) for a cohort who received the standard rest period with interspersed mood ratings (cohort BoredomAfterOnly, n=150) ($t_{598} = 6.28$, $p = 3.23e-10$). This shows that, perhaps unsurprisingly, passage-of-time dysphoria is not universal to all activities. However, the nominal increase in mood during this period (0.130% mood/min) was much smaller than the decrease in mood observed during a typical rest period in our task (-1.89% mood/min). Put another way, each minute in which participants could choose their activity raised their collective mood less than 10% of the mood decline experienced during a minute of rest.

Inter-Individual Differences

While the results above make it clear that the group average slope of mood is negative during rest and simple tasks, there is considerable variation across participants. The range between the 2.5th and 97.5th percentile of subject-level slope estimates for online participants is -7.23 to $1.79\%mood/min$. (The full distribution of slope estimates can be seen in Figure 2.) This means that different individuals can have very different effects of time spent at rest on their mood. Using cohorts that completed the task more than once, we found that these individual differences had significant stability across blocks, days, and weeks (see Supplementary Materials section "Stability Over Time"). We next investigated the relationship between this variability and other traits of clinical and theoretical interest.

Passage-of-Time Dysphoria Is Inversely Related to Depression Risk

First, we investigated whether the slope of a participant's mood correlates with trait-level depressive characteristics. In our online participant LME model, higher depression risk score was significantly associated with lower initial mood ($Mean \pm SE = -18.6 \pm 0.8\%mood$, $t_{877} = -22.4, p < 10^{-6}$) and less negative mood slope (depression-risk * time interaction, $Mean \pm SE = 0.515 \pm 0.109\%mood/min$, $t_{869} = 4.75, p < 10^{-6}$).

This relationship is visually characterised in several ways in figure 3. This includes a comparison of the average mood timecourse of participants at risk of depression and not at risk (A), a scatter plot of individuals' mood slopes against their depression risk (B), and a plot of the proportion of participants at risk of depression who showed significantly positive and negative mood slopes (C). Each analysis supports the relationship between mood slope and trait-level depression.

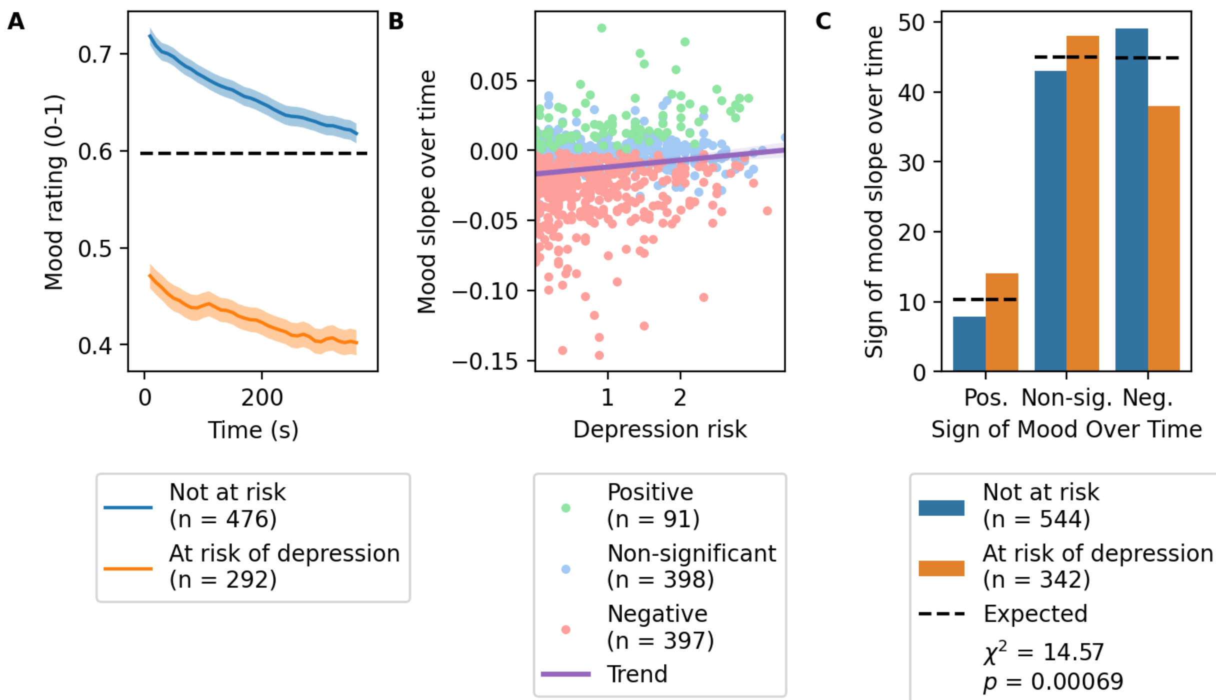

Figure 3: Relationship between passage-of-time dysphoria and depression risk. (A)Mood ratings over time of online participants at risk of depression (defined as MFQ>12 or CES-D>16) vs. those not at risk for the 768 participants with at least 6 minutes of resting mood data (error bars are SEM). The dotted line represents the mean initial rating (mean of cohort means). (B) We fitted simple regressions of time versus mood within each individual and determined significance of the time term with Benjamini-Hochberg false-discovery rate correction ($\alpha = 0.5,\ p < 0.05$) to better understand the relationship between depression risk and the change in mood over time. Depression risk is operationalised as score on the CES-D or MFQ divided by the threshold for depression risk on each measure (16 and 12 respectively). (C) Proportion of individuals with or without risk of depression (i.e., depression risk >1 or <1) with positive (significantly greater than zero), non-significant (not significantly different than zero), and negative (significantly less than 0) slopes of mood over time. 13 more individuals at risk of depression have a positive slope than the 35 expected based on the rates in individuals not at risk of depression.

The inverse relationship between depression risk and mood slope was later replicated in our follow-up cohorts that received boredom or mind-wandering survey questions before and/or after a rest period interspersed with mood ratings. We combined these new cohorts (i.e., cohorts MwBeforeAndAfter, MwAfterOnly, BoredomBeforeAndAfter, and BoredomAfterOnly, n=600) and ran the same linear mixed effects model. As before, a higher depression risk score was significantly associated with lower initial mood ($Mean \pm SE = -18.1 \pm 0.9\% mood$, $t_{593} = -20.3, p < 10^{-6}$) and less negative mood slope (depression-risk * time interaction, $Mean \pm SE = 0.510 \pm 0.140\% mood/min$, $t_{594} = 3.64, p = 2.93e-4$).

This relationship was also observed in the mobile app cohort. Using each participant's life happiness rating as a proxy for (lack of) depression risk, we found a significant negative correlation between life happiness and β_T ($r_s = -0.0658, p = 1.83 * 10^{-22}$). β_T is plotted against life happiness in Figure 4 (left).

We took care to examine the possibility that floor effects were driving these results. On average, individuals reporting greater depressive symptoms reported lower initial mood at the onset of the task. If their mood declined further, they therefore had less of the mood scale available to them to express it. This could lead to "floor effects" where the mood of depressed individuals appears to decline more slowly with time simply because they have reached the bottom of the scale and are forced to level out. To address this possibility, we performed sensitivity analyses in which participants reaching either an absolute or individual mood floor were excluded. In both cases, the interaction effect of depression risk and time remained significant (see Supplementary Materials). The categorical finding that an out-sized proportion of depressed individuals had a positive mood slope ($\chi^2 = 14.57, p = 6.9x10^{-4}$, Figure 3C) further refutes this notion.

Passage-of-Time Dysphoria Is Associated with Sensitivity to Rewards

Depression risk level is also thought to relate etiologically to reward valuation, which prompted us to investigate the relationship between participants' time sensitivity, reward sensitivity, and life happiness in our computational model fits. The time sensitivity parameter β_T had a significant negative correlation with the reward sensitivity parameter β_A, which was fitted simultaneously ($r_s = -0.106, p = 1.73 * 10^{-55}$) (Figure 4, middle). This anticorrelation was weaker in participants with life happiness below the median (i.e., those at greater risk of depression) than it was in those with life happiness greater than or equal to the median ($Z = 6.41, p = 7.13 * 10^{-11}$) (Figure 4, right). This suggests that people more sensitive to the passage of time are also more sensitive to rewards, and that this relationship is less pronounced in those with greater depression risk. Taken together, these results demonstrate relationships between passage-of-time dysphoria and other important individual differences.

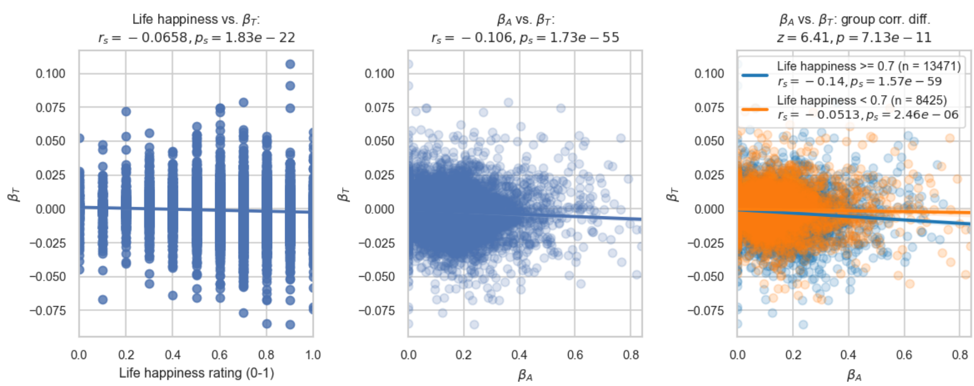

Figure 4: Individual differences in sensitivity to the passage of time relate to other individual differences in the mobile app cohort. The computational model's time sensitivity parameter β_T for each participant in the mobile app cohort is plotted against that participant's life happiness rating (left) and their reward sensitivity parameter β_A (middle). When grouped by life happiness, participants with happiness at or above the median had a stronger $\beta_T - \beta_A$ anticorrelation than participants with happiness below the median (right)

Impact on Behaviour

Participants Receiving Rest Periods Are Less Likely to Gamble

Results thus far have established passage-of-time dysphoria's relationship with individual differences that involve mood directly. To investigate whether rest's effects are limited to subjective mood, we examined the impact of rest and dysphoria on behaviour in the gambling tasks. We observed that gambling (specifically positive closed-loop gambling, in which participants tended to receive positive RPEs) participants who had a preceding rest or visuomotor task block had significantly lower mood at gambling onset than those who did not (2-sided Wilcoxon rank-sum test, $W_{722} = 5.13, p = 2.87 * 10^{-7}$) (Figure 5, top). This effect was no longer significant at the next mood rating, which took place around trial 4 of gambling. We therefore examined gambling behaviour in these first 4 trials. Those who had experienced either a short (350-450 s) or long (500-700 s) opening rest period were significantly less likely to gamble than those who had not (2-sided Wilcoxon rank-sum test, no rest vs short rest: $W_{469} = 4.85, p = 1.21 * 10^{-6}$; no rest vs long rest: $W_{344} = 4.79, p = 1.63 * 10^{-6}$; both less than 0.05/3 controlling for multiple comparisons). (Figure 5, bottom). The long and short rest groups, however, were not significantly different from each other ($W_{629} = 0.52, p = 0.603$). Examination of trial-wise gambling behaviour shows the difference between rest and no-rest groups is most pronounced in the first four trials, much like the differences observed in mood (Figure 5, middle). However, no significant correlation was observed between an individual's mood slope parameter during the rest block and the number of times they chose to gamble in the first 4 trials ($r_s = 0.0317, p_s = 0.427$). An individual's mood at gambling onset, however, did correlate significantly (but weakly) with the choice to gamble in the mobile app cohort ($r_s = 0.0161$, $p_s = 0.0169$). This suggests that mood, rather than differences in mood's sensitivity to time, is most strongly associated with changes in gambling behaviour.

Relationship to Boredom and Mind-Wandering

It is possible that existing terms, such as boredom or mind-wandering (MW), could readily explain the phenomenon we describe in this study, making the introduction of the term "passage-of-time dysphoria" redundant. We conducted a series of preregistered (https://osf.io/gt7a8) experiments to investigate this possibility (see Supplementary Materials for concise results of all preregistered hypotheses). Boredom is typically defined as a state of "low arousal and dissatisfaction" (Mikulas and Vodanovich, 1993). Mind-wandering, often defined as task-unrelated or spontaneous thought (Mrazek et al., 2013; Christoff et al., 2016), tends to be unpleasant, particularly because the emotional content of that thought is disproportionately negative (Killingsworth and Gilbert, 2010; Poerio et al., 2013). In a preregistered data collection and analysis, we collected four new cohorts totaling $n = 600$ participants to examine the relationship between passage-of-time dysphoria and these more established constructs at the state level, state change level, or trait level. Participants were randomised to one of these cohorts (or to the Activities cohort described in a previous section) at the time of participation.

Passage-of-Time Dysphoria is Weakly Related to State Boredom

Two new cohorts were collected to quantify the degree to which passage-of-time dysphoria could be explained by boredom. Each received a rest period with mood ratings 20 seconds apart, followed by the Multidimensional State Boredom Scale's short form (MSBS-SF) (Hunter et al., 2016). The first (cohort BoredomBeforeAndAfter, $n = 150$) completed the MSBS-SF both before and after this rest period. The second (cohort BoredomAfterOnly, $n = 150$) completed the MSBS-SF only after this rest period. Both cohorts completed a survey that included the short boredom proneness scale (SBPS) to assess trait boredom (Struk et al., 2017). Using a one-sided t-test, we determined that repeated administration of the MSBS-SF did affect later responses: that is, participants who *were* asked about boredom before the rest period reported lower boredom after the rest period than those who were *not* asked about boredom before the rest period (Cohen's $d = -0.411$). Because we could not rule out the possibility of a large effect (H0: Cohen's $d < -0.5,\ t_{298} = 0.987,\ p = 0.163$), we did not combine across the two cohorts in subsequent analyses.

We used the BoredomAfterOnly cohort to examine our first boredom-related hypothesis: that final state boredom reported after the rest period explains variance in subject-level passage-of-time dysphoria slope. Results showed that a model including state boredom explained additional variance beyond one excluding it ($\chi^2(2, N = 16) = 8.769,\ p = 0.0125$). But the effect of final state boredom's inclusion on model fit was quite small: the variance explained by the fixed effects in the model increased from $R^2 = 0.341$ to $R^2 = 0.359$

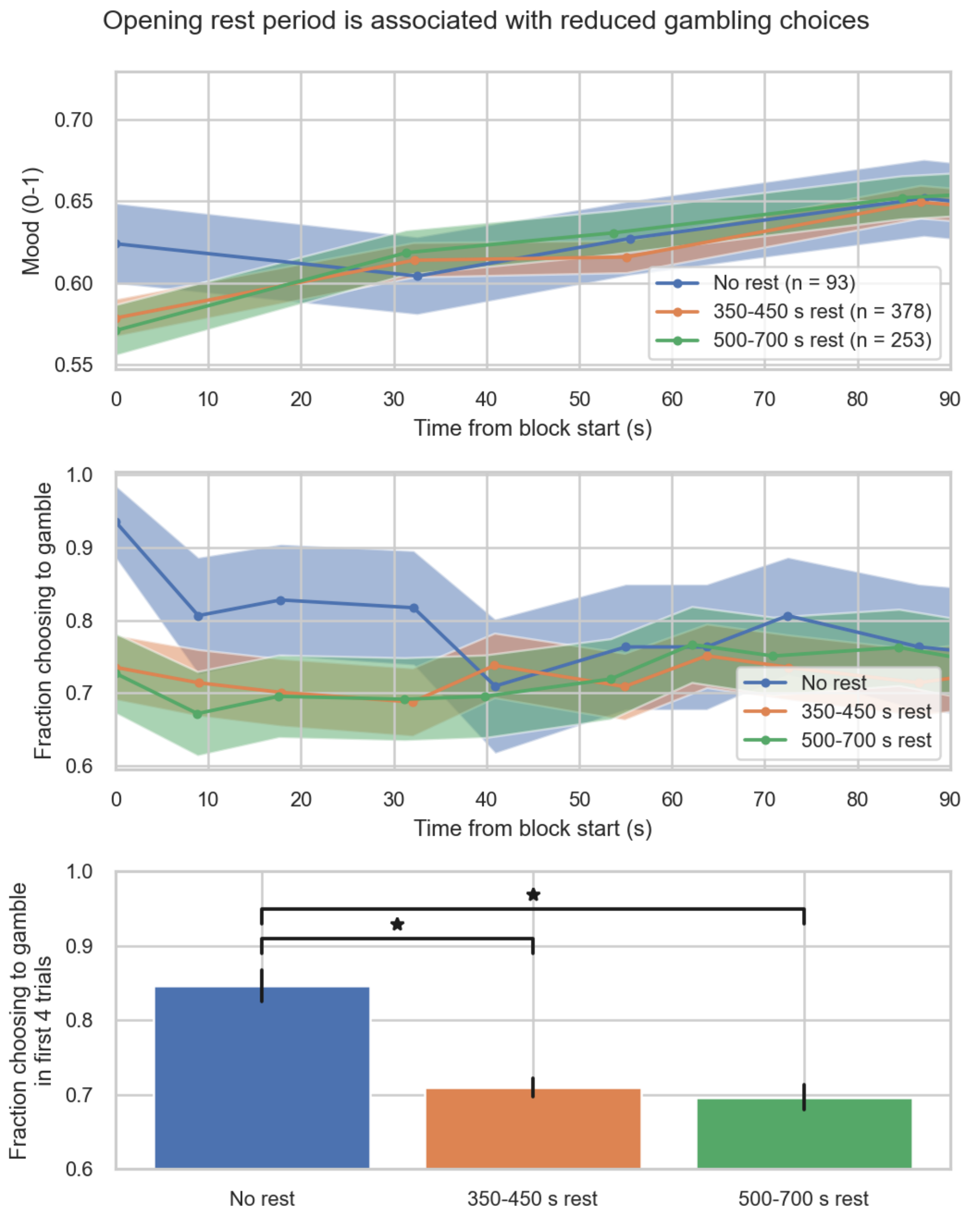

Figure 5: Rest periods decreased the likelihood of choosing to gamble in the first 4 trials after rest ended. Top: mean ± standard error mood ratings across participants in their first block of (positive closed-loop) gambling preceded by different rest period durations. Middle: fraction of participants in each group that chose to gamble on each trial of this first gambling block (error patches are 95 percent confidence intervals derived from a binomial distribution). Bottom: mean ± standard error across participants of the fraction of the first 4 trials of this first gambling block that participants chose to gamble. Stars indicate that a pair of groups was significantly different (2-sided Wilcoxon rank-sum test, $p < 0.05/3$ to correct for multiple comparisons).

($F^2 = 0.0283$).

We next used the BoredomBeforeAndAfter cohort to examine our second hypothesis: that the *change* in state boredom reported before and after the rest block explains variance in subject-level passage-of-time dysphoria slope. Again, results showed that a model including the change in state boredom explained additional variance ($\chi^2(2, N = 16) = 18.6,\ p = 9e-4$). But the effect of boredom change's inclusion on model fit was similarly small: the variance explained by the fixed effects in the model increased from $R^2 = 0.352$ to $R^2 = 0.392$ ($F^2 = 0.0671$).

Finally, we used the BoredomAfterOnly cohort to examine our third hypothesis: that trait boredom explains variance in subject-level passage-of-time dysphoria slope. These results showed that a model including trait boredom failed to explain significant additional variance ($\chi^2(2, N = 16) = 2.37,\ p = 0.305$).

Passage-of-Time Dysphoria is Weakly Related to Mind-Wandering

Two new cohorts were collected to quantify the degree to which passage-of-time dysphoria could be explained by mind-wandering (particularly MW with negative emotional content). Each received a rest period with mood ratings 20 seconds apart, followed by a 13-item Multidimensional Experience Sampling (MDES) as described in Turnbull et al. (2019). The first (cohort MwBeforeAndAfter, $n = 150$) completed the MDES only after this rest period. The second (cohort BoredomAfterOnly, $n = 150$) completed the MDES only after this rest period. As described in Ho et al. (2020), we applied principal components analysis (PCA) on participants' MDES responses to find a component whose primary loading was on the "emotion" item (in which they reported their thoughts as being negative or positive). The "emotion dimension" of each MDES response was then quantified as the amplitude of this component. The sign of PCA components is not meaningful, so we arbitrarily chose that increased emotion dimension would represent *more negative* thoughts. Both cohorts completed a survey that included the 5-item mind-wandering questionnaire (MWQ), which quantifies a person's proneness to mind-wandering without regard to the valence of those spontaneous thoughts (Mrazek et al., 2013). Using two one-sided t-tests, we determined that repeated administration of the MDES did not affect later responses in the emotion dimension: that is, participants did not report different emotional valences after the rest period if they were also asked about their thoughts before the rest period (Cohen's $d = 0.0739$; $H0 : d < -0.5 :\ t_{298} = 7.52,\ p = 2.34e-12$; $H0 : d > 0.5 :\ t_{298} = 5.58,\ p = 5.39e-8$).

We used both the MwBeforeAndAfter and MwAfterOnly cohorts to examine our first MW-related hypothesis: that the final emotion dimension reported after the rest period explains variance in subject-level passage-of-time dysphoria slope. Results showed that a model including final emotion dimension explained additional variance beyond one excluding it ($\chi^2(2, N = 16) = 44.0,\ p = 2.77e-10$). The effect of final emotional dimension's inclusion on model fit was larger than boredom but still modest: the variance explained by the fixed effects in the model increased from $R^2 = 0.275$ to $R^2 = 0.351$ ($F^2 = 0.116$).

We next used the MwBeforeAndAfter cohort to examine our second hypothesis: that the *change* in the emotional valence of thought reported before and after the rest block explains variance in subject-level passage-of-time dysphoria slope. Results showed that a model including change in emotion dimension explained additional variance beyond one excluding it ($\chi^2(2, N = 16) = 7.30,\ p = 0.026$). The effect of change in emotion dimension's inclusion on model fit was small: the variance explained by the fixed effects in the model increased from $R^2 = 0.300$ to $R^2 = 0.312$ ($F^2 = 0.017$).

Finally, we used both the MwBeforeAndAfter and MwAfterOnly cohorts to examine our third hypothesis: that trait MW explains variance in subject-level passage-of-time dysphoria slope. This time, results showed that a model including trait MW did *not* explain significant additional variance ($\chi^2(2, N = 16) = 1.20,\ p = 0.548$). This is perhaps not surprising given past work reporting that MW itself is not aversive, but the negative affective content of MW thought is (Poerio et al., 2013).

Discussion

In this study, we describe the discovery of a highly replicable and large effect which we call passage-of-time dysphoria. Our findings suggest that passage-of-time-dysphoria is distinct from existing constructs such as boredom or mind-wandering. These results call into question the long-held constant affective background assumption in affective neuroscience by demonstrating spontaneous, seemingly linear drifts in mood that occur over time. Our results have implications for the nature of mood and its modeling, and they also have a number of methodological consequences for affective neuroscience experimentation.

Our findings show that participants incorporate information about the passage of time into their ratings of mood. Standard theory holds that self-reports of mood represent an integration of the values of events (rewards and punishments) in the environment, and recent work has demonstrated that mood can be modelled as a weighted average of environmental events (Keren et al., 2021). We provide substantial empirical evidence that human mood is sensitive to the passage of time and that accounting for it improves models of mood. Mood's sensitivity to the passage of time is a long-intuited phenomenon that is widely acknowledged in literature (Nunokawa, 1996; Shattuck, 2001; Proust, 2013) and philosophy (Ciocan, 2010; Ratcliffe, 2013; Heidegger, 1995). It is also relevant to notions of boredom (Raffaelli et al., 2018), temporal discounting theories (Pulcu et al., 2014), and the possible consequences of mind-wandering (Wilson et al., 2014; Killingsworth and Gilbert, 2010). Our results provide robust empirical evidence for this phenomenon and reveal its temporal shape, its variability across individuals, and its level of stability. We also provide evidence that this temporal sensitivity is not a redundant reflection of existing constructs such as boredom, mind-wandering, or the valence of ongoing thoughts: each of these factors explained a small portion of the variance explained by our LME model of mood.

The mechanism that enables humans to be sensitive to the passage of time is not yet known. One possibility is that humans store expectations about the rate of rewards and punishments in the environment and that prolonged periods of monotony violate such expectations. Such a view aligns with the recently articulated theoretical progress in integrating opportunity cost across time, and therefore in estimating the benefits and downsides of delaying opportunity for reward (Agrawal et al., 2020). Momentary mood as measured in our study could function as a signal of opportunity cost. In that sense, an aversive momentary mood may be an adaptive signal that informs decisions to exploit (stay on task) or explore (switch task) (Geana et al., 2016).

Supporting a reward-based interpretation of our findings is our observation that participants at high risk of depression showed less passage-of-time dysphoria. This finding was replicated in multiple cohorts, with different depression measures and different types of participants. This would at first sight seem paradoxical since phenomena such as boredom have traditionally been linked to melancholia and depression (e.g., by Schoppenhaure and Kierkergaard). Yet, it has been argued cogently (Elpidorou, 2014) that such a view conflates negative affect as a trait (e.g., proneness to boredom) with negative affect as a state (a momentary experience). The latter is thought to motivate "the pursuit of a new goal when the current goal ceases to be satisfactory, attractive, or meaningful to the agent" (Elpidorou, 2014), and is as such a normative feeling that helps align ones activity with ones pursuit of rewards. Since valuation of reward is thought to be reduced in depression (Halahakoon et al., 2020), it is possible that misalignment with one's goals and violation of reward expectations—and resultant dysphoria—will, therefore, be less pronounced in depression. This interpretation is supported by our finding that passage-of-time dysphoria is less pronounced in those with lower reward sensitivity, as indexed by our computational parameter β_A. Moreover, we found that the relationship between reward sensitivity and passage-of-time dysphoria is moderated by depression level: the correlation between reward sensitivity and passage-of-time dysphoria (quantified as $-\beta_T$) is significantly larger in those with low depression (or higher life happiness) scores. It is tempting to speculate that reduced passage-of-time dysphoria could contribute to reduced motivation for action or environmental change in those with depression.

We found that mood declined during rest and multiple tasks (including a mobile app more engaging than most paradigms) but not freely chosen activities. This suggests that researchers are subjecting their participants to a somewhat unnatural stressor in their experiments without accounting for it in their analyses or interpretations. Our study's compelling evidence against the notion of a constant affective background suggests that certain methodological challenges have been overlooked in affective neuroscience. First, the inter- and intra-subject variability in the rate of mood changes over time will add to error variance if it is not accounted for in studies of mood or its correlates. Secondly, group differences in the rate of mood changes could lead to apparent group differences in later mood or behaviour. Perhaps most broadly applicable, changes in mood on the scale of tens of minutes prevent these longer blocks of time from being truly interchangeable. This means that changes to experimental procedures that might seem inconsequential could still introduce confounds. We will illustrate these ideas through some examples.

For example, let's consider a large collaborative study that is based on multisite imaging data collection, such as ENIGMA (Thompson et al., 2014). In this dataset, there is much variability across centres in the timing of the resting-state fMRI scan (e.g., the duration of the scan and whether it takes place at the start or end of the scan session) (Adhikari et al., 2019). This could lead to high variability between sites simply because patients at sites with longer scans spent more of the scan in a bad mood. At best, the neural correlates of that decreased mood will be uncorrelated with the effect of interest, increasing noise and reducing statistical power. At worst, they could be mistaken for neural correlates of a certain genotype that is more common in the country where the longer scans took place. One ENIGMA working group studying obsessive-compulsive disorder includes a reward processing task performed after a long period of scanning. This study took care to standardise scan length, but (as in most studies) the time between tasks was not specified (Simpson et al., 2020). If patients tended to take 10 minutes longer to navigate the preceding scans and tasks than healthy controls did, that added period of low stimulation could induce an 13.8% difference in mood at the start of their experimental scans (based on the mean decline of 1.38% mood per minute we observed in individuals at risk of depression). And as we see in Figure 5 and previous studies of reward, this level of mood difference can correspond with significant differences in behaviour.

Similarly, this temporal dependence could contribute to variability between studies. If a study is done at a center that allows participants to watch entertaining videos during collection of anatomical data, and a replication is conducted at a center where participants are instead asked to remain still and quiet, participants in the two studies may have consistently different moods at the start of the experimental scans. Researchers should make a habit of reporting all stimuli (or lack thereof) provided to participants during each scan as well as reporting the length of time required for setup for each experimental group.

In this paper, we introduce the new term passage-of-time dysphoria, and we believe it is important to do so for the following reasons. First, the phenomenon of passage-of-time dysphoria is highly replicable; second, it is of considerable effect size; third, it is relevant to both everyday situations and to scientific experiments that are conducted to shed light on important human conditions such as depression; fourth, and crucially, the phenomenon of passage-of-time dysphoria does not seem redundant: it is not accounted for by other existing terms such as boredom or mind wandering. It is also important to note that we employ the term passage-of-time dysphoria in the spirit of describing a mental phenomenon (Jaspers, 1973; Schneider, 1992; Berrios, 1992), as a first step before explaining or categorising it. As we note above, it is possible that mechanisms for passage-of-time dysphoria are reward sensitivity and opportunity cost, yet the subjective experience and its influence on the outcome of experimental studies seems to require the separate term that we have introduced.

Our study has several strengths, including the use of several independent cohorts, the application of a longitudinal design and the testing of reliability, the adherence to good practice of data analysis with preregistration and replication in left out data, and the use of rigorous computational modeling (including train-test splits and regularization). Importantly, our study included a developmentally diverse sample, demonstrating the effect in adolescents as well as in adults but also showing how the effect differs in people scoring high on depression (a finding that was itself independently replicated). We also conducted a series of control experiments to eliminate potential confounds and test alternative explanations that could have to do with the potential aversive nature of questions about mood or possible regression to the mean effects (See Supplementary Materials).

Yet our study should also be seen in light of some shortcomings.

First, central to our findings is the validity of self-reported momentary mood ratings. Such ratings can be criticised as being subjective and therefore hard to interpret. The use of a single measure makes it difficult to assign these changes to established psychological constructs such as cognitive fatigue, apathy, or affect. However, there are good reasons that momentary mood ratings are central to modern real-world monitoring techniques such as ecological momentary assessment (Ebner-Priemer and Trull, 2009). Decades of previous research show that momentary mood ratings have criterion validity and have been linked to consistent differences in behaviour and brain data (Pavot and Diener, 1993; Pavlickova et al., 2013; de Vries et al., 2008; Huntsinger and Ray, 2016; Mitterschiffthaler et al., 2007; Harrison et al., 2008; Costello and Angold, 1988). In our own experiments, single mood ratings at the beginning of the experiment showed strong association with trait mood ratings, thus underscoring their psychometric validity (Supplementary Figure 11). We also demonstrated that they are not redundant reflections of boredom or the valence of ongoing thought. Most importantly, momentary mood ratings are brief and unobtrusive, which allowed us to gain a dynamic picture of mood's change with time.

Our use of a bounded mood scale has consequences. First, we must consider the possibility that our depression-related findings were driven by floor effects. The effect persisted in categorical analyses (an outsized proportion of depressed participants showed positive mood slopes) and after excluding participants who reached an absolute or individual mood floor. Second, the bounded mood scale prevents the error term of our mood models from being truly Gaussian. Because LMEs are typically robust to such non-Gaussian distributions (Schielzeth et al., 2020), we do not expect this fact to change our LME findings. We chose to maintain the Gaussian assumption because it is well established in existing models, but it is likely that a different assumption would better fit the data. Because very little is known about the true error distribution, exploring alternative models is beyond the scope of this study. We attempted to mitigate the effect of any mismatch by capping the model predictions to the allowable range. We also initialised many parameters to non-normal distributions and restricted several parameters to feasible ranges on every iteration.

We have administered rest and a limited set of simple tasks in this study. Since passage-of-time dysphoria was observed in all of them except freely chosen real-life activities, it is difficult to discern the key contributing factors or the limits of its generalizability. Exploring the mood impact of the full space of possible tasks and situations, of course, is not a tractable problem. We have chosen to focus our attention on a class of paradigms that is extremely common in neuroscience: long, neutral, low-stimulation tasks. The presence of passage-of-time dysphoria in these tasks suggests that many psychological studies are at risk of previously unknown confounds. Most researchers would see these qualities as unobjectionable or even desirable for brain or behaviour studies, even those with clear relationships to mood. We hope that the results of our study will lead researchers to reexamine this idea in their own research.

Our choice to rely on purely behavioral data means that the study did not examine the effects of passage-of-time dysphoria on brain function. Such effects seem plausible given previous knowledge about the effects of mind wandering on brain connectivity (Webb et al., 2020) and the fact that positive affect may decrease after a scan (Gruberger et al., 2013).

To discern whether rest's consequences are limited to subjective mood, we explored a link to behaviour. We showed that participants with a rest period before their gambling game were less likely to gamble than those without one. When combined with the finding that such rest periods are dysphoric, this establishes that rest both decreases mood and decreases gambling behaviour. However, a significant correlation between passage-of-time dysphoria and gambling behaviour was not observed at the individual level. Our results are not able to discern whether the change in behaviour is directly linked to the observed dysphoria or to some other consequence of rest.

The existence of passage-of-time dysphoria is perhaps not surprising, but a detailed account of its shape, size, variability, and clinical correlates is both novel and consequential. Future work may reveal the neural correlates of this dysphoria; link it to other psychological constructs such as boredom, negative affect, and mind-wandering; and explore its relationship to task switching behavior and other explore/exploit decisions.

Methods

Participants

Online Adult Participants

Online adult participants were recruited using Amazon Mechanical Turk (Amazon.com, Inc., Seattle, WA), a service that allows a person needing work done (a "requester") to pay other people ("workers") to do computerised tasks ("jobs") from home (Paolacci et al., 2010). Requesters can use "qualifications" to require certain demographic or performance criteria in their participants. We required that our participants be adults living in the United States, that they have completed over 5,000 jobs for other requesters, and that over 97% of their jobs have been satisfactory to the requester. We also required that participants had not performed any of our tasks (which were relatively similar to the ones in this study) before.

Every online participant received the same written instructions and provided informed consent on a web page where they were required to click "I Agree" to participate. Because we did not obtain information by direct intervention or interaction with the participants and did not obtain any personally identifiable private information, our MTurk studies were classified as not human subjects research and were determined to be exempt from IRB review by the NIH Office of Human Subjects Research Protections (OHSRP). The consent process and task/survey specifics were approved by the OHSRP. For data to be included in the final analyses, participants were required to complete both a task and a survey (described below). Participants submitted a 6-to-10-digit code revealed at the end of each one to prove that they had completed it. Both the task and survey had to be completed in a 90-minute period starting when they accepted the job on Amazon Mechanical Turk.

914 participants completed the task online. Some data files did not save properly due to technical difficulties or the participant closing the task window before being asked to do so. 44 participants whose task or survey data did not save were excluded. Of the 870 remaining Mechanical Turk participants, 390 were female (44.8%). Participants had a mean age of 37.6 years (range: 19-74).

A subset of the online adult participants were invited to return the following day to repeat the same task and survey a second time. Of the 66 individuals who completed both the task and the survey on the first day, 53 (80.3%) completed the task and survey on the second day. Gambling trials were randomised independently so that the subject was not seeing the exact same trials both times. Participants could complete the second task and survey any time in the following three days, but the task and survey had to be done together in the same 90-minute period.

Similarly, a different cohort was invited to return a week after their first run to repeat the same task and survey. These participants could complete the second task and survey any time in the following six days, but the task and survey had to be done together in the same 90-minute period. This cohort was then invited to complete the same task and survey a third time, two weeks after their first run. 196 individuals completed the task and survey the first week. 163 (83.2%) of these completed the task and survey the second week and 158 (80.6%) completed the task and survey the third week. 149 (76.0%) individuals completed the task and survey in all three weeks.

Online Adolescent Participants

Adolescent participants recruited in person at the National Institute of Mental Health were also invited to participate by completing a similar task on their computer at home. These participants completed a different set of questionnaires, developed for adolescents, about their mental health. Every participant received the same scripted instructions and provided informed consent to a protocol approved by the NIH Institutional Review Board.

There were 230 adolescents enrolled in the NIMH depression characterization study who were offered to complete tasks for this study. 129 agreed, a participation rate of 56.1%. 10 adolescents who had not completed all three questionnaires were excluded from the results, as were 3 participants who declined to allow their data to be shared openly. Of the remaining 116 adolescent participants, 77 were female (66.4%). They had a mean age of 16.3 years (range: 12 - 19). 56 participants (48.2%) had been diagnosed with MDD by a clinician at the NIH, and 4 were determined to have sub-clinical MDD (3.4%). Participants had a mean depression score of MFQ = 6.5 (± 5.5 SD) and a mean anxiety score of SCARED = 2.2 (± 3.0 SD).

To assess the stability of findings in this population, the in-person adolescent participants were invited to return each week to complete the same task again, up to three times. 82 (70.6%) individuals completed the task a week later and 4 (3.4%) completed the task a third time the following week. The analyses presented in this paper use only the first run from this cohort.

Boredom, Mind-Wandering, and Activities Participants

In response to reviewer comments, a preregistered follow-up analysis included five new cohorts of MTurk participants who received similar tasks that also included mood ratings, rest periods, and the gambling game. This group was recruited to investigate the impacts of boredom and mind-wandering on mood changes, so they completed surveys about these traits in addition to the demographics, CES-D, and SHAPS questions. Participants were randomised to one of these 5 "follow-up cohorts," summarised in Supplementary Table 1:

- BoredomBeforeAndAfter (n=150), who received a boredom state questionnaire both before and after a 7-minute rest period with 15 s of rest between mood ratings.
- BoredomAfterOnly (n=150), who received a boredom state questionnaire only *after* a 7-minute rest period with 15 s of rest between mood ratings.
- MwBeforeAndAfter (n=150), who received a multidimensional experience sampling (MDES) questionnaire both before and after a 7-minute rest period with 15 s of rest between mood ratings.
- MwAfterOnly (n=150), who received an MDES questionnaire only *after* a 7-minute rest period with 15 s of rest between mood ratings.
- Activities (n=450), who received instructions to leave the task for 7 minutes and perform activities of their choice, completing mood ratings just before and after this period.

After the rest periods described above, each group completed a block of negative closed-loop gambling trials and a block of positive closed-loop gambling trials (as described in the "Gambling Blocks" section). Details of the cohorts' tasks are found in the following sections. A full description of the preregistered tasks and analyses can be found at https://osf.io/gt7a8. 1143 participants completed these tasks online. 93 participants were excluded because their task or survey data was incomplete or did not save, because they completed the task more than once despite instructions to the contrary, or because they failed to answer one or more "catch" questions correctly on the survey. Of the 1050 remaining participants, 463 were female (44.1%). Participants had a mean age of 39.3 years (range: 20-80).

Mobile App Participants

Gambling behaviour and mood rating data were collected from a mobile app called "The Great Brain Experiment", described in (Rutledge et al., 2014). The Research Ethics Committee of University College London approved the study. When participants opened the app for the first time, they gave informed consent by reading a screen of information about the research and clicking "I Agree." They then rated their life satisfaction as an integer between 0 (not at all) and 10 (completely). Any time they used the app after this, participants could then choose between several games, including one called "What makes me happy?" that was used in this research. We used a subset of 26,896 people, primarily from the US and UK, in our analyses. The median life satisfaction of the included participants, which will be used as a proxy for depression risk in this cohort, was 7/10. Age for this cohort was provided in bands. These are the bands and number of individuals in each band in the subset of data used in our analysis: 18-24 (6,500), 25-29 (4,522), 30-39 (7,190), 40-49 (4,829), 50-59 (2,403), 60-69 (1,158), and 70+ (294). 13,168 were female (49.0%).

Mobile app participants were randomly split into an exploratory cohort of 5,000 participants and a confirmatory cohort of all remaining participants. All analyses and hyperparameters involving mobile app participants were optimised using only the exploratory cohort, then tested on the confirmatory cohort. These confirmatory analyses were preregistered on the Open science Framework (https://osf.io/paqf6).

In the linear mixed effects model described below, we made an effort to exclude participants who were outliers in the time they took to complete the task. Such outliers would have a large effect on the LME model's mood slope term, where non-zero slopes would lead to large errors in these outlier participants. Outlier completion times also suggest that the participant was not fully paying attention to the task, either by responding without thinking or leaving the app for an extended period. Mobile app participants with an average task completion time that was less than $Q1 - 1.5 * IQR$ or greater than $Q3 + 1.5 * IQR$ (where Q1 is the 25th percentile, Q3 is the 75th percentile, and IQR = Q3-Q1) were excluded from this linear mixed effects analysis. 4.65% of participants were excluded based on these criteria, leaving $n = 20,877$ mobile app participants.

Task and Survey

The online tasks were created using PsychoPy3 (v2020.1.2) and were uploaded to the task hosting site Pavlovia for distribution to participants. Pavlovia used the javascript package PsychoJS to display tasks in the web browser. Each task used the latest version of Pavlovia and PsychoJS available at the time of data collection. A list of all cohorts collected can be seen in Supplementary Table 1.

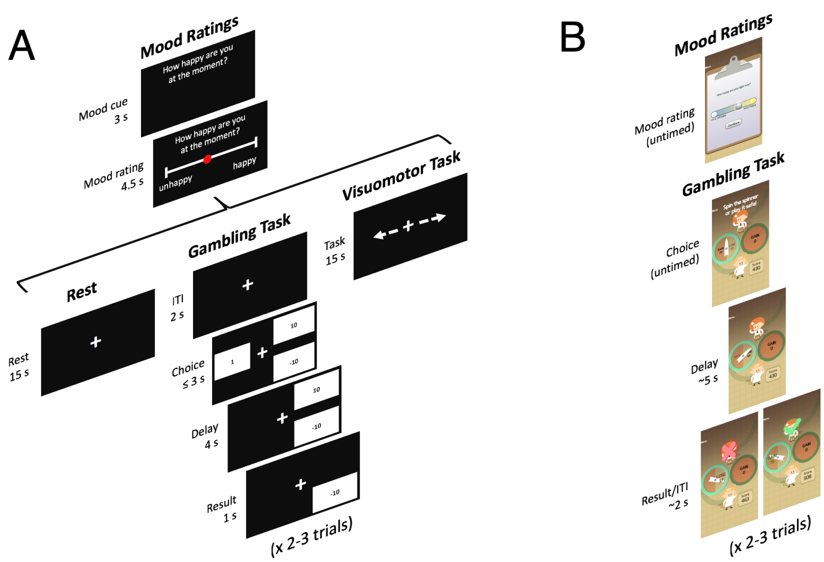

Figure 6: One cycle (mood rating + task) of the administered to (A) online participants and (B) mobile app participants. After completing their first mood rating, participants completed one cycle of the rest, gambling, or visuomotor task, then completed another mood rating, and so on. In the case of the rest and visuomotor tasks, the cycle duration was determined by time. In the case of the gambling task, it was determined by the time taken to complete 2 or 3 (randomised) trials of the gambling task.

Mood Ratings

The task given to online participants is outlined in Figure 6A. Periodically during all tasks, participants were asked to rate their mood. Participants first saw the question "How happy are you at the moment?" for 3 seconds. Then a slider appeared below the question, with a scale whose ends were labeled "unhappy" and "happy." A red circle indicated the current slider position, and it started in the middle for each rating.

Participants could press and hold the left and right arrow keys to move the slider, then spacebar to lock in their response. If the spacebar was not pressed in 4.5 seconds, the current slider position was used as their mood rating.

As part of the instructions at the start of each run, the participant was asked to rate their overall "life happiness" in a similar (but slightly slower) rating. In this case, participants first saw the question "Taken all together, how happy are you with your life these days?" for 4 seconds. The slider then appeared, and the participant had 6.5 seconds to respond.

In one alternative version of the task, participants were asked to rate their mood with a single keypress instead of a slider. They could press a key 1-9 to indicate their current mood, where 1 indicated "very unhappy" and 9 indicated "very happy." This alternative version was used to investigate the possibility that mood effects could be an artefact of the rating method, where participants' ratings converged to the middle because this rating required the least effort.

Rest Blocks

In some blocks, participants were asked to simply rest in between mood ratings. These rest periods consisted of a central fixation cross presented on the screen. The duration of the rest period was 15 seconds for most versions of the experiment. For some versions, this duration was made longer or shorter to disentangle the impacts of rating frequency and elapsed time on mood, investigating the possibility that the mood ratings themselves were aversive.

Thought Probes and Activities Questions

Follow-up versions of the task included thought probes about state boredom or the emotional valence of ongoing thought (including mind-wandering). These groups received rest blocks as described above, but with additional questions just before and/or after it.

Two cohorts were collected to quantify the relationship between passage-of-time dysphoria and boredom. Each received a rest period with mood ratings 20 seconds apart, followed by the Multidimensional State Boredom Scale's short form (MSBS-SF), an 8-item scale of state boredom (Hunter et al., 2016). Participants rated statements like "I feel bored" on a 7-point Likert scale from 1 ("Strongly Disagree") to 7 ("Strongly Agree"). Their level of boredom was quantified as the sum of their ratings on the 8 questions. The first (cohort BoredomBeforeAndAfter, $n = 150$) completed the MSBS-SF both before and after the rest period. The second (cohort BoredomAfterOnly, $n = 150$) completed the MSBS-SF only after the rest period.

Two other cohorts were collected to quantify the relationship between passage-of-time dysphoria and the emotional valence of ongoing thought (including mind-wandering). Each participant in the two mind-wandering cohorts received a rest period with mood ratings 20 seconds apart, followed by a 13-item Multidimensional Experience Sampling (MDES) as described in Turnbull et al. (2019). Participants were asked to respond to a set of questions by clicking on a continuous slider. Most questions, like "my thoughts were focused on the task I was performing", were rated from "not at all" (scored as -0.5) to "completely" (scored as 0.5). The first (cohort MwBeforeAndAfter, $n = 150$) completed the MDES only after the rest period. The second (cohort BoredomAfterOnly, $n = 150$) completed the MDES only after the rest period.

As described in Ho et al. (2020), we applied principal components analysis (PCA) on participants' MDES responses to find a component whose primary loading was on the "emotion" item (in which they reported their thoughts as being negative or positive). The "emotion dimension" of each MDES was then quantified as the amplitude of this component. The sign of PCA components is not meaningful, so we arbitrarily chose that increased emotion dimension would represent *more negative* thoughts.

Another follow-up task investigated the impact on mood of a break period where participants were released to do whatever they wanted. Just before this break period, an alarm sound was played on repeat, and participants were asked to increase the volume on their computer until they could hear the alarm clearly.

Participants were informed that they would have 7 minutes to put the task aside and do something else but should be ready to come back when the alarm sounded at the end. After these instructions and before the break, they rated their mood. During the break, the task window displayed a message saying "this is the break. An alarm will sound when the break is over." After the alarm sounded and participants returned, they rated their mood again. They were then asked 27 questions about how much of the break they spent doing various activities. They were asked to rate each by clicking on a 5-point Likert scale with options labeled "not at all" (scored at 0%), "a little" (scored at 25%), "about half the time" (scored at 50%), "a lot" (scored at 75%), or "the whole time" (scored at 100%). These scores were used to roughly describe the most common activities performed by the participants during the break.

Participants were randomised to one of the follow-up cohorts described in this section at the time of participation.

Task Blocks

In some blocks, participants completed a simple visuomotor task. In this task, the fixation cross moved back and forth across the screen in a sine wave pattern (peak-peak amplitude: 1x screen height, period: 4 seconds). Participants were asked to press the spacebar at the exact moment when the cross was in the center of the screen (as denoted by a small dot). In some blocks, they received feedback on their performance: each time they responded, the white cross turned green for 400 ms if the spacebar was pressed within the middle 40% of the sine wave's position amplitude (i.e., less than 0.262 seconds before or after the actual center crossing).

Gambling Blocks

In each trial of the gambling task, participants saw a central fixation cross for 2 seconds. Three boxes with numbers in them then appeared. Two boxes on the right side of the screen indicated the possible point values they could receive if they chose to gamble (the "win" and "loss" values). On the left side, a single number indicated the points they would receive if they chose not to gamble (the "certain" value). Participants had 3 seconds to press the right or left arrow key to indicate whether they wanted to gamble or not. If no choice was made, gambling was chosen by default. After making their choice, the option(s) not chosen would disappear. If they chose to gamble, both possible gambling outcomes appeared for 4 seconds, then the actual outcome appeared for 1 second. If they chose not to gamble, the certain outcome appeared for 5 seconds. The locations (top/bottom) of the higher and lower gambling options were randomised.

The gambling outcome values were calculated according to several rules depending on the version of the experiment. In each version, the "base" value was a random value between -4 and 4 points. The other value was this base value plus a positive or negative reward prediction error (RPE). If they chose to gamble, participants would always receive the base value + RPE option. To encourage gambling, the "certain" value was set to $(win + 2 * loss)/3$, or 1/3 of the way from the loss value to the win value. (Note that this rule was the same for every subject and was therefore unlikely to drive individual differences in gambling behaviour.)

In the "random" version, the RPE was a random value with uniform distribution between -5.0 and 5.0. RPEs with a magnitude of less than 0.03 were increased to 0.03. If 3 trials in a row happened to have the same outcome (win or loss), the next trial was forced to have the other outcome.

In the "closed-loop" version, RPEs were calculated based on the difference between a participant's mood and a "target mood" of 0 or 1. Some blocks of trials were "positive" blocks in which the participant had a 70% chance of winning on each trial ("positive congruent trials") and a 30% chance of losing ("positive incongruent trials"). Other blocks were "negative" blocks in which the participant had a 70% chance of losing on each trial ("negative congruent trials") and a 30% chance of winning ("negative incongruent trials"). If there had been 3 incongruent trials in a row, the next trial was forced to be congruent. The RPE was calculated as in a Proportional-Integral (PI) controller: a weighted sum of the current difference and the integral across all such differences reported so far in the block. The weightings were different for congruent and incongruent trials. Specifically, the RPE was set to:

$$RPE(t) = \begin{cases} 14 * (M(t-1) - M_{target}) + \sum_{j=1}^{t-1}(M(j) - M_{target}) & congruent\ trial \\ -3.5 * (M(t-1) - M_{target}) + \sum_{j=1}^{t-1}(M(j) - M_{target})/12 & incongruent\ trial \end{cases}$$

Where $t = 1, 2, ..., n$ is the trial index relative to the start of the block, $M(t)$ is the mood reported after trial t, and M_{target} is the target mood for the current block. RPEs with a magnitude of less than 0.03 were assigned a magnitude of 0.03.

During gambling blocks, mood ratings occurred after every 2 or 3 trials (on average, 1 rating every 2.4 trials). Every subject received mood ratings after the same set of trials.

At the end of the task, participants were presented with their overall point total. These point totals were translated into a cash bonus of $1-6 depending on their performance. Bonus cutoffs were determined based on simulations such that any value 1-6 were possible to achieve, but a typical subject gambling at every opportunity could be expected to receive approximately $3. Upon payment, participants received $8 for their participation (this was later increased to $10) plus this bonus.

Survey

After performing the task, online adult participants were asked to complete a series of questionnaires. In the demographics portion, they were asked for their age, gender and location (city and state). They were also asked to indicate their overall status using the MacArthur Scale of Subjective Social Status (Adler et al., 2000). Shown a ten-rung ladder, participants clicked on the rung that represented their overall status relative to others in the United States. This scale is a widely used indicator of subjective social status, and in certain cases, it has been shown to indicate health status better than objective measures of socioeconomic status (Singh-Manoux et al., 2005).

After the demographics portion, online adult participants completed questionnaires including the Center for Epidemiologic Studies Depression Scale (CES-D), a 20-item scale of depressive symptoms (Radloff, 1977). They also completed the Snaith–Hamilton Pleasure Scale (SHAPS), a 14-item scale of hedonic capacity (Snaith et al., 1995).

In-person adolescent participants completed a different set of questionnaires, selected to be age-appropriate and maintain consistency with other ongoing research projects. These questionnaires included the Mood and Feelings Questionnaire (MFQ), a 33-item scale of how the participant has been feeling and acting recently (Costello and Angold, 1988). They also included the Screen for Child Anxiety Related Emotional Disorders (SCARED), a 41-item scale of childhood anxiety (Birmaher et al., 1999). These questionnaires were completed before the subject began completing the online tasks described above.

Participants recruited for follow-up investigations of boredom, mind-wandering, and free time activities also completed the short boredom proneness scale (SBPS), an 8-item scale of an individual's proneness to boredom in everyday life (Struk et al., 2017). They also completed the 5-item mind-wandering questionnaire (MWQ), which quantifies a person's proneness to mind-wandering in everyday life (Mrazek et al., 2013). The SBPS and MWQ were used to quantify trait-level boredom and mind-wandering, respectively.

Mobile App

The task given to mobile app participants is outlined in Figure 6B. Mobile app participants completed 30 trials of a gambling game. In each trial, participants chose between a certain option and a gamble, represented as a spinner in a circle with two possible outcomes. If the participant chose to gamble, the spinner rotated for approximately 5 seconds before coming to rest on one of the two outcomes. Participants were equally likely to win or lose if they chose to gamble. The points were added to or subtracted from the participant's total during an approximately 2-second inter-trial interval before the game advanced to the next trial. After every 2-3 trials (12 times per play), the participant rated their mood. They were presented with the question, "How happy are you right now?". A slider was presented with a range from "very unhappy" to "very happy."

The participant could select a value by moving their finger on the slider and tapping "Continue". No limit was placed on their reaction times.

Each participant received 11 gain trials (with gambles between one positive outcome and one zero), 11 loss trials (one negative outcome and one zero), and 8 mixed trials (one positive and one negative outcome). The possible gambling outcomes were randomly drawn from a list of 60 gain trials, 60 loss trials, and 30 mixed trials. Participants played one of two versions of the app, between which the only difference was the precise win, loss, and certain amounts in these lists. The amounts in the first version are described in detail in the supplementary material of (Rutledge et al., 2014). In the second version, gain trials had 3 certain amounts (35, 45, 55) and 15 gamble amounts (59, 66, 72, 79, 85, 92, 98, 105, 111, 118, 124, 131, 137, 144, 150). As in the first version, the set of loss trials was identical to the gain trials except that the values were negative. Mixed trials has 3 prospective gains (40, 44, 75) and 10 prospective losses (-10, -19, -28, -37, -46, -54, -63, -72, -81, -90). Both versions are described further in (Bedder et al., 2020). The median participant played the game for approximately 5 minutes.

After playing the game, participants saw their score plotted against those of other players, and they were told if their score was a "new record" for them. They could then choose to play again and try to improve their score. We reasoned that introducing the notion of a "new record" would significantly change participants' motivations and behaviour on subsequent runs, and we therefore limited our analysis to the first run from each participant.

Linear Mixed Effects Model

Analyses and statistics were performed using custom scripts written in Python. Participants' momentary subjective mood ratings were fitted with a linear mixed effects (LME) model with rating time as a covariate using the Pymer4 software package (http://eshinjolly.com/pymer4/). Rating times were converted to minutes to satisfy the algorithm's convergence criteria while maintaining interpretability. This method resulted in each participant's data being modeled by a slope and intercept parameter such that:

$$M(t) = M_0 + \beta_T * T(t) \tag{1}$$

where M_0 is the estimated mood at block onset (intercept), β_T is the estimated change in mood per minute (slope), and $T(t)$ is the time in minutes from the start of the block. The LME modeling algorithm also produced a group-level slope and intercept term as well as confidence intervals and statistics testing against the null hypothesis that the true slope or intercept was zero.

The first block of the first run for all online adult and in-person adolescent cohorts experiencing rest or random gambling first were fitted together in a single model, with factors:

$$\begin{aligned} Mood \sim & 1 + Time * (isMale + meanIRIOver20 + totalWinnings + meanRPE + & (2)\\ & fracRiskScore + isAge0to16 + isAge16to18 + isAge40to100) + (Time|Subject) & (3) \end{aligned}$$

isMale is 1 if the participant is male, 0 otherwise. meanIRIOver20 is the mean inter-rating interval across the block(s) of interest (in seconds) minus 20 (a round number near the mean). totalWinnings is the total points won by the participant in the block(s). meanRPE is the mean reward prediction error across the block(s). totalWinnings and meanRPE will be zero for participants who were experiencing rest instead of gambling. fracRiskScore is the participant's clinical depression risk score divided by a clinical cutoff: i.e., their MFQ score divided by 12 or their CES-D score divided by 16.

For reliability analyses, the first block of each run was modeled separately for each cohort/run with the same model shown above. An intraclass correlation coefficient quantifying absolute agreement (ICC(2,1)) between the runs of each cohort, was calculated using R's "psych" package, accessed through the python wrapper package rpy2.

To measure the psychometric validity of the subjective momentary mood ratings, we correlated the initial mood (or "Intercept") parameter of this model with the life happiness ratings. The correlation was highly significant ($r_s = 0.548, p = 1.49 * 10^{-70}$, Supplementary Figure 11, left).

For comparisons with the online data, the same model was also employed in the initial analysis of the mobile app data.

Computational Model

When examining the effect of time on mood during random gambling in the mobile app data, we next attempted to disentangle time's effects from those of reward and expectation using a computational model. The model is based on one described in detail by (Keren et al., 2021) that has been validated on behavioural data from a similar gambling task. The authors found that changes in momentary subjective mood were predicted accurately by a weighted combination of current and past rewards and RPEs in the task. Quantifying RPEs relies on subjective expectations that are formulated according to a "primacy model," in which expected reward is more heavily influenced by early rewards than it is by recent ones.

The model described in (Keren et al., 2021) was modified to include a coefficient β_T that linearly relates time and mood. Our modified model is defined as follows:

$$\hat{M}(t) = M_0 + \beta_A \sum_{u=1}^{t} \lambda^{t-u} A(u) + \beta_E \sum_{u=1}^{t} \lambda^{t-u} E(u) + \beta_T T(t) \tag{4}$$

In the above equation, $t = 1, 2, ...n$ is the trial index, and $\hat{M}(t)$ is the estimated mood rating from trial t. M_0 (the estimated mood at time 0), λ (an exponential discounting factor), and the βs are learned parameters of the model. $A(t)$ is the actual outcome (in hundreds of points) of trial t, $T(t)$ is the time of trial t in minutes, and $E(t)$ is the primacy model of the subject's reward expectation in trial t, defined as:

$$E(t) = \frac{1}{t-1} \sum_{u=1}^{t-1} A(u) \tag{5}$$

If we remove the influence of time (i.e., set our $\beta_T = 0$), the full mood model in (Keren et al., 2021) is equivalent to this one as long as its reward prediction error coefficient is less than its expectation coefficient (i.e., $\beta_R^{Keren} < \beta_E^{Keren}$) and $\beta_E^{Keren} > 0$, where β_R^{Keren} and β_E^{Keren} denote the values β_R and β_E defined in (Keren et al., 2021)). The values in our model can be derived from the values in theirs by setting $\beta_A = \beta_R^{Keren}$ and $\beta_E = \beta_E^{Keren} - \beta_R^{Keren}$.

We used the PyTorch package (Paszke et al., 2019) on a GPU to fit 500 models simultaneously for each subject. β_T was initialised to random values with distribution $\mathcal{N}(0, 1)$. β_E and β_A were initialised to random values with distribution $Lognormal(0, 1)$ and capped to the interval [0,10] on every iteration. M_0 and λ were initialised to random values with normal distributions $\mathcal{N}(0, 1)$, then sigmoid-transformed (to facilitate optimization and conform to the interval [0, 1]) using the standard logistic function:

$$y = \frac{1}{1 + e^{-x}} \tag{6}$$

At the end of 100,000 iterations, the model with the lowest sum of squared errors (i.e., $\sum_{t=1}^{N} (\hat{M}(t) - M(t))^2$) was selected. The time coefficient β_T learned by the model could then be used as a measure of the influence of time on that participant's mood, disentangled from the effects of rewards and RPEs.

End-to-end optimization was carried out using ADAM (Kingma and Ba, 2014) with a learning rate of $\alpha = 0.005$. L2 penalty terms were placed on the β terms and added to the sum of squared errors. This meant that the objective function being minimised was:

$$L = \sum_{t=1}^{n} (\hat{M}(t) - M(t))^2 + \lambda_{EA} * ({\beta_A}^2 + {\beta_E}^2) + \lambda_T * {\beta_T}^2 \tag{7}$$

The regularization hyperparameters λ_{EA} and λ_T were determined from a tuning step, in which the model was trained on the first 10 mood ratings and tested on the last two in each of 5,000 exploratory participants. One model was trained with each combination of λ_{EA} and λ_T ranging from 10^{-4} to 10^3 in 20 steps (evenly spaced on a log scale). The testing loss (median across participants) across penalty terms was fitted to a third degree polynomial using Skikit-Learn's kernel ridge regression with regularization strength $\alpha = 10.0$. The best fitting regularization hyperparameters were defined as those that minimised this smoothed testing loss.

As in the online cohort's LME model, the initial mood parameter M_0 showed psychometric validity. It was significantly correlated with life happiness ($r_s = 0.362, p < 2.23 * 10^{-308}$, Supplementary Figure 11, right).

Control Model

To quantify the effect of including the time-related term, we fitted a control model without β_T. This control model is defined as follows:

$$\hat{M}(t) = M_0 + \beta_A \sum_{u=1}^{t} \lambda^{t-u} A(u) + \beta_E \sum_{u=1}^{t} \lambda^{t-u} E(u) \tag{8}$$

As in the primary model, the regularization hyperparameter λ_{EA} in this control model was tuned using the method described above.

Data Availability

All data used in the manuscript have been made publicly available. Online Participants' data can be found on the Open Science Framework at https://osf.io/xbc6u/ . Mobile App Participants' data can be found on Dryad at https://doi.org/10.5061/dryad.prr4xgxkk (Rutledge, 2021).

Code Availability

The code for each task and survey is available from the corresponding author upon request. Our data analysis software, as well as the means to create a Python environment that automatically installs it on a user's machine, has been made available online at https://github.com/djangraw/PassageOfTimeDysphoria.

Acknowledgments

This research was supported in part by the Intramural Research Program of the National Institute of Mental Health, part of the National Institutes of Health (NIH) (Grant No. ZIA-MH002957-01 [to AS]). This work used the computational resources of the NIH high-performance computing (HPC) Biowulf cluster (http://hpc.nih.gov). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; or decision to submit the manuscript for publication. The views expressed in this article do not necessarily represent the views of the National Institutes of Health, the Department of Health and Human Services, or the United States Government.

References

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., and Platt, M. L. (2017). A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology*, 42(10):1931–1939.

Adhikari, B. M., Jahanshad, N., Shukla, D., Turner, J., Grotegerd, D., Dannlowski, U., Kugel, H., Engelen, J., Dietsche, B., Krug, A., Kircher, T., Fieremans, E., Veraart, J., Novikov, D. S., Boedhoe, P. S. W., van der Werf, Y. D., van den Heuvel, O. A., Ipser, J., Uhlmann, A., Stein, D. J., Dickie, E., Voineskos, A. N., Malhotra, A. K., Pizzagalli, F., Calhoun, V. D., Waller, L., Veer, I. M., Walter, H., Buchanan, R. W., Glahn, D. C., Hong, L. E., Thompson, P. M., and Kochunov, P. (2019). A resting state fMRI analysis pipeline for pooling inference across diverse cohorts: an ENIGMA rs-fMRI protocol. *Brain Imaging and Behavior*, 13(5):1453–1467.

Adler, N. E., Epel, E. S., Castellazzo, G., and Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology*, 19(6):586–592.

Agrawal, M., Mattar, M. G., Cohen, J. D., and Daw, N. D. (2020). The temporal dynamics of opportunity costs: A normative account of cognitive fatigue and boredom. *bioRxiv*.

Bedder, R. L., Vaghi, M. M., Dolan, R. J., and Rutledge, R. B. (2020). Risk taking for potential losses but not gains increases with time of day. *psyarxiv*.

Berrios, G. E. (1992). Phenomenology, psychopathology and Jaspers: a conceptual history. *History of psychiatry*, 3(11):303–327.

Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., and Baugher, M. (1999). Psychometric properties of the screen for child anxiety related emotional disorders (SCARED): A replication study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38(10):1230–1236.

Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N., and Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nature Reviews Neuroscience*, 17(11):718–731.

Ciocan, C. (2010). Heidegger and the Problem of Boredom. *Journal of the British Society for Phenomenology*, 41(1):64–77.

Cohen, J. D., McClure, S. M., and Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):933–942.

Costello, E. J. and Angold, A. (1988). Scales to Assess Child and Adolescent Depression: Checklists, Screens, and Nets. *Journal of the American Academy of Child & Adolescent Psychiatry*, 27(6):726–737.

de Vries, M., Holland, R. W., and Witteman, C. L. M. (2008). In the winning mood: Affect in the Iowa gambling task. *Judgment and Decision Making*, 3(1):42–50.

Ebner-Priemer, U. W. and Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation.

Egloff, B., Tausch, A., Kohlmann, C. W., and Krohne, H. W. (1995). Relationships between time of day, day of the week, and positive mood: Exploring the role of the mood measure. *Motivation and Emotion*, 19(2):99–110.

Elpidorou, A. (2014). The bright side of boredom. *Frontiers in psychology*, 5:1245.

Feingold, A. (2015). Confidence interval estimation for standardized effect sizes in multilevel and latent growth modeling. *Journal of consulting and clinical psychology*, 83(1):157.

Geana, A., Wilson, R., Daw, N. D., and Cohen, J. D. (2016). Boredom, Information-Seeking and Exploration. In *CogSci*.

Gruberger, M., Maron-Katz, A., Sharon, H., Hendler, T., and Ben-Simon, E. (2013). The wandering mood: Psychological and neural determinants of rest-related negative affect. *Frontiers in Psychology*, 4(DEC):1–10.

Halahakoon, D. C., Kieslich, K., O'Driscoll, C., Nair, A., Lewis, G., and Roiser, J. P. (2020). Reward-processing behavior in depressed participants relative to healthy volunteers: A Systematic Review and Meta-analysis. *JAMA psychiatry*.

Harrison, B. J., Pujol, J., Ortiz, H., Fornito, A., Pantelis, C., and Yücel, M. (2008). Modulation of brain resting-state networks by sad mood induction. *PLoS ONE*, 3(3).

Heidegger, M. (1995). *The fundamental concepts of metaphysics: World, finitude, solitude*. Indiana University Press.

Ho, N. S. P., Poerio, G., Konu, D., Turnbull, A., Sormaz, M., Leech, R., Bernhardt, B., Jefferies, E., and Smallwood, J. (2020). Facing up to why the wandering mind: Patterns of off-task laboratory thought are associated with stronger neural recruitment of right fusiform cortex while processing facial stimuli. *NeuroImage*, 214(March):116765.

Hunter, J. A., Dyer, K. J., Cribbie, R. A., and Eastwood, J. D. (2016). Exploring the utility of the Multidimensional State Boredom Scale. *European Journal of Psychological Assessment*, 32(3):241–250.

Huntsinger, J. R. and Ray, C. (2016). A flexible influence of affective feelings on creative and analytic performance. *Emotion*, 16(6):826–837.

Jaspers, K. (1973). Die abnorme Seele in Gesellschaft und Geschichte (Soziologie und Historie der Psychosen und Psychopathien). In *Allgemeine Psychopathologie*, pages 594–623. Springer.

Kahneman, D. (1973). *Attention and effort*, volume 1063. Citeseer.

Keren, H., Zheng, C., Jangraw, D. C., Chang, K., Vitale, A., Rutledge, R. B., Pereira, F., Nielson, D. M., and Stringaris, A. (2021). The temporal representation of experience in subjective mood. *eLife*, 10:1–24.

Killingsworth, M. A. and Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, 330(6006):932.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Mikulas, W. L. and Vodanovich, S. J. (1993). The essence of boredom. *The Psychological Record*, 43(1):3.

Miner, A. G. and Glomb, T. M. (2010). State mood, task performance, and behavior at work: A within-persons approach. *Organizational Behavior and Human Decision Processes*, 112(1):43–57.

Mitterschiffthaler, M. T., Fu, C. H., Dalton, J. A., Andrew, C. M., and Williams, S. C. (2007). A functional MRI study of happy and sad affective states induced by classical music. *Human Brain Mapping*, 28(11):1150–1162.

Mrazek, M. D., Phillips, D. T., Franklin, M. S., Broadway, J. M., and Schooler, J. W. (2013). Young and restless: validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Frontiers in psychology*, 4:560.

Nunokawa, J. (1996). The Importance of Being Bored: The Dividends of Ennui in" The Picture of Dorian Gray". *Studies in the Novel*, 28(3):357–371.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical turk. *Judgment and Decision Making*, 5(5):411–419.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., and Antiga, L. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pages 8026–8037.

Pavlickova, H., Varese, F., Smith, A., Myin-Germeys, I., Turnbull, O. H., Emsley, R., and Bentall, R. P. (2013). The Dynamics of Mood and Coping in Bipolar Disorder: Longitudinal Investigations of the Inter-Relationship between Affect, Self-Esteem and Response Styles. *PLOS ONE*, 8(4):e62514.

Pavot, W. and Diener, E. (1993). The affective and cognitive context of self-reported measures of subjective well-being. *Social Indicators Research*, 28(1):1–20.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier Science.

Poerio, G. L., Totterdell, P., and Miles, E. (2013). Mind-wandering and negative mood: Does one thing really lead to another? *Consciousness and Cognition*, 22(4):1412–1421.

Proust, M. (2013). *Swann's Way: In Search of Lost Time, Volume 1*, volume 1. Yale University Press.

Pulcu, E., Trotter, P. D., Thomas, E. J., McFarquhar, M., Juhász, G., Sahakian, B. J., Deakin, J. F. W., Zahn, R., Anderson, I. M., and Elliott, R. (2014). Temporal discounting in major depressive disorder. *Psychological medicine*, 44(9):1825–1834.

Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, 1(3):385–401.

Raffaelli, Q., Mills, C., and Christoff, K. (2018). The knowns and unknowns of boredom: a review of the literature. *Experimental Brain Research*, 236(9):2451–2462.

Ratcliffe, M. (2013). Why mood matters. *The Cambridge companion to Heidegger's being and time*, pages 157–176.

Robison, M. K., Miller, A. L., and Unsworth, N. (2020). A multi-faceted approach to understanding individual differences in mind-wandering. *Cognition*, 198(September 2019):104078.

Rutledge, R. B. (2021). Risky decision and happiness task: The Great Brain Experiment smartphone app.

Rutledge, R. B., Skandali, N., Dayan, P., and Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences of the United States of America*, 111(33):12252–12257.

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., and Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, 11(9):1141–1152.

Schneider, K. (1992). *Klinische Psychopathologie*. Georg Thieme Verlag, Stuttgart, 14 edition.

Shattuck, R. (2001). *Proust's way: A field guide to In Search of Lost Time*. WW Norton & Company.

Simpson, H. B., van den Heuvel, O. A., Miguel, E. C., Reddy, Y. C. J., Stein, D. J., Lewis-Fernández, R., Shavitt, R. G., Lochner, C., Pouwels, P. J. W., Narayanawamy, J. C., Venkatasubramanian, G., Hezel, D. M., Vriend, C., Batistuzzo, M. C., Hoexter, M. Q., de Joode, N. T., Costa, D. L., de Mathis, M. A., Sheshachala, K., Narayan, M., van Balkom, A. J. L. M., Batelaan, N. M., Venkataram, S., Cherian, A., Marincowitz, C., Pannekoek, N., Stovezky, Y. R., Mare, K., Liu, F., Otaduy, M. C. G., Pastorello, B., Rao, R., Katechis, M., Van Meter, P., and Wall, M. (2020). Toward identifying reproducible brain signatures of obsessive-compulsive profiles: rationale and methods for a new global initiative. *BMC Psychiatry*, 20(1):68.

Singh-Manoux, A., Marmot, M. G., and Adler, N. E. (2005). Does subjective social status predict health and change in health status better than objective status? *Psychosomatic Medicine*, 67(6):855–861.

Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., and Trigwell, P. (1995). A scale for the assessment of hedonic tone. The Snaith-Hamilton Pleasure Scale. *British Journal of Psychiatry*, 167(JULY):99–103.

Struk, A. A., Carriere, J. S. A., Cheyne, J. A., and Danckert, J. (2017). A short boredom proneness scale: Development and psychometric properties. *Assessment*, 24(3):346–359.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., Wright, M. J., Martin, N. G., Agartz, I., Alda, M., Alhusaini, S., Almasy, L., Almeida, J., Alpert, K., Andreasen, N. C., Andreassen, O. A., Apostolova, L. G., Appel, K., Armstrong, N. J., Aribisala, B., Bastin, M. E., Bauer, M., Bearden, C. E., Bergmann, Ø., Binder, E. B., Blangero, J., Bockholt, H. J., Bøen, E., Bois, C., Boomsma, D. I., Booth, T., Bowman, I. J., Bralten, J., Brouwer, R. M., Brunner, H. G., Brohawn, D. G., Buckner, R. L., Buitelaar, J., Bulayeva, K., Bustillo, J. R., Calhoun, V. D., Cannon, D. M., Cantor, R. M., Carless, M. A., Caseras, X., Cavalleri, G. L., Chakravarty, M. M., Chang, K. D., Ching, C. R. K., Christoforou, A., Cichon, S., Clark, V. P., Conrod, P., Coppola, G., Crespo-Facorro, B., Curran, J. E., Czisch, M., Deary, I. J., de Geus, E. J. C., den Braber, A., Delvecchio, G., Depondt, C., de Haan, L., de Zubicaray, G. I., Dima, D., Dimitrova, R., Djurovic, S., Dong, H., Donohoe, G., Duggirala, R., Dyer, T. D., Ehrlich, S., Ekman, C. J., Elvsåshagen, T., Emsell, L., Erk, S., Espeseth, T., Fagerness, J., Fears, S., Fedko, I., Fernández, G., Fisher, S. E., Foroud, T., Fox, P. T., Francks, C., Frangou, S., Frey, E. M., Frodl, T., Frouin, V., Garavan, H., Giddaluru, S., Glahn, D. C., Godlewska, B., Goldstein, R. Z., Gollub, R. L., Grabe, H. J., Grimm, O., Gruber, O., Guadalupe, T., Gur, R. E., Gur, R. C., Göring, H. H. H., Hagenaars, S., Hajek, T., Hall, G. B., Hall, J., Hardy, J., Hartman, C. A., Hass, J., Hatton, S. N., Haukvik, U. K., Hegenscheid, K., Heinz, A., Hickie, I. B., Ho, B.-C., Hoehn, D., Hoekstra, P. J., Hollinshead, M., Holmes, A. J., Homuth, G., Hoogman, M., Hong, L. E., Hosten, N., Hottenga, J.-J., Hulshoff Pol, H. E., Hwang, K. S., Jack, C. R., Jenkinson, M., Johnston, C., Jönsson, E. G., Kahn, R. S., Kasperaviciute, D., Kelly, S., Kim, S., Kochunov, P., Koenders, L., Krämer, B., Kwok, J. B. J., Lagopoulos, J., Laje, G., Landen, M., Landman, B. A., Lauriello, J., Lawrie, S. M., Lee, P. H., Le Hellard, S., Lemaître, H., Leonardo, C. D., Li, C.-s., Liberg, B., Liewald, D. C., Liu, X., Lopez, L. M., Loth, E., Lourdusamy, A., Luciano, M., Macciardi, F., Machielsen, M. W. J., MacQueen, G. M., Malt, U. F., Mandl, R., Manoach, D. S., Martinot, J.-L., Matarin, M., Mather, K. A., Mattheisen, M., Mattingsdal, M., Meyer-Lindenberg, A., McDonald, C., McIntosh, A. M., McMahon, F. J., McMahon, K. L., Meisenzahl, E., Melle, I., Milaneschi, Y., Mohnke, S., Montgomery, G. W., Morris, D. W., Moses, E. K., Mueller, B. A., Muñoz Maniega, S., Mühleisen, T. W., Müller-Myhsok, B., Mwangi, B., Nauck, M., Nho, K., Nichols, T. E., Nilsson, L.-G., Nugent, A. C., Nyberg, L., Olvera, R. L., Oosterlaan, J., Ophoff, R. A., Pandolfo, M., Papalampropoulou-Tsiridou, M., Papmeyer, M., Paus, T., Pausova, Z., Pearlson, G. D., Penninx, B. W., Peterson, C. P., Pfennig, A., Phillips, M., Pike, G. B., Poline, J.-B., Potkin, S. G., Pütz, B., Ramasamy, A., Rasmussen, J., Rietschel, M., Rijpkema, M., Risacher, S. L., Roffman, J. L., Roiz-Santiañez, R., Romanczuk-Seiferth, N., Rose, E. J., Royle, N. A., Rujescu, D., Ryten, M., Sachdev, P. S., Salami, A., Satterthwaite, T. D., Savitz, J., Saykin, A. J., Scanlon, C., Schmaal, L., Schnack, H. G., Schork, A. J., Schulz, S. C., Schür, R., Seidman, L., Shen, L., Shoemaker, J. M., Simmons, A., Sisodiya, S. M., Smith, C., Smoller, J. W., Soares, J. C., Sponheim, S. R., Sprooten, E., Starr, J. M., Steen, V. M., Strakowski, S., Strike, L., Sussmann, J., Sämann, P. G., Teumer, A., Toga, A. W., Tordesillas-Gutierrez, D., Trabzuni, D., Trost, S., Turner, J., Van den Heuvel, M., van der Wee, N. J., van Eijk, K., van Erp, T. G. M., van Haren, N. E. M., van 't Ent, D., van Tol, M.-J., Valdés Hernández, M. C., Veltman, D. J., Versace, A., Völzke, H., Walker, R., Walter, H., Wang, L., Wardlaw, J. M., Weale, M. E., Weiner, M. W., Wen, W., Westlye, L. T., Whalley, H. C., Whelan, C. D., White, T., Winkler, A. M., Wittfeld, K., Woldehawariat, G., Wolf, C., Zilles, D., Zwiers, M. P., Thalamuthu, A., Schofield, P. R., Freimer, N. B., Lawrence, N. S., Drevets, W., and the Alzheimer's Disease Neuroimaging Initiative IMAGEN Consortium, Saguenay Youth Study (SYS) Group, E. C. (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior*, 8(2):153–182.

Turnbull, A., Wang, H. T., Schooler, J. W., Jefferies, E., Margulies, D. S., and Smallwood, J. (2019). The ebb and flow of attention: Between-subject variation in intrinsic connectivity and cognition associated with the dynamics of ongoing experience. *NeuroImage*, 185(September 2018):286–299.

van Hooff, M. L. and van Hooft, E. A. (2014). Boredom at work: Proximal and distal consequences of affective work-related boredom. *Journal of Occupational Health Psychology*, 19(3):348–359.

Vodanovich, S. J., Verner, K. M., and Gilbride, T. V. (1991). Boredom proneness: Its relationship to positive and negative affect. *Psychological reports*, 69(3):1139–1146.

Webb, C. A., Israel, E. S., Belleau, E., Appleman, L., Forbes, E. E., and Pizzagalli, D. A. (2020). *Mind-wandering in adolescents predicts worse affect and is linked to aberrant default mode network – salience network connectivity*. American Academy of Child & Adolescent Psychiatry.

Wilson, T. D., Reinhard, D. A., Westgate, E. C., Gilbert, D. T., Ellerbeck, N., Hahn, C., Brown, C. L., and Shaked, A. (2014). Just think: the challenges of the disengaged mind. *Science*, 345(6192):75–7.

Supplementary Materials

Cohorts

A list and summary of the cohorts used in this study can be found in Supplementary Table 1.

Linear Mixed Effects Model

A large-scale linear mixed effects (LME) model was used to quantify the passage-of-time dysphoria observed in the online participants. The model is discussed in the Methods section, and many results are described in the Results section. Additional results are included below.

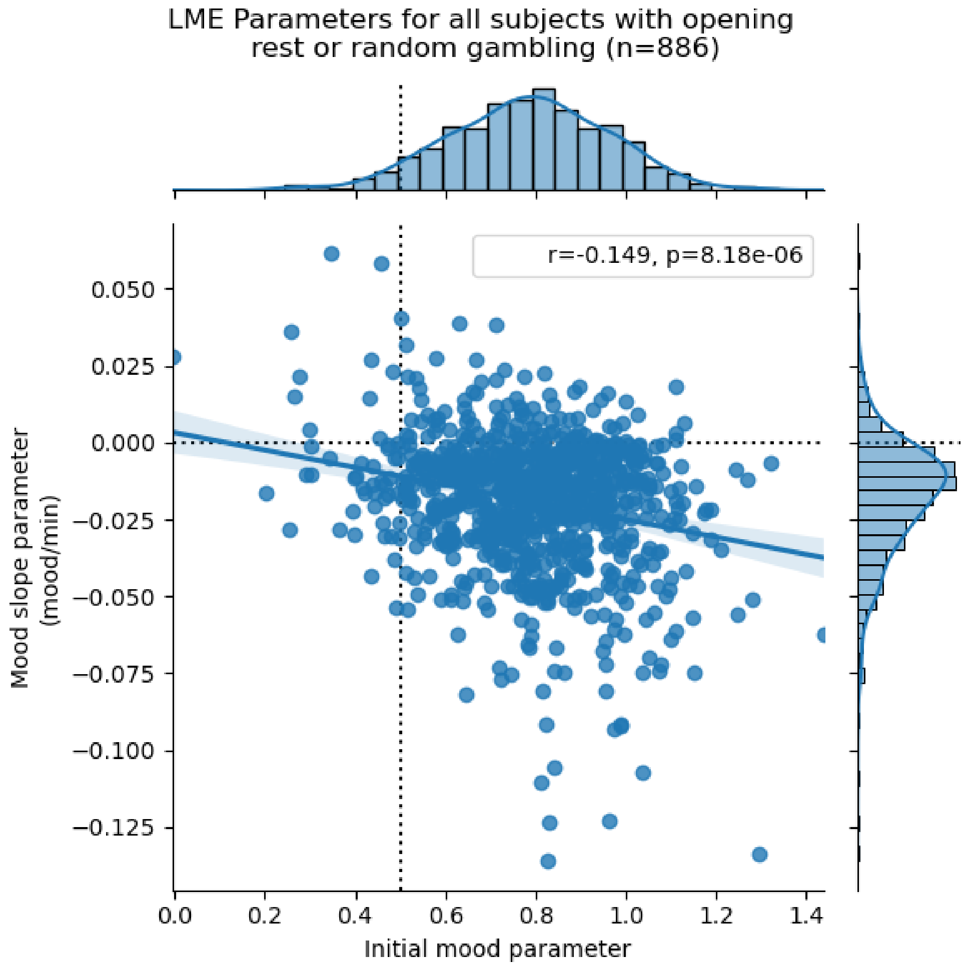

Figure 1: Joint plot of LME slope and intercept parameters for all online participants receiving opening rest periods. The r and p in the legend refer to a Spearman correlation.

Passage-of-Time Dysphoria's Uncertain Relationship to Age

Our large-scale LME model reported that participants with ages 16-18 had a significantly lower initial mood ($-8.8 \pm 2.8\% mood$, $t_{879} = -3.1, p = 0.002$) and higher slope ($0.9 \pm 0.4\% mood/min$, $t_{898} = 2.31, p = 0.021$) than those with ages 18-40. No other age group had significant differences in these parameters. The slope

Opening Rest Cohort	nParticipants	Block 0	Block 1	Block 2	Block 3
15sRestBetween	40	rest15 * 30	closed+ * 54		
30sRestBetween	37	rest30 * 18	closed+ * 54		
7.5sRestBetween	38	rest7.5 * 45	closed+ * 54		
60sRestBetween	39	rest60 * 10	closed+ * 54		
AlternateRating	32	rest15 * 30	closed+ * 54		
Expectation-7mRest	64	rest15 * 18	random * 22	closed- * 22	closed+ * 22
Expectation-12mRest	67	rest15 * 18	random * 22	closed- * 22	closed+ * 22
RestDownUp	58	rest15 * 18	closed- * 33	closed+ * 33	
Daily-Rest-01	66	rest15 * 18	closed+ * 18	rest15 * 18	closed+ * 18
Daily-Rest-02	53	rest15 * 18	closed+ * 18	rest15 * 18	closed+ * 18
Weekly-Rest-01	196	rest15 * 18	closed+ * 22	closed- * 22	closed+ * 22
Weekly-Rest-02	164	rest15 * 18	open+ * 22	open- * 22	open+ * 22
Weekly-Rest-03	160	rest15 * 18	open+ * 22	open- * 22	open+ * 22
Adolescent-01	116	rest15 * 18	closed+ * 22	closed- * 22	closed+ * 22
Opening Task Cohort					
Visuomotor	37	task15 * 30	closed+ * 54		
Visuomotor-Feedback	30	task15 * 30	closed+ * 54		
Opening Gambling Cohort					
RestAfterWins	25	closed+ * 54	rest15 * 30		
Daily-Closed-01	68	closed+ * 32	closed- * 32	closed+ * 32	
Daily-Random-01	66	random * 32	random * 32	random * 32	
App-Exploratory	5000	random * 30			
App-Confirmatory	21896	random * 30			
Follow-Up Cohorts					
BoredomBeforeAndAfter	150	rest15 * 18	closed- * 33	closed+ * 33	
BoredomAfterOnly	150	rest15 * 18	closed- * 33	closed+ * 33	
MwBeforeAndAfter	150	rest15 * 18	closed- * 33	closed+ * 33	
MwAfterOnly	150	rest15 * 18	closed- * 33	closed+ * 33	
Activities	450	break420 * 1	closed- * 33	closed+ * 33	

Table 1: A list and description of cohorts collected. nParticipants contains the number of participants who completed both the task and survey in this cohort. The columns beginning with "Block" denote the type, parameter, and number of trials used in that block of trials. "Rest" denotes looking at a fixation cross, and "task" denotes a simple visuomotor task in which a cross moves predictably across the screen and the subject is asked to press a button when it crosses the center line. The number that follows these labels is the time in seconds between mood ratings. "Break" denotes a free period where participants could leave to do anything they chose. "Closed" and "random" denote the closed-loop and random gambling task conditions described in the Methods section. ("open" denotes open-loop gambling not described in this paper; these blocks were not used in analyses). The + or - after the "closed" label indicates whether mood was being manipulated upwards (+) or downwards (-). The number after the * indicates how many trials of this type were included in the block. Certain cohort names also contain information. The AlternateRating cohort rated their mood with a single button press rather than moving a slider. The Expectation cohorts received opening instructions stating that the upcoming rest period would be up to 7 minutes or 12 minutes. Groups beginning with "Daily" or "Weekly" returned 1 day or 1 week apart to complete a similar task again (e.g., the Daily-Rest-02 cohort is the same participants as Daily-Rest-01, returning to complete the same task one day later). The Adolescent-01 cohort is a group of adolescents recruited in person rather than on Amazon Mechanical Turk.

Factor	Estimate	2.5_ci	97.5_ci	SE	DF	T-stat	P-val	Sig
(Intercept)	0.784	0.756	0.812	0.0141	875	55.6	$< 10^{-6}$	***
Time	-0.0189	-0.0226	-0.0153	0.00185	864	-10.3	$< 10^{-6}$	***
isMale	-0.0144	-0.0395	0.0107	0.0128	877	-1.12	0.262	
meanIRIOver20	0.000698	-0.000585	0.00198	0.000655	901	1.07	0.287	
totalWinnings	-0.000332	-0.00435	0.00369	0.00205	898	-0.162	0.872	
meanRPE	0.158	-0.0104	0.326	0.0859	898	1.84	0.0662	.
fracRiskScore	-0.186	-0.202	-0.169	0.00828	877	-22.4	$< 10^{-6}$	***
isAge0to16	-0.0456	-0.108	0.0168	0.0318	879	-1.43	0.152	
isAge16to18	-0.0883	-0.144	-0.0325	0.0285	879	-3.1	0.002	**
isAge40to100	-0.00712	-0.0351	0.0208	0.0143	877	-0.5	0.617	
Time:isMale	0.00159	-0.00171	0.00488	0.00168	869	0.944	0.345	
Time:meanIRIOver20	-0.000103	-0.000267	$6.1 * 10^{-5}$	$8.4 * 10^{-5}$	810	-1.23	0.219	
Time:totalWinnings	$-1.9 * 10^{-5}$	-0.000566	0.000529	0.00028	$1.04 * 10^{3}$	-0.0664	0.947	
Time:meanRPE	-0.00743	-0.0304	0.0155	0.0117	$1.05 * 10^{3}$	-0.634	0.526	
Time:fracRiskScore	0.00515	0.00303	0.00728	0.00109	869	4.75	$2 * 10^{-6}$	***
Time:isAge0to16	-0.00144	-0.00967	0.00678	0.0042	895	-0.344	0.731	
Time:isAge16to18	0.00869	0.00131	0.0161	0.00376	898	2.31	0.0212	*
Time:isAge40to100	0.00302	-0.000638	0.00668	0.00187	865	1.62	0.106	

Table 2: Results of the LME model trained on all naïve online adult and adolescent participants who received opening rest, visuomotor task, or random gambling periods; as produced by the pymer software package. The first column lists each factor in the model as described in the Methods section. Factors beginning with "is" are binary (0 or 1). "Time" is the mood slope parameter we use to quantify passage-of-time dysphoria. Mood ratings ranged from 0-1, and time was in minutes. totalWinnings and meanRPE were in points, whose monetary value is unknown to naïve subjects. fracRiskScore was the score on a clinical depression questionnaire divided by a clinical cutoff. Age was in years. Factors preceded by "Time:" indicate the interaction of that parameter and the elapsed time. The next four columns describe the effect size: "Estimate" is the estimated coefficient of each factor in the model, 2.5 and 97.5 ci are the 95 percent confidence interval of the estimate, and SE is its standard error. DF is the degrees of freedom, T-stat is the t statistic, and P-val is the p value. All values are rounded to 3 decimal places. The Sig (significance) column contains . if p<0.1, * if p<0.05, ** if p<0.01, and *** if p<0.001.

parameters produced by an LME without age factors included are plotted against age in Supplementary Figure 2. The relationship between age and mood slope was not clear from these plots; more research will be required to clarify the relationship between passage-of-time dysphoria and age.

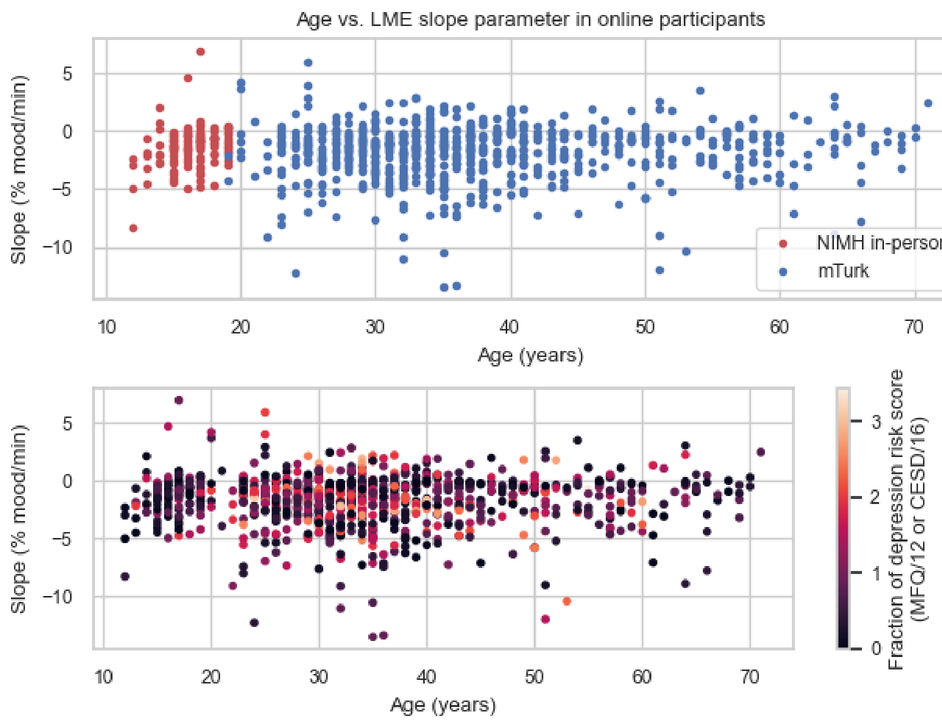

Figure 2: Mood slopes (produced by an LME model with age-related terms removed) plotted against participant age.

Eliminating Methodological Confounds

Because this finding is new, we wanted to examine the impact of possible methodological confounds. We therefore created slightly modified versions of the task to see whether the observed decline in mood ratings might be due to:

1. The aversive nature of rating one's mood
2. The method of rating mood and its susceptibility to fatigue
3. The expected duration of the rest period
4. The aversive nature of multitasking or task switching

Passage-of-Time Dysphoria Is Not a Product of Aversive Mood Ratings

To investigate whether the decline in mood might be driven by the ratings themselves, we varied the frequency of mood ratings. We reasoned that, if mood ratings were decreasing mood, more frequent ratings would cause mood to decline more quickly. We observed that participants with 60 s, 30 s, 15 s, and 7.5 s of rest between ratings (cohorts 60sRestBetween, 30sRestBetween, 15sRestBetween, and 7.5sRestBetween, in Table 1) all had mood ratings that declined at roughly the same rate (Figure 1C). This finding was later confirmed by our multi-cohort LME model, in which a participant's mean inter-rating interval did not have a significant relationship with their slope parameter ($-0.0103 \pm 0.0084\% mood$, $t_{810} = -1.23, p = 0.219$. From this, we conclude that mood ratings were not aversive enough that an increase in mood rating frequency led to an increase in passage-of-time dysphoria.

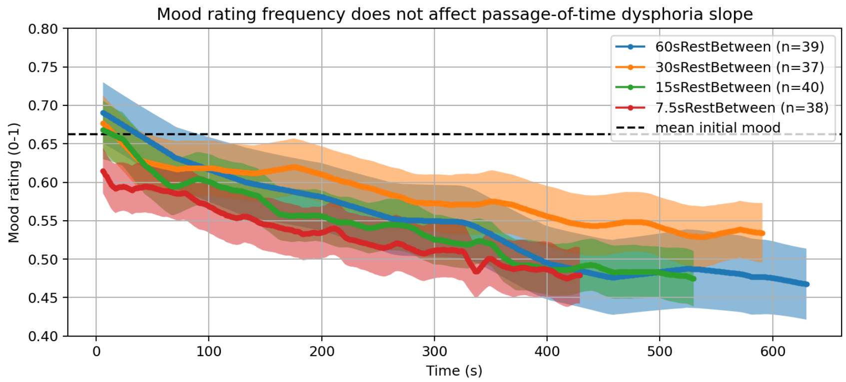

Figure 3: The magnitude of passage-of-time dysphoria did not vary with the frequency of mood ratings.

Passage-of-Time Dysphoria Is Not an Artefact of the Rating Method

Participants had thus far rated their mood with a slider that started in the middle of the scale (0.5). We therefore wondered whether participants' mood ratings were converging on 0.5 because they were becoming more fatigued and ratings near the middle of the slider required the least effort. In another modified version of the task, we asked participants (cohort AlternateRating in Table 1) to press a single number key (1-9) to indicate their happiness during the mood ratings, where 1 was "unhappy" and 9 was "happy". In this way, we made each mood require roughly equal time and effort. We found that LME slope parameters collected from this task were not significantly different from those of the original cohort ($t_{70} = 0.427, p = 0.671$).

Passage-of-Time Dysphoria Is Not Driven by Expectations

We examined whether the mood ratings might be affected by the expected duration of the rest period. This would suggest that the dysphoria observed during rest was a product of rumination about the amount of rest time remaining. To test this, we gave identical tasks to two groups, preceded by slightly different instructions: one was told that the initial rest period would be up to 7 minutes (cohort Expectation-7mRest, $n = 64$), and the other was told it would be up to 12 minutes (cohort Expectation-12mRest, $n = 67$). After these instructions, both groups actually received rest periods of approximately 6.4 minutes. LME slope parameters were not significantly different between these two groups ($t_{104} = 0.185, p = 0.854$).

Passage-of-Time Dysphoria Is Not Driven by MultiTasking

Passage-of-time dysphoria's generalizability across task conditions speaks to the concern that online participants were multitasking on their computers or phones during rest periods. Online participants included in the large-scale LME moved or locked in their mood rating slider on 97.7% of rest trials, suggesting that any multitasking was not so engaging as to stop them from noticing the next mood rating. Cohorts with short rest periods or task demands between mood ratings likely had to make responses too frequently to multitask, but their level of passage-of-time dysphoria was not significantly different from the cohorts with longer rest periods. This evidence does not rule out that people were multitasking, but it suggests that any multitasking taking place did not reliably change the observed levels of dysphoria.

Regression to the Mean

We were concerned that our results concerning depression and passage-of-time dysphoria might be an artefactual result of regression to the mean: for a purely random process, values starting high will tend to go down over time, and values starting low will tend to go up over time. Thus, slope parameters might be less negative for people with higher depression risk simply because their initial mood happened to be lower. We examined two pieces of evidence to investigate this possibility: stability and time-of-day effects.

Stability Over Time

First, we examined the stability of the LME intercept and slope parameters within an individual. One cohort (Daily-Rest-01 in Supplementary Table 1) repeated a task with a rest block lasting 6.8 minutes on average, a closed-loop positive gambling block lasting 3.5 minutes on average, another 6.8-minute rest block, and another 3.5-minute closed-loop positive gambling block. This cohort was invited to return the following day to complete the same task again (Daily-Rest-02). This allowed us to assess stability both (a) across blocks within a run, and (b) across days. A second cohort (Weekly-Rest-01) completed an initial rest block lasting 6.8 minutes on average, followed by three 4.3-minute closed-loop gambling blocks (1 positive, 1 negative, 1 positive). They were invited back one and two weeks later to complete the same task again (Weekly-Rest-02/03). This allowed us to assess stability across weeks.

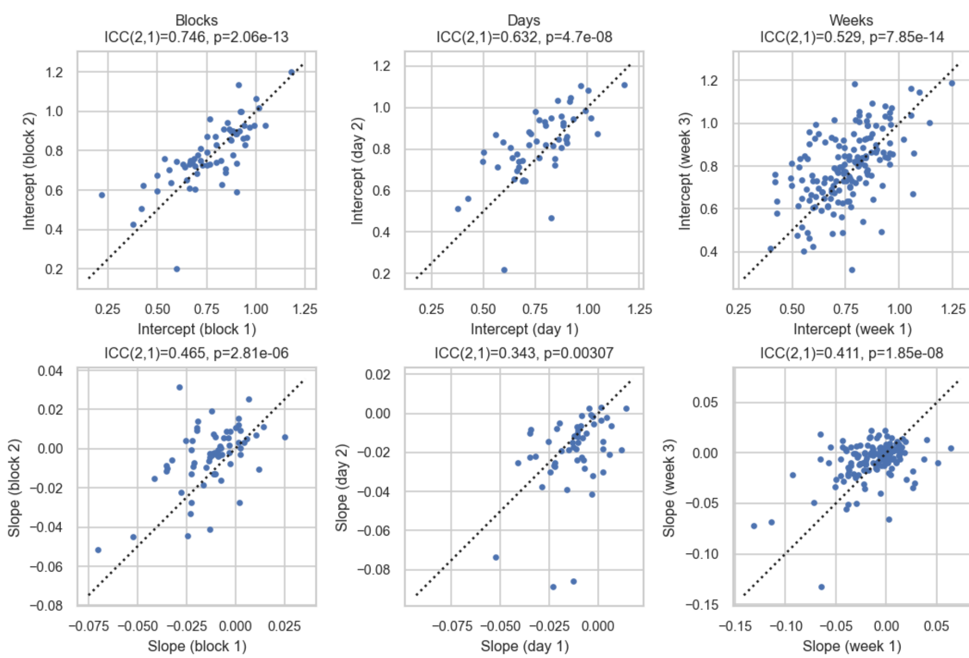

Figure 4: Stability of LME coefficients estimating the initial mood (top) and slope of mood over time (bottom) for each participant across rest periods one block apart (left), 1 day apart (middle), and 2 weeks apart (right). ICC denotes the intra-class correlation coefficient for each comparison.

The LME intercept parameter (i.e., initial mood) showed high stability across blocks ($ICC(2,1) = 0.746, p = 2.1 * 10^{-13}$), days ($ICC(2,1) = 0.632, p = 4.7 * 10^{-8}$), and weeks ($ICC(2,1) = 0.529, p = 7.9 * 10^{-14}$), confirming the stability of subjective momentary mood ratings. The Slope parameter showed moderate stability that was statistically significant, across blocks ($ICC(2,1) = 0.465, p = 2.8 * 10^{-6}$), days ($ICC(2,1) = 0.343, p = 3.1 * 10^{-3}$), and weeks ($ICC(2,1) = 0.411, p = 1.9 * 10^{-8}$). Scatter plots are shown in Supplementary Figure 4. This level of stability suggests that inter-individual differences in initial mood and slope are driven by stable traits rather than random fluctuations.

Time-of-Day Effects

We also examined the specific effect of time of day on mood. Past research has shown that affective ratings vary consistently with time of day, with reports of pleasantness being lowest in the morning and highest in the evening (Egloff et al., 1995). Time of day also impacts loss sensitivity during risky decision-making (Bedder et al., 2020). If time of day were related to initial mood or mood slope, our individual difference results could possibly be explained by depressed individuals participating at different times of day than non-depressed participants. In the dataset of online participants, however, we did not observe a significant relationship between the time of day when the task was completed and the intercept or slope parameter (Supplementary Figure 5). This suggests that inter-individual differences in initial mood and slope were not driven by periodic daily fluctuations in mood.

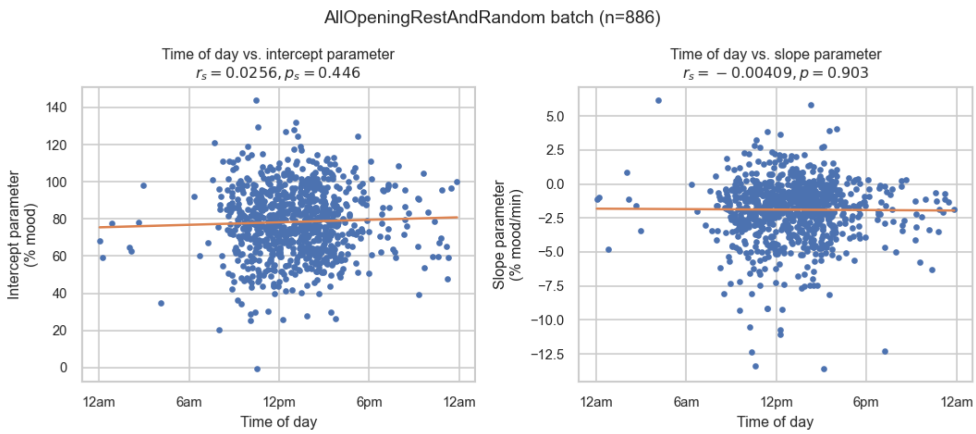

Figure 5: Intercept and slope parameters learned by the LME model, plotted against time of day in the online cohorts.

Examining Floor Effects in the Depression-Time Interaction

Individuals reporting greater depressive symptoms on average reported lower initial mood at the onset of the task. If their mood declined further, they therefore had less of the mood scale available to them to express it. This could lead to "floor effects" where the mood of depressed individuals appears to decline more slowly with time simply because they have reached the bottom of the scale and are forced to level out.

In a sensitivity analysis, we excluded the 27/600 participants in the follow-up cohorts (See Supplementary Table 1) who reached the floor of the mood scale (i.e., mood = 0) at any time during the rest period. We then re-fit the LME model of mood. The significant effect of the interaction between depression risk and time (i.e., the relationship between depression risk and passage-of-time dysphoria) persisted in this analysis. ($t_{566} = 4.06, p = 5.65e - 5$). Thus, the effect is not driven by depressed participants reaching the absolute minimum of the scale.

We also considered whether participants might be reluctant to reach the floor of the scale but could still reach a sort of "individual" mood floor, a point under which they would be reluctant to rate themselves. In our follow-up cohorts, rest periods were followed a period of negative mood induction (via increasing the probability of monetary losses in a block of trials). We have demonstrated before (Keren et al., 2021) that this form of mood induction produces potent changes in mood with effect sizes of Cohen's $d = -1.75$. We took the lowest point during this mood induction to represent a (conservative) individual mood floor. This allowed us to check whether participants reached an individual mood floor during the preceding rest period. In a sensitivity analysis, we excluded the 101/600 participants who reached such an "individual mood floor" (i.e., we excluded all those participants who during resting state reached the minimum mood that they had reached during the negative mood induction). This sensitivity analysis also had minimal effect on our results, in which the interaction effect of depression risk and time remained significant. ($t_{493} = 3.43, p = 6.65e-4$).

Computational Model

Our computational model was based on the one described and validated in (Keren et al., 2021), which accurately modeled subjective mood ratings in a very similar gambling game. The computational model fit the data well for most of our mobile app participants. In the tuning step, the hyperparameters minimizing testing loss were determined to be $\lambda_{EA} = 0.483, \lambda_T = 33.6$. The relationship between these hyperparameters and the smoothed testing loss is shown in Figure 6.

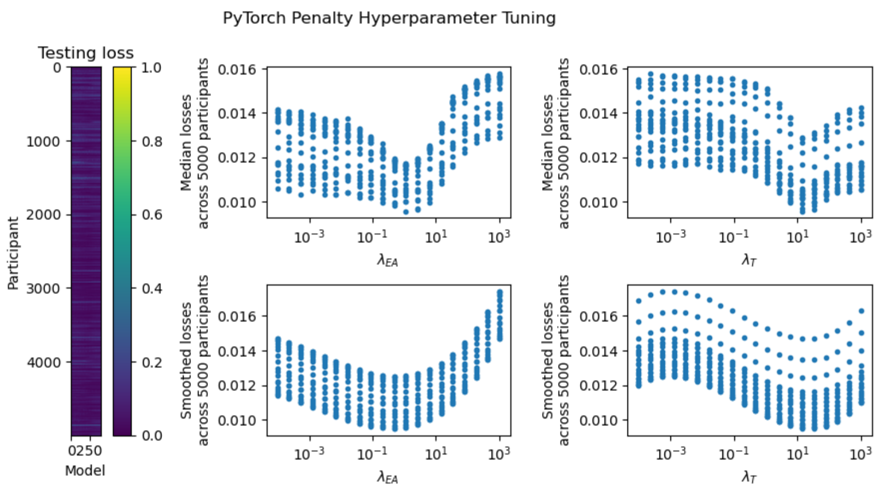

Figure 6: Tuning of penalty term hyperparameters. The two penalty parameters λ_{EA} and λ_T were varied systematically, and the computational model was fit to all but the final two ratings for each participant. Top graphs show the median testing loss (i.e., the sum of squared errors on the final two ratings) across participants. Bottom graphs show these same losses after smoothing with a polynomial fit. The parameters with the lowest smoothed loss on this exploratory mobile app cohort were used in the final model fit to the confirmatory mobile app cohort.

When using these hyperparameters, the median testing loss (defined as the mean squared errors for the 2 testing trials) across the 5,000 exploratory/tuning participants used to tune parameters was 0.00486. When those hyperparameters were used on the 21,896 confirmatory app participants, the median loss on testing trials was 0.00325. The mean (across participants) Spearman correlation coefficient between each participant's model fits and actual mood ratings was $r_s = 0.715$, 95% CI = (0.754,0.759).

Sample fits are shown in Supplementary Figure 7. Histograms of the learned parameters are shown in Supplementary Figure 8. Relationships between β_T and the other model parameters are shown in Supplementary Figures 9 and 10.

Linking Subjective Momentary Mood Ratings to Life Happiness Ratings

To measure the psychometric validity of the subjective momentary mood ratings, we correlated the initial mood (or "Intercept") parameter of the online cohort's LME model (left) and the mobile app cohort's computational model (right) with the life happiness ratings. Results showed that both estimates of initial mood correlated significantly with ratings of life happiness (Supplementary Figure 11)

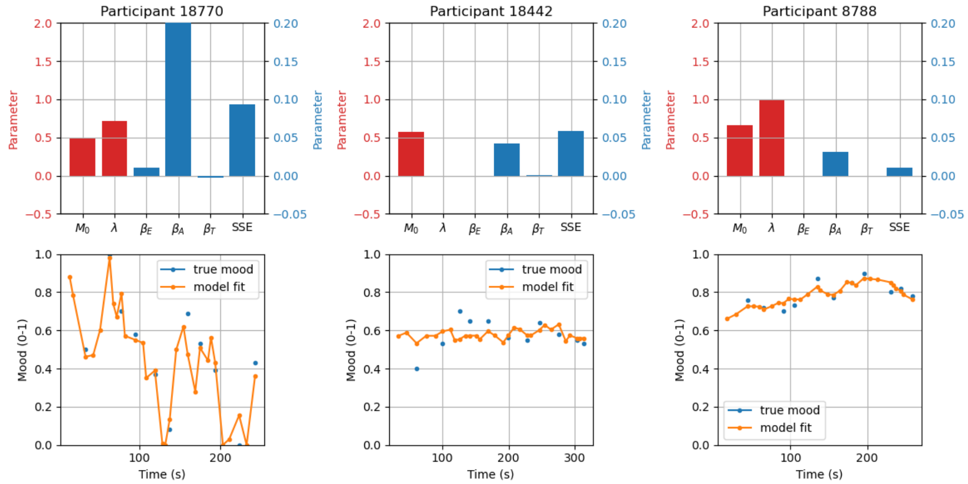

Figure 7: Sample fits of the computational model for three random subjects in the confirmatory mobile app cohort. SSE = sum squared error, a measure of goodness of fit to the training data. In the top plots, the red bars are in units of the left-hand y axis, and the blue bars are in units of the right-hand y axis.

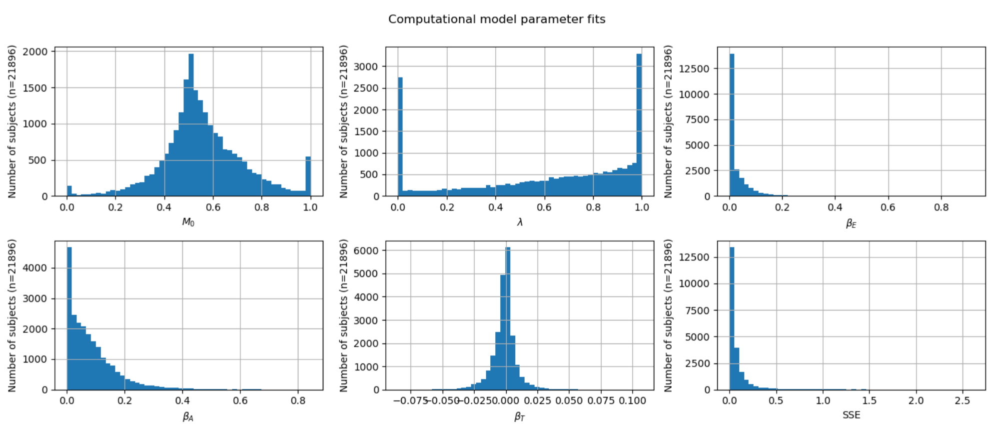

Figure 8: Histogram of computational model parameters across the 21,896 confirmatory mobile app subjects.

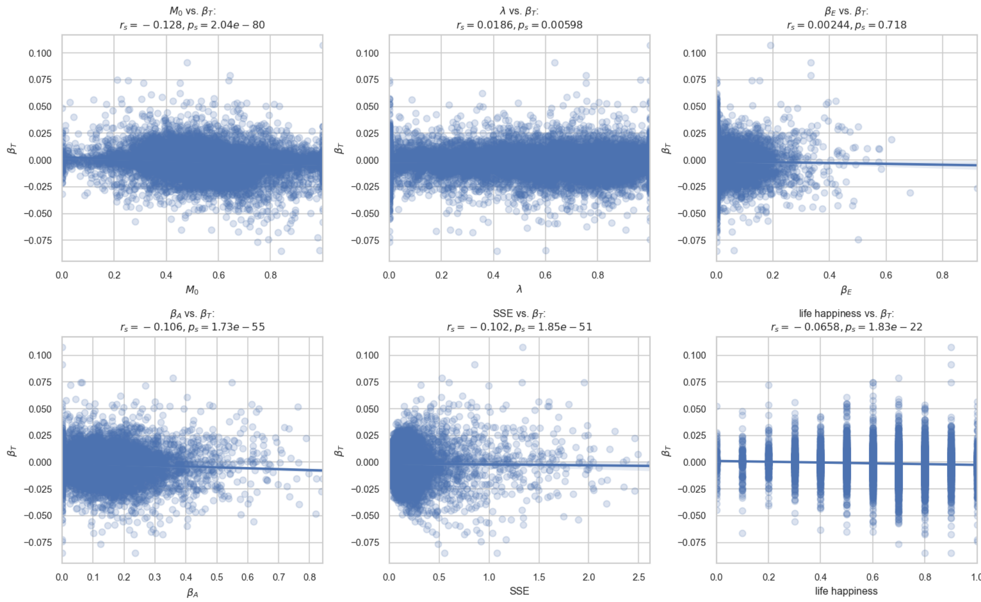

Figure 9: Time sensitivity parameter β_T vs. other parameters in the confirmatory mobile app cohort.

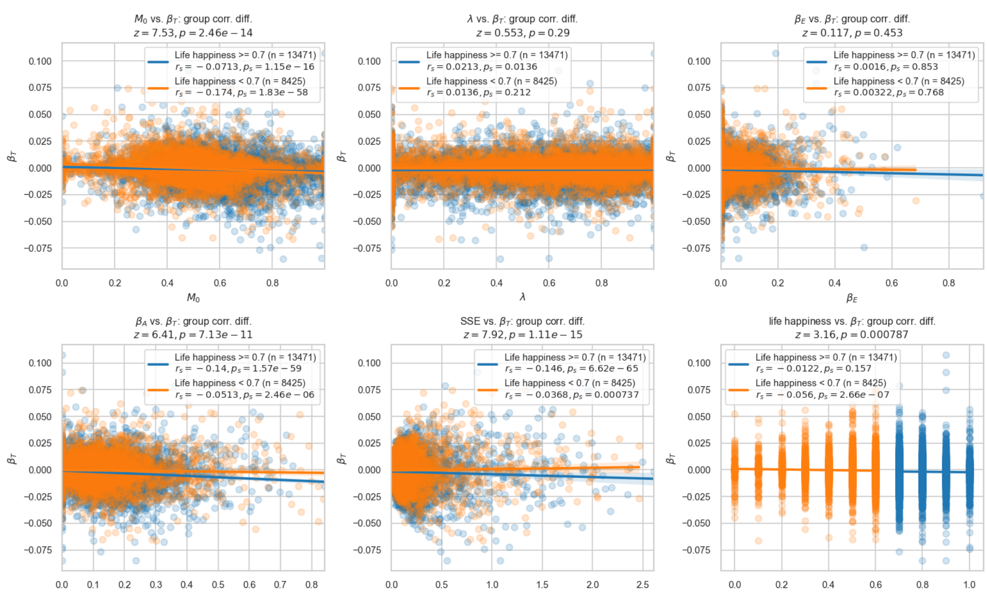

Figure 10: Time sensitivity parameter β_T vs. other parameters in the confirmatory mobile app cohort, in 2 groups separated by high (blue) or low (orange) life happiness.

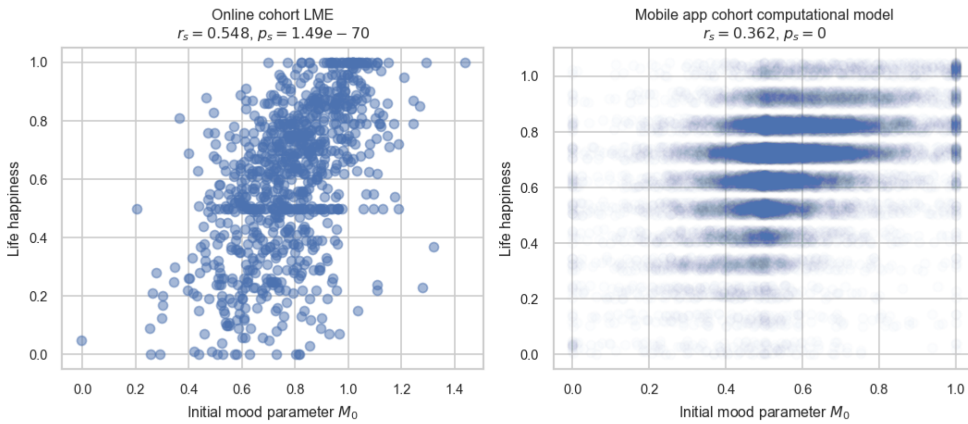

Figure 11: Initial mood parameter vs. life happiness rating in the online cohort (left) and the confirmatory mobile app cohort (right).

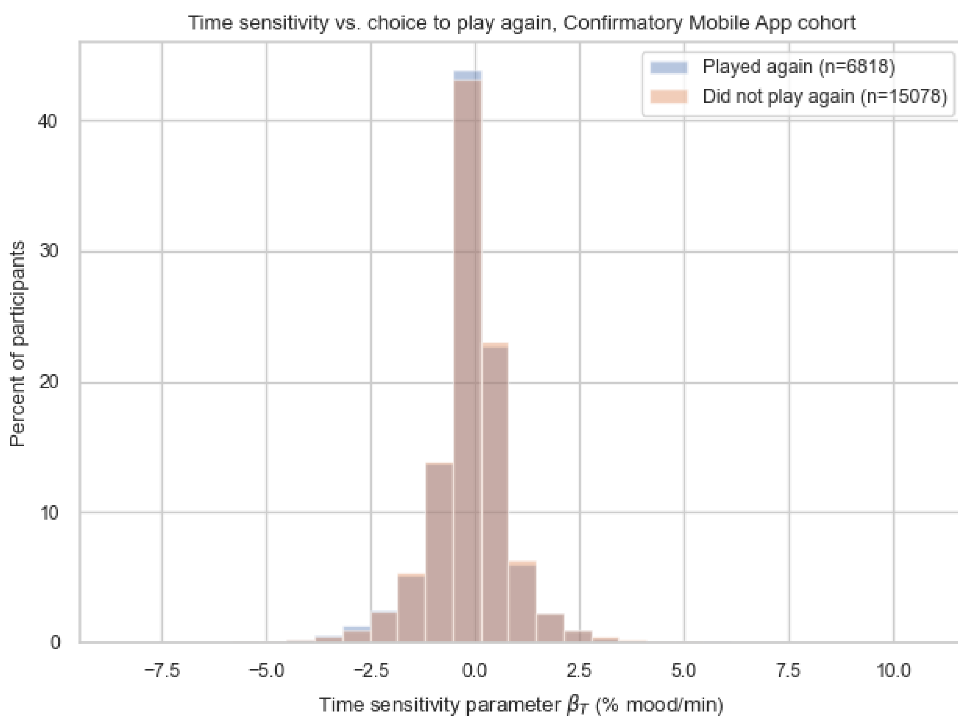

Figure 12: Histogram of the computational model time sensitivity parameter for subsets of the confirmatory mobile app cohort that chose to play again later (blue) and those that did not (orange). No significant difference in the distributions was observed (2-sided Wilcoxon rank-sum test, $W_{21894} = 0.804, p = 0.402$).

Impact of Methodological Choices on Mobile App Slope Estimates

Results showed that mobile app participants experienced significantly less passage-of-time dysphoria than online participants. This difference is larger if we use the computational model's time sensitivity parameter rather than the LME analysis' slope parameter. This is likely related to the regularization hyperparameter used in the computational model but not the LME analysis. If an LME analysis is used on both cohorts instead of the computational model, the difference between the two groups' medians is considerably smaller, shrinking from $1.49\% mood/min$ to $0.774\% mood/min$ (Supplementary Figure 13). It is also possible that participants experiencing greater dysphoria "self-selected" out of the mobile app game: frustrated mobile app participants could exit at any time without penalty, whereas online participants would lose compensation if they dropped out. However, no relationship was observed between the time sensitivity parameter of our computational model and the number of times a participant played the game (Supplementary Figure 12). Finally, since no participants are known to have participated in both experiments, we cannot rule out more general cohort effects: the participants choosing to play the mobile app game could simply have different sensitivity to time on task than those participating in the online experiment.

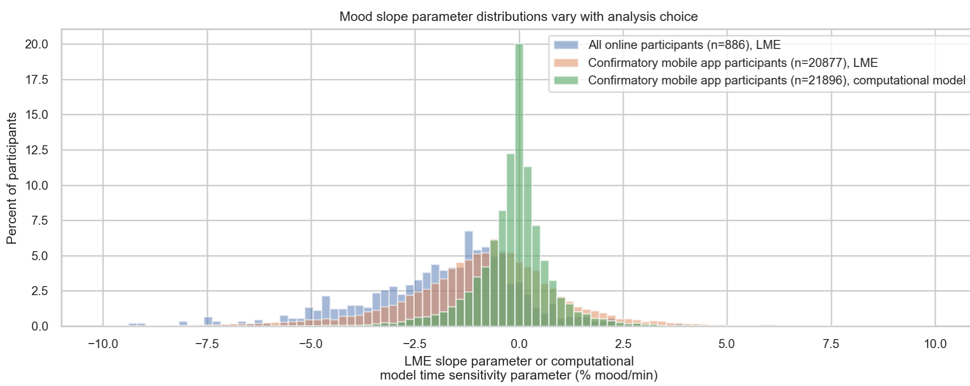

Figure 13: Histogram of the LME mood slope parameters for the online cohort (blue) and the confirmatory mobile app cohort (orange), along with the computational model time sensitivity parameter for the confirmatory mobile app cohort (green). Mobile app participants with outlier task completion times were excluded from the LME analysis (see methods section). Note that the use of LME modeling to analyze the mobile app data significantly lowered the distribution of slopes compared to when the computational model was used (2-sided Wilcoxon rank-sum test, $W_{42771} = -54.2, p < 2.23 * 10^{-308}$), but the LME slopes from the mobile app were still significantly greater than those of the online cohort (2-sided Wilcoxon rank-sum test, $W_{21761} = 14.5, p = 2.03 * 10^{-47}$)

Sensitivity analysis: Excluding First Rating

We chose to include the first mood rating in our linear trend estimation, despite the fact that this rating appeared to be an outlier in our exploratory cohort's computational model fits (Supplementary Figure 14, left). To check the sensitivity of our conclusions to this choice, we performed the same analyses while excluding this first mood rating from our model fitting procedure.

In our confirmatory cohort, this pattern (in which the first rating was an outlier) was not observed (Supplementary Figure 14, right). Nevertheless, we preregistered this sensitivity analysis, and we therefore report the results for the confirmatory cohort below.

- Model tuning:
  - best fitting penalty hyperparameters (model WITH β_T): [$\lambda_{EA} = 0.483$, $\lambda_T = 33.6$]
  - best fitting penalty coefficients (model WITHOUT β_T): $\lambda_{EA} = 0.207$

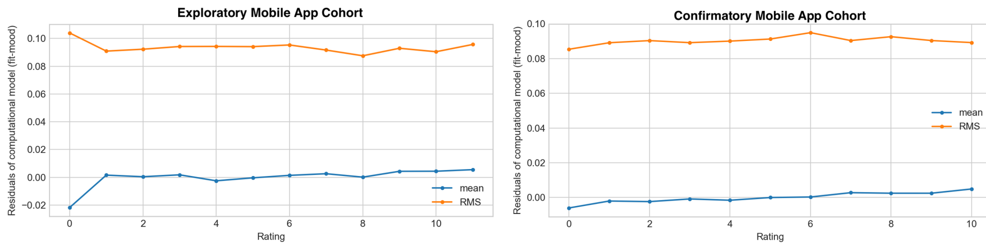

Figure 14: Mean (blue) and root-mean-square (RMS, orange) residuals across the exploratory (left) and confirmatory (right) mobile app subjects of the computational model fit for each rating number. In the exploratory cohort, the first rating appeared to be an outlier, inspiring our preregistered sensitivity analysis. In the confirmatory cohort (right), this pattern was not observed. But we still report our preregistered sensitivity analysis on the confirmatory cohort.

  - median MSE (model WITH β_T): 0.0032388
  - median MSE (model WITHOUT β_T): 0.0033644
  - 2-sided Wilcoxon sign-rank test on the difference between models with and without β_T: $W_{499} = 0.0, p = 1.26 * 10^{-83}$
- Distribution of β_T:
  - Mean ± standard error β_T: -0.129% mood/min ± 0.00667
  - 2-sided Wilcoxon sign-rank test on β_T vs. 0: $W_{21895} = 1.00 * 10^8, p = 5.15 * 10^{-90}$
  - 2-sided Wilcoxon rank-sum test of LME time coefficients vs. Computational Model β_T: $W_{42771} = -18.4, p = 9.47 * 10^{-76}$
- Individual differences:
  - life happiness vs. β_T: $r_s = -0.0654$, $p_s = 3.24 * 10^{-22}$
  - β_A vs. β_T: $r_s = -0.106$, $p_s = 2.23 * 10^{-55}$
  - β_A vs. β_T (life happiness>=0.7): $r_s = -0.140$, $p_s = 1.63 * 10^{-59}$
  - β_A vs. β_T (life happiness<0.7): $r_s = -0.0510$, $p_s = 2.78 * 10^{-06}$
  - β_A vs. β_T correlation difference between high and low life happiness groups: $z = 6.43$, $p = 6.32 * 10^{-11}$

Freely Chosen Activities Varied Widely

Participants were allowed to choose their own activities during a 7-minute rest period, as described in the main text. Afterwards, participants could indicate how much time they spent on each activity using a slider ranging from "Not at all" (scored at 0%) to "The whole time" (scored at 100%). Their rating of each activity (in the order in which they were rated) is shown in Table 3.

Results of Preregistration on Boredom, Mind-Wandering, and Freely Chosen Activities

We performed a follow-up set of preregistered tasks and analyses on boredom, mind-wandering, and freely chosen activities (https://osf.io/gt7a8). The methods and results section of the main text describe the motivation and detailed results of each piece of this preregistration. Below, we reproduce the hypotheses enumerated in that preregistration. We follow each with a concise summary of whether the hypothesis was supported by the data.

1.1) In the validation of short interval state boredom scale repeat administration, we hypothesize that the effect of including an initial administration will have an absolute effect size (cohen's d) less than 0.5. We will test this with two, one-sided t-tests (TOST). With an alpha of 0.01 and sample size of 150 participants per

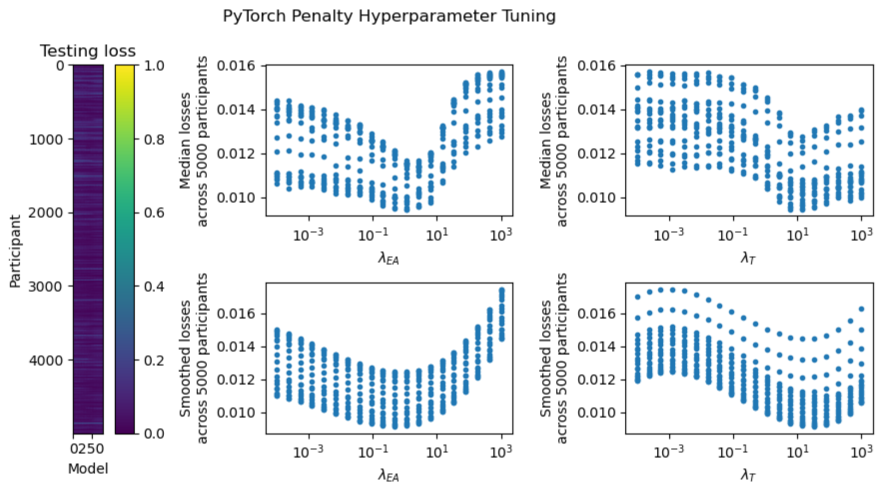

Figure 15: Sensitivity analysis with first rating excluded from model fit: Tuning of penalty term hyperparameters. The two penalty parameters λ_{EA} and λ_T were varied systematically, and the computational model was fit to all but the final two ratings for each participant. Top graphs show the median testing loss (i.e., the sum of squared errors on the final two ratings) across participants. Bottom graphs show these same losses after smoothing with a polynomial fit. The parameters with the lowest smoothed loss on this exploratory mobile app cohort were used in the final model fit to the confirmatory mobile app cohort.

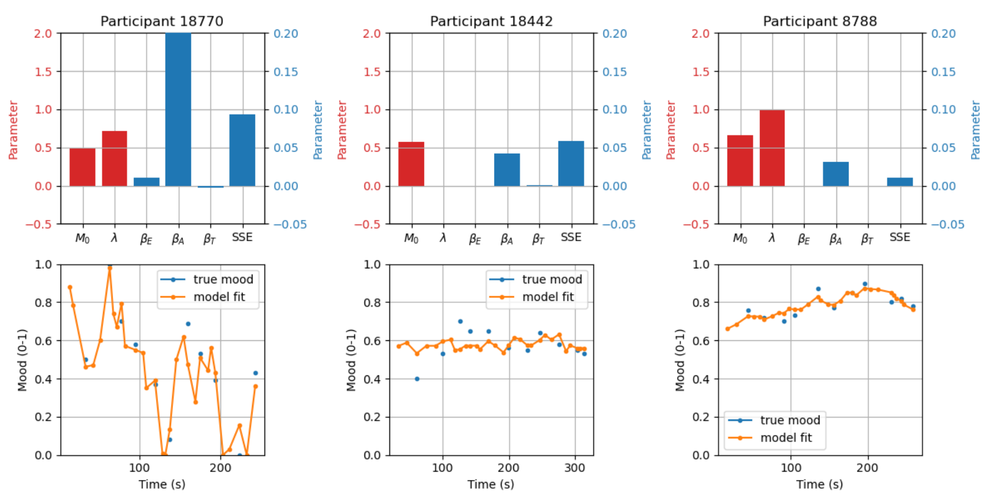

Figure 16: Sensitivity analysis with first rating excluded from model fit: Sample fits of the computational model for three random subjects. SSE = sum squared error, a measure of goodness of fit to the training data. In the top plots, the red bars are in units of the left-hand y axis, and the blue bars are in units of the right-hand y axis.

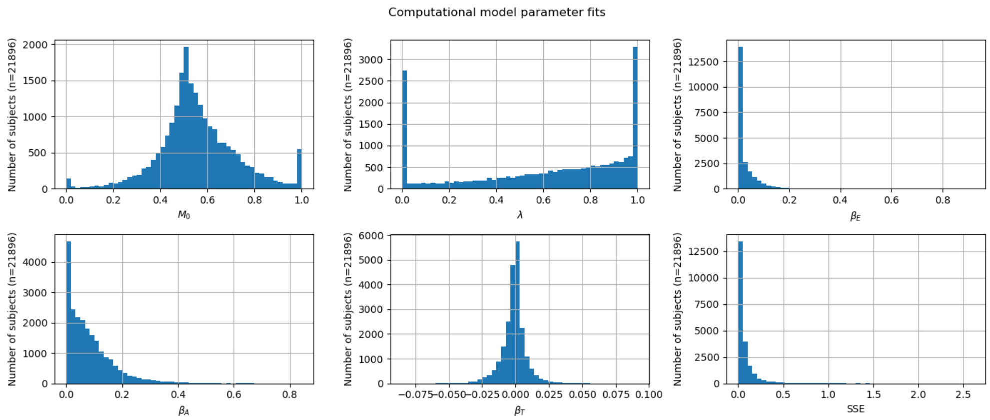

Figure 17: Sensitivity analysis with first rating excluded from model fit: Histogram of computational model parameters across the confirmatory mobile app subjects.

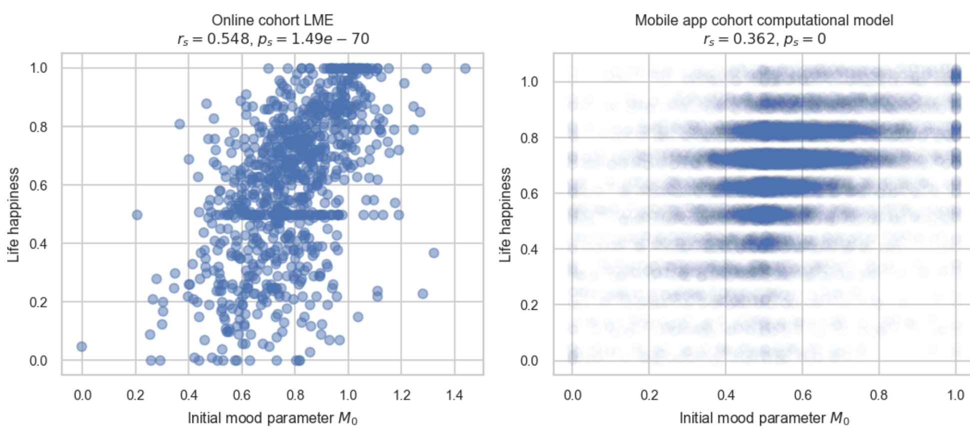

Figure 18: Sensitivity analysis with first rating excluded from model fit: Initial mood parameter vs. life happiness rating in the online cohort (left) and the confirmatory mobile app cohort (right).

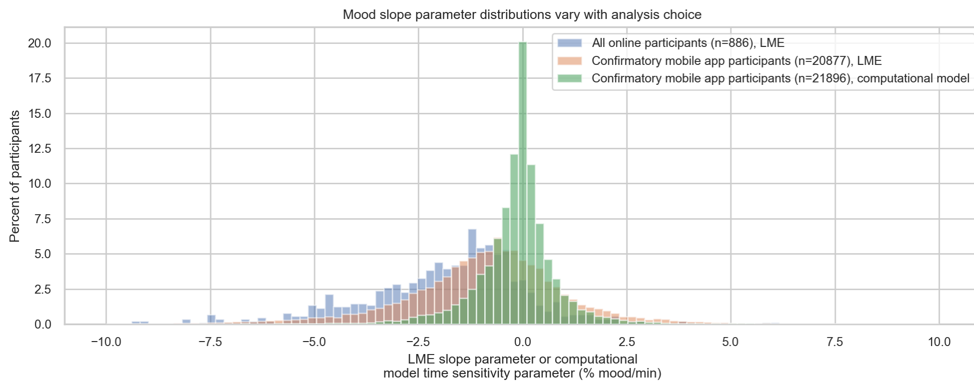

Figure 19: Sensitivity analysis with first rating excluded from model fit: Histogram of the LME mood slope parameters for the online cohort (blue) and the confirmatory mobile app cohort (orange), along with the computational model time sensitivity parameter for the confirmatory mobile app cohort (green). Mobile app participants with an inter-rating interval (IRI) > 38 seconds were excluded from analysis. Note that the use of LME modeling to analyze the mobile app data significantly lowered the distribution of slopes compared to when the computational model was used (2-sided Wilcoxon rank-sum test, $W_{42771} = -18.4, p = 9.47 * 10^{-76}$), but the LME slopes from the mobile app were still significantly greater than those of the online cohort (2-sided Wilcoxon rank-sum test, $W_{21761} = 18.9, p = 1.33 * 10^{-79}$)

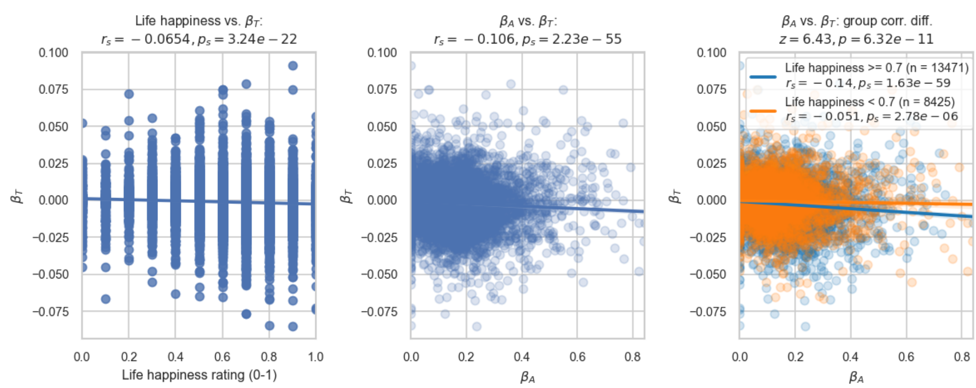

Figure 20: Sensitivity analysis with first rating excluded from model fit: Individual differences in sensitivity to the passage of time relate to other individual differences. The computational model's time sensitivity parameter β_T for each participant in the confirmatory mobile app cohort is plotted against that participant's life happiness rating and their reward sensitivity parameter β_A.

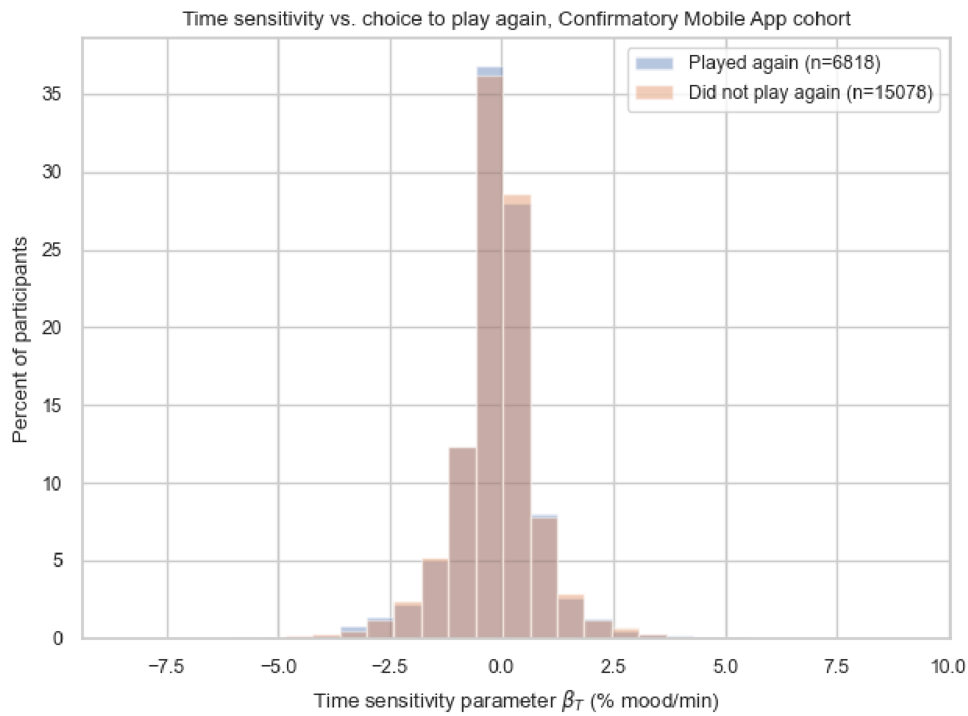

Figure 21: Sensitivity analysis with first rating excluded from model fit: Histogram of the computational model time sensitivity parameter for subsets of the confirmatory mobile app cohort that chose to play again later (blue) and those that did not (orange). No significant difference in the distributions was observed (2-sided Wilcoxon rank-sum test, $W_{21894} = 0.838, p = 0.402$).

Order	Activity	Frequency
1.	I thought.	50.2%
2.	I consumed the news.	28.2%
3.	I looked at photos.	20.2%
4.	I listened to music, podcasts, or radio.	23.5%
5.	I did some work for my (non-MTurk) job.	16.3%
6.	I looked for a (non-MTurk) job.	10.4%
7.	I paid bills, banked, or invested.	10.2%
8.	I did something else on my computer or phone.	44.7%
9.	I read texts or emails.	22.5%
10.	I wrote something.	12.2%
11.	I watched videos.	18.5%
12.	I went on social media.	20.3%
13.	I shopped.	9.44%
14.	I did something on MTurk.	15.4%
15.	I called/videochatted with someone.	8.22%
16.	I played a computer/phone game.	13.6%
17.	I did something on my computer/phone not listed here.	15.6%
18.	I read something NOT on a computer/phone.	11.8%
19.	I wrote something NOT on a computer/phone.	8.5%
20.	I watched TV.	12.8%
21.	I ate or drank something.	21.6%
22.	I spoke with someone in person.	13.5%
23.	I did a craft.	8.17%
24.	I stood up.	26.2%
25.	I did something physically active.	15.5%
26.	I went to the restroom.	14.1%
27.	I did something OFF a computer/phone not listed here.	17.6%

Table 3: Activities reported during the rest period by the (n=450) participants in the Activities cohort (in the order in which the activities were rated).

arm, TOST has 99.22% power to reject the null hypothesis of an absolute effect greater than 0.5 and 83.04% power for an absolute effect greater than 0.35.

This hypothesis was NOT confirmed.

- BoredomBeforeAndAfter vs. BoredomAfterOnly: Cohens D=-0.411
- Is BoredomBeforeAndAfter < BoredomAfterOnly with Cohens $D > -0.5: \; T_2 98 = 0.987, \; p = 0.163$
- Is BoredomBeforeAndAfter > BoredomAfterOnly with Cohens $D < 0.5: \; T_2 98 = -10.1, \; p = 5.37e-19$
- Presenting boredom questions before start of task leads to DECREASED responses after block0. because we cannot exclude H0:|D|>=0.5, we will use only the BoredomAfterOnly cohort in subsequent analyses.

1.2) We hypothesize that final state boredom will explain variance in subject-level POTD slope. This is a one-sided hypothesis.

This hypothesis was confirmed ($\chi^2(2, N = 16) = 8.769138, \; p = 0.012468$).

1.3) We hypothesize that the change in boredom will explain variance in subject-level POTD slope. This is a one-sided hypothesis.

This hypothesis was confirmed ($\chi^2(2, N = 16) = 18.640841, \; p = 0.00009$).

1.4) We hypothesize that trait boredom will explain variance in subject-level POTD slope. This is a one-sided hypothesis.

This hypothesis was NOT confirmed ($\chi^2(2, N = 16) = 2.374599, \; p = 0.305044$).

2.1) In the validation of short interval state MDES repeat administration, we hypothesize that the effect of including an initial administration will have an absolute effect size (cohen's d) less than 0.5. We will test this with two, one-sided t-tests (TOST). With an alpha of 0.01 and sample size of 150 participants per arm, TOST has 99.22% power to reject the null hypothesis of an absolute effect greater than 0.5 and 83.04% power for an absolute effect greater than 0.35.

This hypothesis was confirmed.

- MwBeforeAndAfter vs. MwAfterOnly: Cohens D=0.0739
- Is MwBeforeAndAfter < MwAfterOnly with Cohens $D > -0.5: \; T_2 98 = 7.52, \; p = 2.34e-12$
- Is MwBeforeAndAfter > MwAfterOnly with Cohens $D < 0.5: \; T_2 98 = -5.58, \; p = 5.39e-08$
- Presenting MW questions before start of task DOES NOT change responses after block0. Because we can exclude H0:|D|>0.5, we will use both MW cohorts in subsequent analyses.

2.2) We hypothesize that the final emotion dimension score will explain variance in subject-level POTD slope. This is a one-sided hypothesis.

This hypothesis was confirmed ($\chi^2(2, N = 16) = 44.013078, \; p = 2.771287e-10$).

2.3) We hypothesize that the change in emotion dimension score will explain variance in subject-level POTD slope. This is a one-sided hypothesis.

This hypothesis was confirmed ($\chi^2(2, N = 16) = 7.297105, \; p = 0.026029$).

2.4) We hypothesize that trait mind wandering will explain variance in subject-level POTD slope. This is a one-sided hypothesis.

This hypothesis was NOT confirmed ($\chi^2(2, N = 16) = 1.203324, \; p = 0.5479$).

3.1) We hypothesize that final mood ratings will be lower on average than the initial mood ratings in our real-world task. This is a one-sided hypothesis.

This hypothesis was NOT confirmed.

- Mean pre-break mood: 65.7%, post-break mood: 66.6%, change in mood: 0.909% (0.13%/min)
- happinessBeforeActivities < happinessAfterActivities (PAIRED): T=-1.33, p=0.0918
- happinessBeforeActivities > happinessAfterActivities (PAIRED): T=-1.33, p=0.908
- Free time break DOES NOT change mood ratings in block 0.

3.2) We hypothesize that the decrease in mood ratings will be smaller than that observed in the boredom task. This is a one-sided hypothesis.

This hypothesis was confirmed.

- activities < boredom: T=6.28, p=1
- activities > boredom: T=6.28, p=3.23e-10
- Free time break happiness change is GREATER than boredom happiness change in block 0.