

A highly replicable decline in mood during rest and simple tasks

Received: 17 May 2021

Accepted: 4 January 2023

Published online: 27 February 2023

 Check for updates

David C. Jangraw  ^{1,2}✉, Hanna Keren³, Haorui Sun², Rachel L. Bedder^{4,5}, Robb B. Rutledge  ^{4,5,6}, Francisco Pereira¹, Adam G. Thomas  ¹, Daniel S. Pine¹, Charles Zheng  ¹, Dylan M. Nielson  ^{1,9} & Argyris Stringaris  ^{7,8,9}

Does our mood change as time passes? This question is central to behavioural and affective science, yet it remains largely unexamined. To investigate, we intermixed subjective momentary mood ratings into repetitive psychology paradigms. Here we demonstrate that task and rest periods lowered participants' mood, an effect we call 'Mood Drift Over Time'. This finding was replicated in 19 cohorts totalling 28,482 adult and adolescent participants. The drift was relatively large (-13.8% after 7.3 min of rest, Cohen's $d = 0.574$) and was consistent across cohorts. Behaviour was also impacted: participants were less likely to gamble in a task that followed a rest period. Importantly, the drift slope was inversely related to reward sensitivity. We show that accounting for time using a linear term significantly improves the fit of a computational model of mood. Our work provides conceptual and methodological reasons for researchers to account for time's effects when studying mood and behaviour.

An important but implicit notion among behavioural and affective scientists is that each participant has a baseline mood or affective state that will remain constant during an experiment or vary only with emotionally salient events¹. Mood is modelled as a discounted sum of rewards and punishments^{2,3}, but many models hold that the timescale over which these events unfold is irrelevant and the passage of time itself has no effect on mood.

This assumption of a constant affective background has profound methodological implications for psychological experiments. First, consider a 'resting state' functional brain scan in which a participant is asked to stare at a fixation cross. On the basis of the constant affective background assumption, comparisons of resting-state neuro-imaging data between (for example) depressed and non-depressed participants are thought to reveal differences in their task-general traits, rather than their response to experimentally imposed rest periods. Second, consider an event-related design, such as a gambling or

face recognition task, during which participants experience stimuli (wins or losses) that elicit emotional reactions. When analysing these data, 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.

While convenient, this assumption of a constant affective background contradicts evidence from multiple fields that time impacts mood and behaviour. Affective chronometry research has demonstrated that affect changes systematically with time after an affective stimulus^{4–7}, and that individuals vary in the rates at which positive or negative affect decays after an event^{8,9}. Such individual differences may be linked to mental health. For instance, psychopathologists theorize that anhedonia, a symptom of both depression and schizophrenia, arises from a failure to sustain reward responses for a normative period of time¹⁰. And studies of attention deficit hyperactivity disorder suggest that hyperactivity's impulsive behaviour results from

¹National Institute of Mental Health, Bethesda, MD, USA. ²Department of Electrical and Biomedical Engineering, University of Vermont, Burlington, VT, USA. ³Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel. ⁴Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, UK. ⁵Wellcome Centre for Human Neuroimaging, University College London, London, UK. ⁶Departments of Psychology and Psychiatry, Yale University, New Haven, CT, USA. ⁷Department of Psychiatry, National and Kapodistrian University of Athens, Athens, Greece. ⁸Faculty of Brain Sciences, Division of Psychiatry, University College London, London, UK. ⁹These authors contributed equally: Dylan M. Nielson, Argyris Stringaris.

✉e-mail: djangraw@uvm.edu

delay aversion, the idea that a delay is itself unpleasant and impulsivity is simply a rational choice to avoid it^{11–13}.

Economists speak of the opportunity cost of time, suggesting that time spent performing one activity incurs the cost of other alternatives they might have chosen instead (such as paid work or leisure)^{14–16}. This idea is fundamental to the explore/exploit question that has recently preoccupied neuroscientists^{17–19}. Affect is central to this question: it is currently thought that negative affective states (such as boredom) building over time provide the subjective motivation to switch to a different activity^{20,21}.

When participants are engaged in a psychological task or rest period, they are committed to exploiting that task environment and are unable to explore other activities. This sense of constraint, or reduced agency, is considered central to feelings of boredom and its associated negative affect²². We might therefore conceive of a psychological task's behavioural constraint as a sort of negative affective stimulus that could gradually draw mood downward.

If this is true and the constant affective background assumption is violated, this could be problematic given evidence that spontaneous affective changes vary systematically between the individuals and groups being compared in affective science. For example, spontaneous negative thoughts are known to occur and vary substantially between humans, as highlighted by extensive work in mind wandering (MW)^{23–26}. Similarly, it is well known from occupational psychology that periods of low or relatively constant stimulation (as occurs in rest or repetitive experimental tasks) can induce varying levels of boredom^{27,28}. These insights raise the possibility that mood states will follow a similar pattern of inter-individual variability, creating potential confounds for resting-state and event-related experiments. However, the size, stability and clinical correlates of this variability remain unexplored.

In this Article, to answer these fundamental questions, we examine how the passage of time affects mood in a variety of experiments across studies, participants and settings. We find that participants' mood worsened considerably during rest periods and simple tasks, an effect we call 'Mood Drift Over Time' ('mood drift' for short). This downward mood drift was replicated in 19 large and varied cohorts, totalling 116 healthy and depressed adolescents recruited in person, 1,913 adults recruited online from across the United States, and 26,896 participants performing a gambling task in a mobile app. It was not observed when participants freely chose their own activities. We show that mood drift is related to, but not a trivial extension of, the existing constructs of boredom and thought content (including the task-unrelated thought often considered central to MW). We show that mood drift slopes are positively correlated with reward sensitivity and that this relationship is moderated by overall life happiness. These findings may have profound implications for experimental design and interpretation in affective science.

Results

Characterizing the effect

The results to follow characterize the average person's gradual decline in mood during rest and simple tasks, a phenomenon we call 'Mood Drift Over Time' ('mood drift' for short). This effect was initially observed in a task where participants were periodically asked to rate their mood (Fig. 1a). 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 min 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 characterize 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, in which they chose between an uncertain gamble or a certain outcome. This task is a standard one commonly used to examine mood^{3,29–31}. It was included

to observe the effects of rest on rational behaviour, to maintain links with previous studies of mood and reward^{2,3,32}, and to enable related analyses on a large cohort of participants ($n = 26,896$) playing a similar game on their smartphones³³ (Fig. 1b). A list of the cohorts we examined is in Extended Data Table 1). To quantify time's effect on mood, we created a linear mixed effects (LME) model with terms for initial mood and mood slope (that is, 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 (Methods). 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 cut-off, that is, 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 mood drift. The distribution of slopes was assumed to be Gaussian³⁴, but LME models are robust to violations of this assumption³⁵. All statistical tests used were two-sided unless otherwise specified.

As the smartphone game cohort was 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 (Methods). The model's time sensitivity parameter for each participant was used to quantify their mood drift.

Mood Drift Over Time is sizeable during rest. Our first objective was to estimate the size of the effect. In our initial cohort (called 15sRest-Between in Extended Data 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 (Fig. 2a, blue line). After 9.7 min of rest, the change in mood was considerable (mean \pm standard error (SE) $22.4 \pm 4.15\%$ of the mood scale). We replicated this in five other adult MTurk cohorts that received shorter opening rest periods (Fig. 2a, other lines).

Mood Drift Over Time is robust to methodological choices. To examine possible methodological confounds, we created slightly modified versions of the task to see whether the observed decline in mood ratings might be due to the following:

1. The aversive nature of rating one's mood: we did not find evidence that more frequent ratings changed mood drift (inter-rating interval \times time interaction -0.0103% mood, 95% confidence interval (CI) -0.0267% to 0.0061% , $t_{810} = -1.23$, $P = 0.219$, two-sided, Extended Data Fig. 1).
2. The method of rating mood and its susceptibility to fatigue: we did not find evidence that making every mood rating require an equally easy single keypress changed mood drift (-2.22 versus $-2.45\% \text{ mood min}^{-1}$, 95% CI -0.772 to 1.23 , $t_{70} = 0.427$, $P = 0.671$, two-sided).
3. The expected duration of the rest period: groups expecting different rest durations did not have different mood drift (-1.47 versus $-1.53\% \text{ mood min}^{-1}$, 95% CI -0.613 to 0.743 , $t_{104} = 0.185$, $P = 0.854$, two-sided).
4. Multitasking or task switching: participants moved their mood rating slider on 97.7% of trials.

The results of these control analyses suggested that mood drift cannot be explained by these methodological factors (Supplementary Note C).

Mood Drift Over Time 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

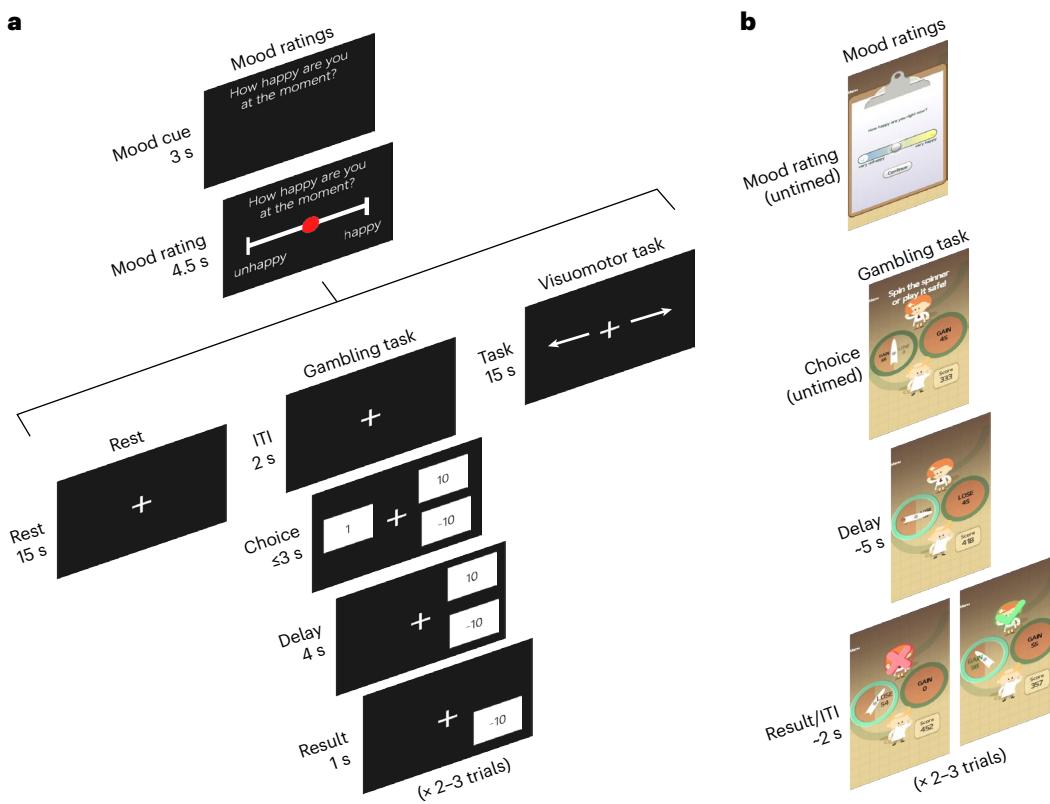


Fig. 1 | One cycle (mood rating + task). **a,b,** Cycle administered to online participants (**a**) and mobile app participants (**b**). After completing their first mood rating, participants completed one cycle of the rest, gambling or visuomotor task, then completed another mood rating, and so on. ITI denotes

inter-trial interval. 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 two or three (randomized) trials of the gambling task.

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 centreline, and each response would make the cross turn green if the response was accurate or red if it was too early or late (Methods). 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 we did not find evidence of a difference between the LME slope parameters of this group and those of the original cohort (-2.19 versus $-2.45\text{ \%mood min}^{-1}$, 95% CI -0.876 to 1.40 , $t_{68} = 0.437$, $P = 0.663$ for visuomotor task, -1.91 versus $-2.45\text{ \%mood min}^{-1}$, 95% CI -0.453 to 1.52 , $t_{104} = 1.07$, $P = 0.287$ for random gambling, both two-sided) (Fig. 2b).

Mood Drift Over Time is generalizable. We next investigated the generalizability of this result across age groups and recruitment methods. To do this, 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 (Methods). This group (Adolescent-01, $n = 116$) showed a pattern of declining mood similar to that observed in the MTurk cohort (Fig. 2c) (-1.69 versus $-1.93\text{ \%mood min}^{-1}$, 95% CI -0.122 to 0.599 , $t_{884} = 1.09$, $P = 0.275$, two-sided).

To more precisely characterize the effect, we fitted a large LME model to the complete cohort of online participants (both adults and

adolescents) completing rest or simple tasks in the first block (Extended Data Table 2). The mood drift parameter (rate of mood decline with time) for these 886 participants was $\text{mean} \pm \text{SE } -1.89 \pm 0.185\text{ \%mood min}^{-1}$, which was significantly less than 0 ($t_{864} = -10.3$, $P < 0.001$). After 7.3 min (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 to a Cohen's d of 0.574, with a 95% CI of 0.464 to 0.684 (ref. ³⁶).

Mood drift is diminished in a mobile app gambling game. We next tested whether mood drift 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 pre-registration (<https://osf.io/paqf6>). We applied the LME modelling procedure to this confirmatory cohort and again found a slope parameter that was significantly below zero at the group level ($\text{mean} \pm \text{SE } -0.881 \pm 0.0613\text{ \%mood min}^{-1}$, $t_{22,804} = -14.4$, $P < 0.001$). 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 penalized for failing to finish), mood tended to decrease with time spent on task.

We note, however, that mood drift was significantly smaller in this cohort (median -0.752 , inter-quartile range (IQR) $2.10\text{ \%mood min}^{-1}$) than in the combined cohort of online participants (median -1.53 , IQR $2.34\text{ \%mood min}^{-1}$, two-sided Wilcoxon rank-sum test, $W_{21,761} = -14.5$, $P < 0.001$; Extended Data Fig. 2). In total, 87.5% of online participants had negative slopes in the LME analysis, whereas only 70.2% of mobile app participants did. A histogram of the LME slope parameters for

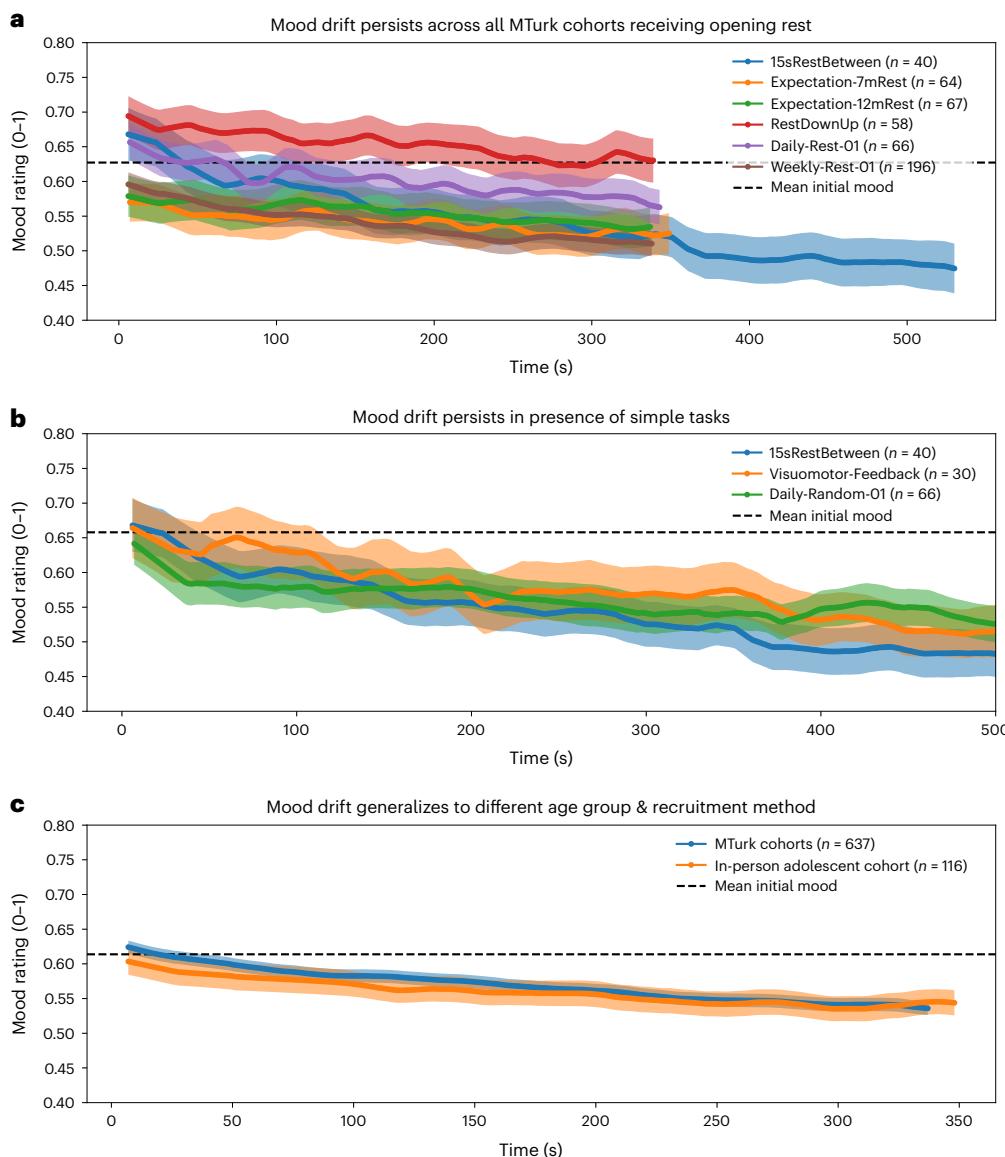


Fig. 2 | The timecourse of mood drift is consistently present across many cohorts and task modulations. These plots each show the mean timecourse of mood across participants in various online cohorts for the first block of the task. Each participant's mood between ratings was linearly interpolated before 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 (Extended Data Table 1). 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. Mood drift was discovered in one cohort (blue line) and replicated in five independent naïve cohorts.

b, Mood drift 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 RPEs and winnings (green). **c**, Mood drift was observed both in adults recruited on MTurk (combining across all MTurk participants that received opening rest or visuomotor task periods) (blue) and in adolescents recruited in person (orange).

online and mobile app participants is plotted in Fig. 3. This shows that, as one might expect, mood drift is sensitive to task context.

Next, to disentangle mood drift from the effects of reward and RPE in this dataset, we fitted the computational model described in Methods 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 (IQR 0.00197%, two-sided Wilcoxon signed-rank test, $W_{499} = 0$, $P < 0.001$). 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 Extended Data Figs. 3 and 4. The distribution of participants' time sensitivity parameters β_T (which can be interpreted as mood drift independent of reward effects) was centred significantly below zero (mean \pm SE $-0.128 \pm 0.00668\% \text{mood min}^{-1}$, two-sided Wilcoxon signed-rank test $W_{21,895} = 1.00 \times 10^8$, $P < 0.001$).

Mood Drift Over Time is absent in freely chosen activities. After the surprising finding that mood drift 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 pre-registered (<https://osf.io/gt7a8>) a task in which the

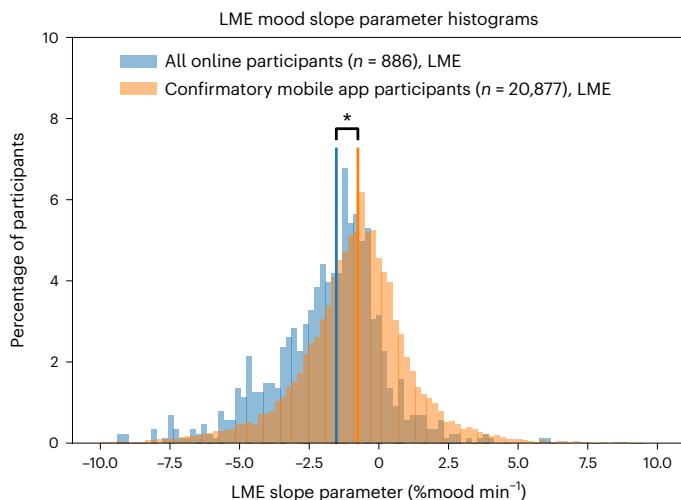


Fig. 3 | Individual subject LME slope parameters for online participants (blue) and mobile app participants (orange). The online participants had slopes below zero on average ($\text{mean} \pm \text{SE} = -1.89 \pm 0.185\% \text{mood min}^{-1}$, $t_{864} = -10.3$, $P < 0.001$), as did the mobile app participants ($\text{mean} \pm \text{SE} = -0.881 \pm 0.0613\% \text{mood min}^{-1}$, $t_{22,804} = -14.4$, $P < 0.001$). Mood drift was significantly less negative in the mobile app participants (median -0.752 , IQR $2.10\% \text{mood min}^{-1}$) than in the online participants (median -1.53 , IQR $2.34\% \text{mood min}^{-1}$, two-sided Wilcoxon rank-sum test, $W_{21,761} = -14.5$, $P < 0.001$). Vertical lines represent group medians. Stars indicate $P < 0.05$.

initial rest period was replaced with 7 min 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. The most frequent activities reported were thinking, reading the news and standing up (Supplementary Table 3).

This group was the first sample investigated in this study that did not exhibit mood drift. The mood ratings just after the free period were not statistically different from the mood ratings before the free period (66.6% versus 65.7%, 95% CI -2.15% to 97% , $t_{449} = -1.33$, one-tailed $P_{H0:\text{decrease}} = 0.0918$, $P_{H0:\text{increase}} = 0.908$). This change in mood was significantly greater than that of a cohort who received the standard rest period with interspersed mood ratings (cohort BoredomAfter-Only, $n = 150$) (0.909% versus -8.11% , 95% CI 5.95% to 12.1% , $t_{598} = 6.28$, $P < 0.001$, two-sided). This shows that, perhaps unsurprisingly, mood drift is not universal to all activities. However, the nominal increase in mood during this period ($0.130\% \text{mood min}^{-1}$) was much smaller than the decrease in mood observed during a typical rest period ($-1.89\% \text{mood min}^{-1}$). 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

Having characterized the effect at the group level, we next turned our attention to the individual. The motivation for this line of analysis is that, if an individual's mood slope is different from that of others in a way that remains stable over days or weeks, it may be linked to traits of clinical and theoretical interest. While the group average mood drift is negative during rest and simple tasks, there is considerable variation across participants (2.5th–97.5th percentile of subject-level mood drift for online participants -7.23 to $1.79\% \text{mood min}^{-1}$) (Fig. 3). Using an intraclass correlation coefficient (ICC) on cohorts that completed the task more than once, we found that these individual differences had moderate, statistically significant stability across blocks ($\text{ICC}(2, 1) = 0.465$, $P < 0.001$), days ($\text{ICC}(2, 1) = 0.343$, $P = 0.0031$) and weeks ($\text{ICC}(2, 1) = 0.411$, $P < 0.001$, one-sided since ICC values are

expected to be positive) (Extended Data Fig. 5, Supplementary Note D). We therefore investigated the relationship between this variability and other traits of clinical and theoretical interest.

Mood drift is associated with sensitivity to rewards. Mood is central to depression, which is thought to relate aetiologically to reward responsiveness^{37,38}. The idea that mood drift might be related to this responsiveness prompted us to investigate the relationship between participants' mood drift, reward sensitivity and life happiness in our computational model fits. The time sensitivity/mood drift parameter β_T was anticorrelated with the reward sensitivity parameter β_A ($r_s = -0.106$, $P < 0.001$, two-sided) (Fig. 4, left). This anticorrelation was weaker in participants with life happiness below the median (that is, those at greater risk of depression) than it was in those at/above it ($r_s = -0.0513$ versus -0.14 , $Z = 6.41$, $P < 0.001$, two-sided) (Fig. 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.

The direct relationship between depression risk and mood drift was significant, but its effect on model fit was very small. In our online participant LME model, higher depression risk score was significantly associated with less negative mood drift (depression risk \times time interaction, $\text{mean} \pm \text{SE} = 0.515 \pm 0.109\% \text{mood min}^{-1}$, $t_{869} = 4.75$, $P < 0.001$; Extended Data Fig. 6). While the model fit improved, the within-individual variance explained by the addition of this interaction term was very small ($f^2 = 0.00289$) (refs. ^{39,40}). Nevertheless, the interaction term's significance was replicated in two more independent cohorts (including the mobile app cohort, where time sensitivity and life happiness were weakly anticorrelated; Extended Data Fig. 7, bottom right) and was robust to methodological artefacts such as floor effects (Supplementary Notes E–G). Taken together, these results demonstrate relationships between mood drift and other important individual differences: depression risk, life happiness and reward sensitivity.

Impact on behaviour

Participants are less likely to gamble after rest periods. To investigate whether mood drift's effects extend to behaviours beyond subjective mood reports, we examined the impact of rest and mood drift on behaviour in the gambling tasks. Past research has shown that a participant's choice between a certain outcome and a more exciting but uncertain gamble is affected by mood as induced by unexpected gifts^{41,42}, music⁴³ and feedback³¹. We asked whether mood drift would influence this behaviour in a similar way.

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 (median 0.55 versus 0.66, IQR 0.28 versus 0.31, two-sided Wilcoxon rank-sum test, $W_{722} = 2.08$, $P = 0.0377$) (Fig. 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 four 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 (median 3, IQR 2 for both short and long rest, two-sided Wilcoxon rank-sum test, no rest versus short rest: $W_{469} = 4.85$, $P < 0.001$; no rest versus long rest: $W_{344} = 4.79$, $P < 0.001$; both $<0.05/3$ controlling for multiple comparisons; Fig. 5, bottom). However, we did not find evidence of a difference between the long and short rest groups (two-sided $W_{629} = 0.52$, $P = 0.603$). Trial-wise gambling behaviour differences between rest and no-rest groups are most pronounced in the first four trials, much like the differences observed in mood (Fig. 5, middle). However, no significant correlation was observed between an individual's mood drift parameter during the preceding rest block and the number of times they chose to gamble in the first four trials ($r_s = 0.0317$, $P = 0.427$, two-sided).

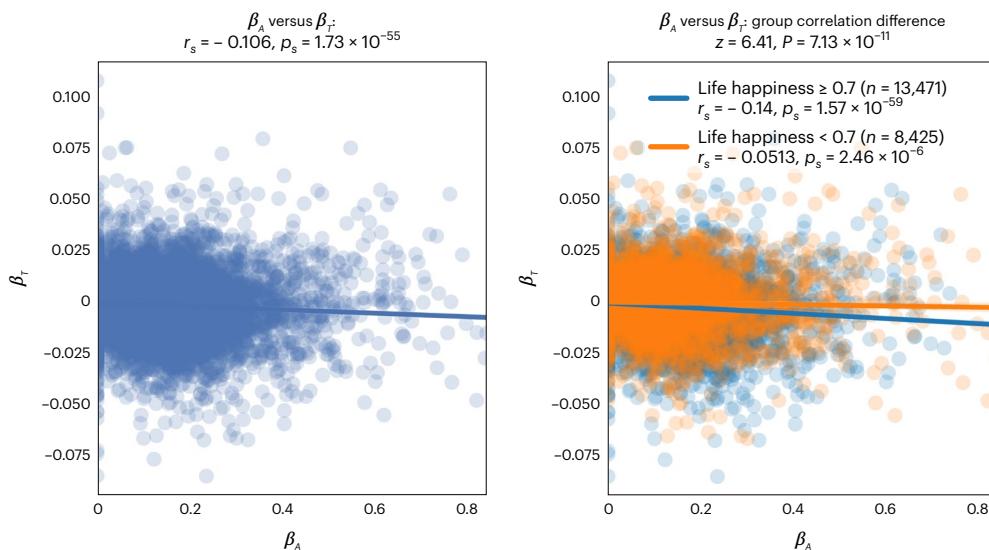


Fig. 4 | Individual differences in sensitivity to the passage of time relate to other individual differences in the mobile app cohort. The computational model's time sensitivity parameter β_T for each participant in the mobile app cohort is plotted against that participant's reward sensitivity parameter β_A (left). r_s and p_s denote Spearman correlation coefficient and corresponding P value. 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). Each dot is a participant ($n = 21,896$). Each line is a linear best fit, and patches show the 95% CI of this fit. The group difference in Spearman correlations was statistically confirmed using a z statistic. P values shown are two-sided with no correction for multiple comparisons.

Relationship to boredom and thought content

We next examined whether the existing construct of boredom or MW could trivially explain mood drift. In a pre-registered (<https://osf.io/gt7a8>) data collection and analysis, we examined the relationship between mood drift and these more established constructs at the state level, state change level and trait level (Supplementary Notes L and M). Participants were randomized to a boredom, MW or Activities cohort (described previously) at the time of participation.

Mood Drift Over Time is weakly related to state boredom. We assessed whether mood drift could be explained by boredom. Participants completed a rest block with interspersed mood ratings, plus a state boredom questionnaire (the Multidimensional State Boredom Scale's short form, MSBS-SF)⁴⁴ afterwards (cohort BoredomAfterOnly, $n = 150$), or before and afterwards (cohort BoredomBeforeAndAfter, $n = 150$), and a trait-boredom questionnaire (the short boredom proneness scale, SBPS)⁴⁵.

In our LME model of mood, we added a factor for final state boredom (that is, at the end of the rest block). We then compared this baseline model with one that further added the interaction between final boredom and time. The difference represents the ability of boredom to account for mood drift. While the model fit improved, the added within-individual variance explained by the addition of this new interaction term was very small ($ff^{22} = 0.00578$). The change in state boredom across the rest block produced similar results ($ff^{22} = 0.0111$).

Including time's interaction with trait boredom in the model did not explain significant additional variance in mood (likelihood ratio test: $\chi^2(1, N = 16) = 0.0253, P = 0.874$).

Mood Drift Over Time is weakly related to thought content. We also assessed whether mood drift could be explained by the content of ongoing thought, including the task-unrelated thought, stimulus-independent thought and spontaneity often considered in definitions of MW⁴⁶. We note that such content-based definitions of MW are controversial and do not capture the dynamics-based definition espoused by some researchers^{47,48}. New participants completed a rest block with interspersed mood ratings, plus a Multidimensional

Experience Sampling (MDES) questionnaire⁴⁹ afterwards (cohort MwAfterOnly, $n = 150$), or before and afterwards (cohort MwBeforeAndAfter, $n = 150$), and a trait-MW questionnaire (the mind-wandering questionnaire (MWQ)⁵⁰). MDES results produce 13 principal components that attempt to capture the content of ongoing thought. We investigated how well this complete collection of components explains within-individual mood variance.

In our LME model of mood, we added 13 factors for 'final' MDES components (that is, at the end of the rest block). We then compared this baseline model with one that further added the 13 interactions between these final MDES components and time. The difference represents the ability of MDES components to account for mood drift. While the model fit improved, the within-individual variance explained by the addition of these new interaction terms was small ($ff^{22} = 0.0227$). The change in MDES components across the rest block produced similar results ($ff^{22} = 0.0380$).

Including time's interaction with trait MW in the model did not explain significant additional variance in mood ($\chi^2(1, N = 16) = 0.305, P = 0.581$).

Discussion

In this study, we describe the discovery of a highly replicable and relatively large effect that we call Mood Drift Over Time: the average participant's mood gradually declined with time as they completed simple tasks or rest periods. Mood's sensitivity to the passage of time is a long-intuited phenomenon that is widely acknowledged in literature^{51–53} and philosophy^{54–56}. Our results provide robust empirical evidence for this phenomenon and reveal its temporal structure, its variability across individuals and its level of stability. These results call into question the long-held constant affective background assumption in behavioural and affective science.

The mechanism that enables mood 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 to guide behaviour²¹. Lower

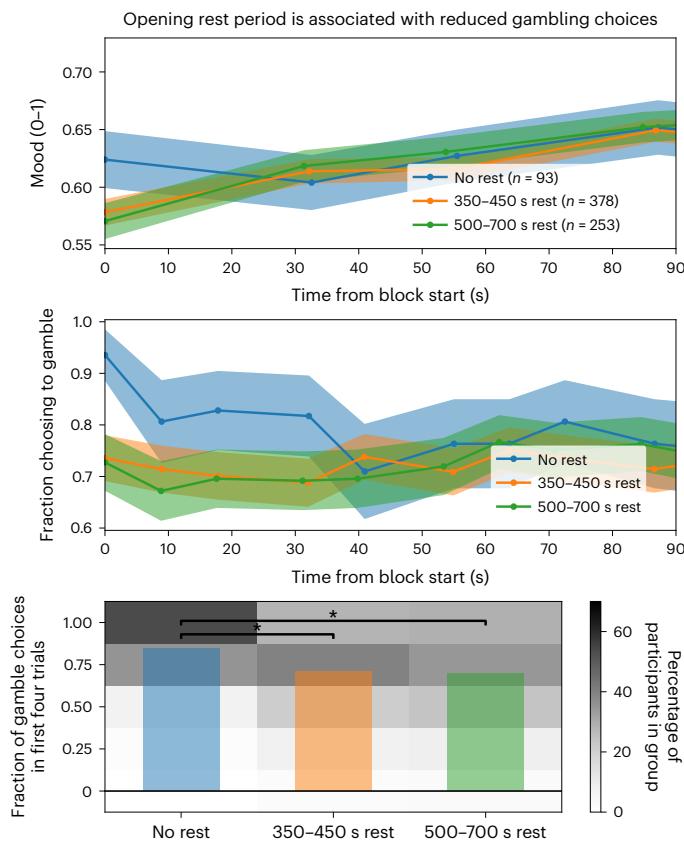


Fig. 5 | Rest periods decreased the likelihood of choosing to gamble in the first four trials after rest ended. Top: mean \pm standard error mood ratings across participants in their first block of (positive closed-loop) gambling preceded by different rest period durations. Middle: fraction of participants in each group that chose to gamble on each trial of this first gambling block (mean \pm 95% CIs derived from a binomial distribution). Bottom: bars show mean across participants of the fraction of the first four trials of this first gambling block that participants chose to gamble. Histogram shows the distribution of choices (that is, to gamble on zero, one, two, three or four trials) within each group. Stars indicate that a pair of groups was significantly different (two-sided Wilcoxon rank-sum test, no rest versus short rest: $W_{469} = 4.85, P < 0.001$; no rest versus long rest: $W_{344} = 4.79, P < 0.001$; both $<0.05/3$ controlling for multiple comparisons). Sample sizes are: no rest group: $n = 93$ participants, 350–450 s rest group: $n = 378$ participants, 500–700 s rest group: $n = 253$ participants).

mood could function as an estimate of that opportunity cost, making mood drift an adaptive signal that informs decisions to exploit (stay on task) or explore (switch task)²⁰.

Supporting this reward/cost-based interpretation of our findings is our observation that depressed participants showed less negative mood drift. This would at first seem paradoxical since phenomena such as boredom have traditionally been linked to melancholia and depression (for example, by Schopenhauer⁵⁷ and Kierkegaard⁵⁸). Yet it has been argued cogently⁵⁹ that such a view conflates negative affect as a trait (for example, proneness to boredom) with negative affect as a state (a momentary experience). As valuation of reward is thought to be reduced in depression^{37,38}, it is possible that misalignment with one's goals and violation of reward expectations—and resultant downward mood drift—will be less pronounced in depression. This interpretation is supported by our finding that mood drift is less pronounced in those with lower reward sensitivity, and that the relationship between reward sensitivity and mood drift was moderated by depression risk (Fig. 4). It is tempting to speculate that reduced mood drift could contribute to reduced motivation for action or environmental change in those with depression.

We found that mood declined during rest and tasks (including a mobile app more engaging than most experiments) but not freely chosen activities. This suggests that researchers are subjecting their participants to an unnatural stressor in their experiments without accounting for it in their analyses or interpretations. Changes in mood on the scale of tens of minutes prevent these longer blocks of time from being truly interchangeable. This means that variations in experimental procedures that might seem inconsequential could still introduce confounds.

For example, let us consider a large collaborative study that is based on multisite imaging data collection, such as ENIGMA⁶⁰. In this dataset, centres vary in the duration of the resting-state functional magnetic resonance imaging (fMRI) scan and whether it takes place at the start or end of the scan session⁶¹. This could lead to high variability between sites simply because patients at sites with longer or later 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. (We do not imply that mood drift lowers reliability in resting-state fMRI^{62–64}; we simply point out its role as a potential confound when drawing inferences about mood and brain states during/after rest.)

In this paper, we introduce the new term Mood Drift Over Time for the following reasons. First, the phenomenon is highly replicable; second, it is of considerable effect size; third, it is relevant to both everyday situations and to scientific experiments; fourth, mood drift does not seem to be captured by existing terms such as boredom or MW. We employ the term mood drift in the spirit of describing a mental phenomenon^{65–67}, as a first step before explaining or categorizing it. It is possible that mechanisms for mood drift are reward sensitivity and opportunity cost, yet the subjective experience and its influence on the outcome of experimental studies seem to require the separate term that we have introduced.

The distinction between mood drift and boredom requires special consideration due to their apparent similarities. State boredom assessed using the MSBS-SF⁴⁴ accounted for modest variance beyond other factors. Of course, the MSBS is only one (relatively well established) way of measuring boredom; moreover, there is debate about the very conceptualization of boredom and its heterogeneity^{22,59,68}. Therefore, we cannot conclude purely from these results that boredom is not driving mood drift. Future work might instead ask participants to directly report their boredom⁶⁹, enabling more frequent assessment of boredom as an emotion⁷⁰.

Importantly, we show that accounting for time using a linear term significantly improves the fit of a computational model of mood. A linear term may be unrealistic as we expect that, on a bounded mood scale, the effect will eventually saturate. However, we propose that, until alternative models have been established, the linear term may be a good-enough way to account for the substantial effects of mood drift on the timescale of most experiments.

Our study has several strengths, including adherence to good data analysis practices such as pre-registration and replication, the addition of a longitudinal design to test reliability and the use of rigorous computational modelling (including train–test splits and regularization). Our study demonstrated the effect in adolescents as well as adults and showed how the effect differs in people with varying reward sensitivity and depression risk. We used control experiments to eliminate potential confounds and test alternative explanations (Supplementary Notes C–G). Yet our study should also be seen in light of some shortcomings.

First, this study uses self-reported momentary mood ratings as in previous studies with similar methodology^{2,3}. Such ratings can be criticized as being subjective and difficult to interpret. However, mood is a well-established construct of central importance to affective science. Its definition as a long-duration affective state that is not

immediately responsive to stimuli^{71,72} makes it central to the study of mood disorders defined by long-term affect⁷³. Mood is distinct from emotion, in part, by being less temporally responsive^{74–76}. Mood's links to long-term context makes it the more useful construct to describe gradual changes in affect.

Despite its subjectivity, self-report remains the gold standard for the measurement of mood and emotion^{76–78}. It is widely used in clinical⁷⁹, epidemiological⁸⁰ and psychological research (including ecological momentary assessment⁸¹). Other physiological 'markers' of affect are typically benchmarked against these self-reports. And evidence suggests that these candidates lack the reliability of self-reports: different emotions cannot be distinguished by their autonomic nervous system signatures⁸², facial expressions^{83,84} or neural activity⁸⁵. In our experiments, initial mood ratings showed strong association with trait mood ratings, underscoring their psychometric validity (Extended Data Fig. 8). Our study cannot conclusively determine mood drift's behavioural consequences. On average, rest induces downward mood drift (Fig. 2) and decreases gambling behaviour (Fig. 5). However, a significant correlation between individuals' mood drift and gambling behaviour was not observed. Our results are not able to discern whether the change in behaviour is directly linked to mood drift or to some other consequence of rest.

Our study's limited set of tasks, all of which induced mood drift, makes it difficult to discern the phenomenon's key contributing factors. We chose to focus on a category that is extremely common in neuroscience: long, neutral, low-stimulation tasks. Most researchers would see these qualities as unobjectionable or even desirable. We hope that the results of this study will lead researchers to re-examine this idea in their own research.

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 computerized tasks ('jobs') from home⁸⁶. 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. As 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 institutional review board review by the National Institutes of Health (NIH) Office of Human Subjects Research Protections. The consent process and task/survey specifics were approved by the Office of Human Subjects Research Protections. 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 six-to-ten-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 min period starting when they accepted the job on Amazon Mechanical Turk.

The consent form included a description of the tasks they were about to perform, but cohorts were blinded to the specific cohort to which they had been assigned. Most cohorts were collected in series, but some were randomized to a cohort at the time of participation (we have specified these in Methods or Results). In the initial cohorts, no statistical methods were used to pre-determine sample sizes, but

our cohort sample sizes are similar to those reported in ref.², and our combined cohorts are much larger.

A total of 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. Forty-four 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 years).

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 randomized 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 3 days, but the task and survey had to be done together in the same 90 min 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 6 days, but the task and survey had to be done together in the same 90 min period. This cohort was then invited to complete the same task and survey a third time, 2 weeks after their first run. A total of 196 individuals completed the task and survey the first week. Among these, 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. In total, 149 (76.0%) individuals completed the task and survey in all 3 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 enroled in the NIMH depression characterization study who were offered to complete tasks for this study. Among these, 129 agreed, a participation rate of 56.1%. Ten adolescents who had not completed all three questionnaires were excluded from the results, as were three 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 years). Fifty-six participants (48.2%) had been diagnosed with major depressive disorder (MDD) by a clinician at the NIH, and 4 were determined to have subclinical MDD (3.4%). Participants had a mean depression score of MFQ of 6.5 (± 5.5 standard deviation (s.d.)) and a mean anxiety score of SCARED of 2.2 (± 3.0 s.d.).

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. Eighty-two (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, MW and Activities participants. In response to reviewer comments, a pre-registered 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 MW on mood changes, so they completed surveys about these traits in addition to the demographics, CES-D and SHAPS questions. Participants were randomized to one of these five 'follow-up cohorts', summarized in Extended Data Table 1:

- BoredomBeforeAndAfter ($n = 150$), who received a boredom state questionnaire both before and after a 7 min rest period with 15 s of rest between mood ratings.
- BoredomAfterOnly ($n = 150$), who received a boredom state questionnaire only after a 7 min rest period with 15 s of rest between mood ratings.
- MwBeforeAndAfter ($n = 150$), who received an MDES questionnaire both before and after a 7 min rest period with 15 s of rest between mood ratings.
- MwAfterOnly ($n = 150$), who received an MDES questionnaire only after a 7 min rest period with 15 s of rest between mood ratings.
- Activities ($n = 450$), who received instructions to leave the task for 7 min 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 pre-registered tasks and analyses can be found at <https://osf.io/gt7a8>, registered on 18 November 2021. A total of 1,143 participants completed these tasks online. Ninety-three 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 1,050 remaining participants, 463 were female (44.1%). Participants had a mean age of 39.3 years (range 20–80 years).

The above sample sizes were selected using power calculations described in detail in the pre-registration. For the scale validation experiments, a sample size of 150 in each group with an α of 0.01 gives 99.02 power to detect a medium effect ($d = 0.5$) and 83.04% power to detect an intermediate effect ($d = 0.3$) assuming the effect truly is null at a population level. Power for linear multiple regression tests were calculated in G*Power⁸⁷. In the boredom and MW cohorts, samples of 150 participants were selected to provide 80% power to detect a 7.99% increase in variance explained with the inclusion of a single parameter ($\alpha = 0.01$, 20 total predictors) and a 95% power to detect a 12.18% change in variance explained. In analyses using a pair of cohorts, 300 participants gives 80% power to detect a 3.93% increase in variance explained and a 95% power to detect a 6.01% increase in variance explained. An Activities cohort of 450 participants was chosen to provide 80% power to detect a difference between the Activities and MTurk cohorts of Cohen's d of 0.2, and it also provides 80% power to detect a decrease in mood in the Activities cohort of Cohen's d of 0.15.

Mobile app participants. Gambling behaviour and mood rating data were collected from a mobile app called 'The Great Brain Experiment', described in ref.³. 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 United States and the 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). In total, 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 optimized using only the exploratory cohort, then tested on the confirmatory cohort. These confirmatory analyses were pre-registered on the Open Science Framework (<https://osf.io/paqf6>, registered on 29 January 2021).

In the LME 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 $Q1 - 1.5 \times IQR$ or greater than $Q3 + 1.5 \times IQR$ (where $Q1$ is the 25th percentile, $Q3$ is the 75th percentile and $IQR = Q3 - Q1$) were excluded from this LME analysis. In total, 4.65% of participants were excluded on the basis of these criteria, leaving $n = 20,877$ mobile app participants.

Task and survey

The online tasks were created using PsychoPy3 (v2020.1.2) and 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 Extended Data Table 1.

Mood ratings. The task given to online participants is outlined in Fig. 1a. 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 s. Then a slider appeared below the question, with a scale whose ends were labelled '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 s, 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 s. The slider then appeared, and the participant had 6.5 s 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 s 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 MW). 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 mood drift and boredom. Each received a rest period with mood ratings 20 s apart, followed by the MSBS-SF, an eight-item scale of

state boredom⁴⁴. Participants rated statements like ‘I feel bored’ on a seven-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 eight 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 mood drift and the emotional valence of ongoing thought (including MW). Each participant in the two MW cohorts received a rest period with mood ratings 20 s apart, followed by a 13-item MDES as described by Turnbull et al.⁴⁹. 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 MwAfterOnly, $n = 150$) completed the MDES only after the rest period.

As described by Ho et al.⁸⁸, we used principal components analysis (PCA) to quantify the affective valence of thought at each administration of MDES. We first compiled the MDES responses of all participants in the MwAfterOnly group into a matrix with 13 (the number of items in each administration) columns and 450 (the number of administrations) rows. We then used scikit-learn’s PCA function to find 13 orthogonal dimensions explaining the MDES variance. The use of PCA orthogonalizes the MDES responses, which is desirable for their use as explanatory variables in an LME³⁵.

For a pre-registered analysis, we focused on the emotional content of ongoing thought (this approach was later abandoned in favour of examining the collective predictive power of all 13 MDES components, Supplementary Notes L and M). By examining the component matrix, we identified the component that loaded most strongly onto the ‘emotion’ item of the MDES (in which they reported their thoughts as being negative or positive). The ‘emotion dimension’ of each MDES (in both MW cohorts) was then quantified as the amplitude of this component, calculated by applying this pre-learned PCA transformation to the data and extracting the corresponding column. 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 min 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 five-point Likert scale with options labelled ‘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 randomized 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: 1× screen height, period: 4 s). Participants were asked to press the spacebar at the exact moment when the cross was in the centre 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 (that is, less than 0.262 s before or after the actual centre crossing).

Gambling blocks. In each trial of the gambling task, participants saw a central fixation cross for 2 s. 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 s 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 s, then the actual outcome appeared for 1 s. If they chose not to gamble, the certain outcome appeared for 5 s. The locations (top/bottom) of the higher and lower gambling options were randomized.

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 RPE. If they chose to gamble, participants would always receive the base value + RPE option. To encourage gambling, the ‘certain’ value was set to $(\text{win} + 2 \times \text{loss})/3$, 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. RPE magnitudes of less than 0.03 were increased to 0.03. If three 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 on the basis of 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 three incongruent trials in a row, the next trial was forced to be congruent. The RPE was calculated as in a Proportional-Integral 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:

$$\text{RPE}(t) =$$

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

where $t = 1, 2, \dots, n$ 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 two or three trials (on average, one 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 cut-offs were determined on the basis of simulations such that any value 1–6 was 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 Status⁸⁹. 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 socio-economic status⁹⁰.

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 symptoms⁹¹. They also completed the Snaith–Hamilton Pleasure Scale (SHAPS), a 14-item scale of hedonic capacity⁹².

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 Short Child Self-Report Mood and Feelings Questionnaire (MFQ), a 13-item scale of how the participant has been feeling and acting recently^{79,93}. They also included the Screen for Child Anxiety Related Emotional Disorders (SCARED), a 41-item scale of childhood anxiety⁹⁴. These questionnaires were completed before the subject began completing the online tasks described above.

Participants recruited for follow-up investigations of boredom, MW, and free time activities also completed the SBPS, an eight-item scale of an individual's proneness to boredom in everyday life⁴⁵. They also completed the five-item MWQ, which quantifies a person's proneness to MW in everyday life⁵⁰. The SBPS and MWQ were used to quantify trait-level boredom and MW, respectively.

Mobile app. The task given to mobile app participants is outlined in Fig. 1b. 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 s 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 s inter-trial interval before the game advanced to the next trial. After every two to three 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 ref. ³. In the second version, gain trials had three certain amounts (35, 45 and 55) and 15 gamble amounts (59, 66, 72, 79, 85, 92, 98, 105, 111, 118, 124, 131, 137, 144 and 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 have three prospective gains (40, 44 and 75) and ten prospective losses (-10, -19, -28, -37, -46, -54, -63, -72, -81

and -90). Both versions are described further in ref. ³³. The median participant played the game for approximately 5 min.

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.

LME model

Analyses and statistics were performed using custom scripts written in Python 3. Participants' momentary subjective mood ratings were fitted with a LME model with rating time as a covariate using the Pymer4 software package (<http://eshinjolly.com/pymer4/>)⁹⁵. 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 modelled by a slope and intercept parameter such that:

$$M(t) = M_0 + \beta_T \times T(t) \quad (1)$$

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 modelling algorithm also produced a group-level slope and intercept term as well as CIs 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:

$$\text{Mood} \sim 1 + \text{Time} \times (\text{isMale} + \text{meanIRIOver20} + \text{totalWinnings} + \text{meanRPE} + \text{fracRiskScore} + \text{isAge0to16} + \text{isAge16to18} + \text{isAge40to100}) + (\text{Time} | \text{Subject}) \quad (2)$$

`isMale` is 1 if the participant reported their gender as '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 RPE 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 cut-off: that is, their MFQ score divided by 12 or their CES-D score divided by 16.

While the bounded mood scale prevents the error term of our mood models from being truly Gaussian, LMEs are typically robust to such non-Gaussian distributions³⁵.

For reliability analyses, the first block of each run was modelled 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 ($r_s = 0.548$, $P < 0.001$, two-sided; Extended Data Fig. 8, left).

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

LME model comparisons

To compare the ability of additional terms like depression risk and state boredom to explain variance in our model of mood, we employed an

analysis of variance (ANOVA) that compared two models: a reduced model with the factor but without its interaction with time, and an expanded model with both the factor and its interaction with time. All factors in equation (2) were included in both models (except in the case of depression risk, where the reduced model contained `fracRiskScore` but not its interaction with Time). We then used R's ANOVA function to compare the expanded and reduced model. The degrees of freedom were quantified as the difference in the number of parameters in the two models.

To examine the impact of including a factor(s) on mood variance explained, we used the within-individual and between-individual variance explained (R_1^2 and R_2^2) as defined in refs. ^{96,97}. This calculation required a null model including only an intercept and random effects, which we defined as:

$$\text{Mood} \sim 1 + (1 + \text{Time}|\text{Subject}) \quad (3)$$

The within-individual variance R_1^2 of each model was defined as:

$$R_1^2 = 1 - \frac{\sigma_e^2 + \sigma_a^2}{\sigma_{e0}^2 + \sigma_{a0}^2} \quad (4)$$

where σ_e^2 is the variance of the residuals of the model, σ_a^2 is the variance of the random effects, σ_{e0}^2 is the variance of the residuals of the null model, and σ_{a0}^2 is the variance of the random effects in the null model. The variance of the random effects in a model was calculated using R's `MuMin` library⁹⁸, taking into account the correlation between model factors.

The between-individual variance R_2^2 of each model was defined as:

$$R_2^2 = 1 - \frac{\sigma_e^2 + \sigma_a^2/k}{\sigma_{e0}^2 + \sigma_{a0}^2/k} \quad (5)$$

where k was defined as the harmonic mean of the number of mood ratings being modelled for each participant.

As the depression risk, boredom and MW factors were constant for each subject, we focus primarily on the between-individual variance explained R_2^2 .

To compare the variance explained by the expanded and reduced models as a measure of effect size, we used Cohen's f^2 ²² statistic^{39,40}, defined as:

$$f^2 = \frac{R_{AB}^2 - R_A^2}{1 - R_{AB}^2} \quad (6)$$

where R_{AB}^2 is the variance explained by the expanded model and R_A^2 is the variance explained by the reduced model. Separate f^2 ²² values can be calculated using the within-individual or between-individual variances. Using Cohen's guidelines³⁹, $f^2 \geq 0.02$ is considered a small effect, $f^2 \geq 0.15$ is considered a medium effect and $f^2 \geq 0.35$ is considered a large effect.

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 ref. ² 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 ref. ² was modified to include a coefficient β_T that linearly relates time and mood. Our modified model is defined as follows:

$$\hat{M}(t) = M_0 + \beta_A \sum_{u=1}^t \lambda^{t-u} A(u) + \beta_E \sum_{u=1}^t \lambda^{t-u} E(u) + \beta_T T(t) \quad (7)$$

In the above equation, $t = 1, 2, \dots, n$ 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 β s 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:

$$E(t) = \frac{1}{t-1} \sum_{u=1}^{t-1} A(u) \quad (8)$$

If we remove the influence of time (that is, set our $\beta_T = 0$), the full mood model in ref. ² is equivalent to this one as long as its RPE coefficient is less than its expectation coefficient (that is, $\beta_R^{\text{Keren}} < \beta_E^{\text{Keren}}$) and $\beta_E^{\text{Keren}} > 0$, where β_R^{Keren} and β_E^{Keren} denote the values β_R and β_E defined in ref. ²). The values in our model can be derived from the values in theirs by setting $\beta_A = \beta_R^{\text{Keren}}$ and $\beta_E = \beta_E^{\text{Keren}} - \beta_R^{\text{Keren}}$.

We used the PyTorch package⁹⁹ on a GPU to fit 500 models simultaneously for each subject. β_T was initialized to random values with distribution $\mathcal{N}(0, 1)$. β_E and β_A were initialized to random values with distribution Lognormal(0, 1) and capped to the interval [0, 10] on every iteration. M_0 and λ were initialized 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:

$$y = \frac{1}{1 + e^{-x}} \quad (9)$$

At the end of 100,000 iterations, the model with the lowest sum of squared errors (SSE) (that is, $\sum_{t=1}^N (\hat{M}(t) - M(t))^2$) 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 ADAM¹⁰⁰ 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 minimized was:

$$L = \sum_{t=1}^n (\hat{M}(t) - M(t))^2 + \lambda_{EA} \times (\beta_A^2 + \beta_E^2) + \lambda_T \times \beta_T^2 \quad (10)$$

The regularization hyperparameters λ_{EA} and λ_T were determined from a tuning step, in which the model was trained on the first ten 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 minimized this smoothed testing loss.

As in the LME, the bounded mood scale prevents the error term of our mood models from being truly Gaussian. Our computational model attempted to mitigate the effect of non-Gaussianity by capping mood predictions to the allowable range, initializing parameters to non-normal distributions, and restricting parameters to feasible ranges on every iteration.

As in the online cohort's LME model, the initial mood parameter M_0 showed psychometric validity. It was significantly correlated with life happiness ($r_s = 0.362, P < 0.001$; Extended Data Fig. 8, 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:

$$\hat{M}(t) = M_0 + \beta_A \sum_{u=1}^t \lambda^{t-u} A(u) + \beta_E \sum_{u=1}^t \lambda^{t-u} E(u) \quad (11)$$

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

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

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/km69z/>. Mobile app participants' data can be found on Dryad at <https://doi.org/10.5061/dryad.prr4xgxkk> (ref. ¹⁰¹).

Code availability

The code for the task and survey is available on GitLab at <https://gitlab.pavlovia.org/mooddrift>. 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/MoodDrift>.

References

- Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J. & Nichols, T. E. *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (Elsevier Science, 2011).
- Keren, H. et al. The temporal representation of experience in subjective mood. *eLife* **10**, 1–24 (2021).
- Rutledge, R. B., Skandali, N., Dayan, P. & Dolan, R. J. A computational and neural model of momentary subjective well-being. *Proc. Natl Acad. Sci. USA* **111**, 12252–12257 (2014).
- Frijda, N., Mesquita, B., Sonnemans, J. & Goozen, S. The duration of affective phenomena or emotions, sentiments and passions. *Int. Rev. Stud. Emotion* **1**, 187–225 (1991).
- Scherer, K. R. & Wallbott, H. G. Evidence for universality and cultural variation of differential emotion response patterning. *J. Pers. Soc. Psychol.* **66**, 310–328 (1994).
- Davidson, R. J. Affective style and affective disorders: perspectives from affective neuroscience. *Cogn. Emot.* **12**, 307–330 (1998).
- Davidson, R. J. Comment: affective chronometry has come of age. *Emot. Rev.* **7**, 368–370 (2015).
- Gilboa, E. & Revelle, W. *Personality and the Structure of Affective Responses* (Psychology Press, 1994).
- Hemenover, S. H. Individual differences in rate of affect change: studies in affective chronometry. *J. Pers. Soc. Psychol.* **85**, 121 (2003).
- Kring, A. M. & Barch, D. M. The motivation and pleasure dimension of negative symptoms: neural substrates and behavioral outputs. *Eur. Neuropsychopharmacol.* **24**, 725–736 (2014).
- Sonuga Barke, E. J. S., Taylor, E., Sembi, S. & Smith, J. Hyperactivity and delay aversion-I. The effect of delay on choice. *J. Child Psychol. Psychiatry* **33**, 387–398 (1992).
- Solanto, M. V. et al. The ecological validity of delay aversion and response inhibition as measures of impulsivity in AD/HD: a supplement to the NIMH multimodal treatment study of AD/HD. *J. Abnorm. Child Psychol.* **29**, 215–228 (2001).
- Sonuga Barke, E. J. S., Cortese, S., Fairchild, G. & Stringaris, A. Annual research review: transdiagnostic neuroscience of child and adolescent mental disorders-differentiating decision making in attention deficit/hyperactivity disorder, conduct disorder, depression, and anxiety. *J. Child Psychol. Psychiatry* **57**, 321–349 (2016).
- McRae, T. W. Opportunity and incremental cost: an attempt to define in systems terms. *Account. Rev.* **45**, 315–321 (1970).
- Hoskin, R. E. Opportunity cost and behavior. *J. Account. Res.* **21**, 78–95 (1983).
- Palmer, S. & Raftery, J. Opportunity cost. *BMJ* **318**, 1551–1552 (1999).
- Cohen, J. D., McClure, S. M. & Yu, A. J. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. B* **362**, 933–942 (2007).
- Constantino, S. M. & Daw, N. D. Learning the opportunity cost of time in a patch-foraging task. *Cogn. Affect. Behav. Neurosci.* **15**, 837–853 (2015).
- Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L. & Platt, M. L. A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology* **42**, 1931–1939 (2017).
- Geana, A., Wilson, R., Daw, N. D. & Cohen, J. D. Boredom, Information-Seeking and Exploration. *Proc. 38th Annual Conference of the Cognitive Science Society* (2016).
- Agrawal, M., Mattar, M. G., Cohen, J. D. & Daw, N. D. The temporal dynamics of opportunity costs: a normative account of cognitive fatigue and boredom. *Psychol. Rev.* **129**, 564–585 (2022).
- Eastwood, J. D., Frischen, A., Fenske, M. J. & Smilek, D. The unengaged mind: defining boredom in terms of attention. *Perspect. Psychol. Sci.* **7**, 482–495 (2012).
- Robison, M. K., Miller, A. L. & Unsworth, N. A multi-faceted approach to understanding individual differences in mind-wandering. *Cognition* **198**, 104078 (2020).
- Killingsworth, M. A. & Gilbert, D. T. A wandering mind is an unhappy mind. *Science* **330**, 932 (2010).
- Fox, K. C., Thompson, E., Andrews-Hanna, J. R. & Christoff, K. Is thinking really aversive? A commentary on Wilson et al.'s "Just think: the challenges of the disengaged mind". *Front. Psychol.* **5**(DEC), 10–13 (2014).
- Fox, K. C. et al. Affective neuroscience of self-generated thought. *Ann. N. Y. Acad. Sci.* **1426**, 25–51 (2018).
- van Hooff, M. L. & van Hooft, E. A. Boredom at work: proximal and distal consequences of affective work-related boredom. *J. Occup. Health Psychol.* **19**, 348–359 (2014).
- Miner, A. G. & Glomb, T. M. State mood, task performance, and behavior at work: a within-persons approach. *Organ. Behav. Hum. Dec. Process.* **112**, 43–57 (2010).
- Camille, N. et al. The involvement of the orbitofrontal cortex in the experience of regret. *Science* **304**, 1167–1170 (2004).
- Eldar, E., Rutledge, R. B., Dolan, R. J. & Niv, Y. Mood as representation of momentum. *Trends Cogn. Sci.* **20**, 15–24 (2016).
- Vinckier, F., Rigoux, L., Oudiette, D. & Pessiglione, M. Neuro-computational account of how mood fluctuations arise and affect decision making. *Nat. Commun.* **9**, 1708 (2018).
- Liuzzi, L. et al. Magnetoencephalographic correlates of mood and reward dynamics in human adolescents. *Cerebr. Cortex* **32**, 3318–3330 (2022).
- Bedder, R. L., Vaghi, M. M., Dolan, R. J. & Rutledge, R. B. Risk taking for potential losses but not gains increases with time of day. *PsyArXiv* <https://doi.org/10.31234/osf.io/3qdnx> (2020).
- Grilli, L. & Rampichini, C. Specification of random effects in multilevel models: a review. *Qual. Quant.* **49**, 967–976 (2015).

35. Schielzeth, H. et al. Robustness of linear mixed effects models to violations of distributional assumptions. *Methods Ecol. Evol.* **11**, 1141–1152 (2020).
36. Feingold, A. Confidence interval estimation for standardized effect sizes in multilevel and latent growth modeling. *J. Consult. Clin. Psychol.* **83**, 157 (2015).
37. Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G. & Fava, M. Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task. *J. Psychiatr. Res.* **43**, 76–87 (2008).
38. Halahkoon, D. C. et al. Reward-processing behavior in depressed participants relative to healthy volunteers: a systematic review and meta-analysis. *JAMA Psychiatr.* **77**, 1286–1295 (2020).
39. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Routledge, 2013).
40. Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D. & Mermelstein, R. J. A practical guide to calculating Cohen's f^2 , a measure of local effect size, from PROC MIXED. *Front. Psychol.* **3**, 111 (2012).
41. Isen, A. M. & Patrick, R. The effect of positive feelings on risk taking: when the chips are down. *Organ. Behav. Hum. Perform.* **31**, 194–202 (1983).
42. Arkes, H. R., Herren, L. T. & Isen, A. M. The role of potential loss in the influence of affect on risk-taking behavior. *Organ. Behav. Hum. Dec. Process.* **42**, 181–193 (1988).
43. Schulreich, S. et al. Music-evoked incidental happiness modulates probability weighting during risky lottery choices. *Front. Psychol.* **4**, 981 (2014).
44. Hunter, J. A., Dyer, K. J., Cribbie, R. A. & Eastwood, J. D. Exploring the utility of the Multidimensional State Boredom Scale. *Eur. J. Psychol. Assess.* **32**, 241–250 (2016).
45. Struk, A. A., Carriere, J. S. A., Cheyne, J. A. & Danckert, J. A short boredom proneness scale: development and psychometric properties. *Assessment* **24**, 346–359 (2017).
46. Seli, P. et al. Mind-wandering as a natural kind: a family-resemblances view. *Trends Cogn. Sci.* **22**, 479–490 (2018).
47. Christoff, K. et al. Mind-wandering as a scientific concept: cutting through the definitional haze. *Trends Cogn. Sci.* **22**, 957–959 (2018).
48. Seli, P. et al. The family-resemblances framework for mind-wandering remains well clad. *Trends Cogn. Sci.* **22**, 959–961 (2018).
49. Turnbull, A. et al. The ebb and flow of attention: between-subject variation in intrinsic connectivity and cognition associated with the dynamics of ongoing experience. *NeuroImage* **185**, 286–299 (2019).
50. Mrazek, M. D., Phillips, D. T., Franklin, M. S., Broadway, J. M. & Schooler, J. W. Young and restless: validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Front. Psychol.* **4**, 560 (2013).
51. Nunokawa, J. The importance of being bored: the dividends of ennui in "The Picture of Dorian Gray". *Studies in the Novel* **28**, 357–371 (1996).
52. Shattuck, R. *Proust's Way: A Field Guide to In Search of Lost Time* (WW Norton & Company, 2001).
53. Proust, M. *Swann's Way: In Search of Lost Time* Vol. 1 (Yale Univ. Press, 2013).
54. Ciocan, C. Heidegger and the problem of boredom. *J. Br. Soc. Phenomenol.* **41**, 64–77 (2010).
55. Ratcliffe, M. in *The Cambridge Companion to Heidegger's Being and Time* (ed. Wrathall, M. A.) 157–176 (Cambridge University Press, 2013).
56. Heidegger, M. *The Fundamental Concepts of Metaphysics: World, Finitude, Solitude* (Indiana Univ. Press, 1995).
57. Schopenhauer, A. in *Parerga und Paralipomena*, Vol. 1 217 (Virtual Library, 1851).
58. Kierkegaard, S. *Either/Or: A Fragment of Life* (Penguin Classics, 1992).
59. Elpidorou, A. The bright side of boredom. *Front. Psychol.* **5**, 1245 (2014).
60. Thompson, P. M. et al. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).
61. Adhikari, B. M. et al. A resting state fMRI analysis pipeline for pooling inference across diverse cohorts: an ENIGMA rs-fMRI protocol. *Brain Imaging Behav.* **13**, 1453–1467 (2019).
62. Birn, R. M. et al. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage* **83**, 550–558 (2013).
63. Noble, S. et al. Influences on the test-retest reliability of functional connectivity MRI and its relationship with behavioral utility. *Cerebr. Cortex* **27**, 5415–5429 (2017).
64. Noble, S., Scheinost, D. & Constable, R. T. A decade of test-retest reliability of functional connectivity: a systematic review and meta-analysis. *NeuroImage* **203**, 116157 (2019).
65. Jaspers, K. in *Die abnorme Seele in Gesellschaft und Geschichte (Soziologie und Historie der Psychosen und Psychopathien)* 594–623 (Springer, 1973).
66. Schneider, K. *Klinische Psychopathologie* 14 edn (Georg Thieme Verlag, 1992).
67. Berrios, G. E. Phenomenology, psychopathology and Jaspers: a conceptual history. *Hist. Psychiatry* **3**, 303–327 (1992).
68. Westgate, E. C. & Wilson, T. D. Boring thoughts and bored minds: the MAC model of boredom and cognitive engagement. *Psychol. Rev.* **125**, 689 (2018).
69. Barrett, L. F. Feelings or words? Understanding the content in self-report ratings of experienced emotion. *J. Pers. Soc. Psychol.* **87**, 266–281 (2004).
70. Westgate, E. C. & Steidle, B. Lost by definition: why boredom matters for psychology and society. *Soc. Pers. Psychol. Compass* **14**, e12562 (2020).
71. Frijda, N. H. in *The Oxford Companion to Emotion and the Affective Sciences* (eds Sander, D. & Scherer, K. R.) 258–259 (Oxford Univ. Press, 2009).
72. Ekkekakis, P. *The Measurement of Affect, Mood, and Emotion: A Guide for Health-Behavioral Research* (Cambridge University Press, 2013).
73. Rottenberg, J. Mood and emotion in major depression. *Curr. Dir. Psychol. Sci.* **14**, 167–170 (2005).
74. Nowlis, V. & Nowlis, H. H. The description and analysis of mood. *Ann. N. Y. Acad. Sci.* **65**, 345–355 (1956).
75. Ekman, P. An argument for basic emotions. *Cogn. Emot.* **6**, 169–200 (1992).
76. Watson, D. *Mood and Temperament* (Guilford Press, 2000).
77. Diener, E. Subjective well-being: the science of happiness and a proposal for a national index. *Am. Psychol.* **55**, 34 (2000).
78. Robinson, M. D. & Clore, G. L. Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychol. Bull.* **128**, 934–960 (2002).
79. Costello, E. J. & Angold, A. Scales to assess child and adolescent depression: checklists, screens, and nets. *J. Am. Acad. Child Adolesc. Psychiatry* **27**, 726–737 (1988).
80. Pavot, W. & Diener, E. The affective and cognitive context of self-reported measures of subjective well-being. *Soc. Indic. Res.* **28**, 1–20 (1993).
81. Ebner-Priemer, U. W. & Trull, T. J. Ecological momentary assessment of mood disorders and mood dysregulation. *Psychol. Assess.* **21**, 463 (2009).
82. Siegel, E. H. et al. Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychol. Bull.* **144**, 343 (2018).

83. Gendron, M., Roberson, D. & Barrett, L. F. Cultural variation in emotion perception is real: a response to Sauter, Eisner, Ekman, and Scott (2015). *Psychol. Sci.* **26**, 357–359 (2015).
84. Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M. & Pollak, S. D. Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychol. Sci. Public Interest* **20**, 1–68 (2019).
85. Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E. & Barrett, L. F. The brain basis of emotion: a meta-analytic review. *Behav. Brain Sci.* **35**, 121–143 (2012).
86. Paolacci, G., Chandler, J. & Ipeirotis, P. G. Running experiments on Amazon mechanical turk. *Judgm. Dec. Making* **5**, 411–419 (2010).
87. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
88. Ho, N. S. P. et al. 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**, 116765 (2020).
89. Adler, N. E., Epel, E. S., Castellazzo, G. & Ickovics, J. R. Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. *Health Psychol.* **19**, 586–592 (2000).
90. Singh-Manoux, A., Marmot, M. G. & Adler, N. E. Does subjective social status predict health and change in health status better than objective status? *Psychosom. Med.* **67**, 855–861 (2005).
91. Radloff, L. S. The CES-D scale: a self-report depression scale for research in the general population. *Appl. Psychol. Meas.* **1**, 385–401 (1977).
92. Snaith, R. P. et al. A scale for the assessment of hedonic tone. The Snaith–Hamilton Pleasure Scale. *Br. J. Psychiatry* **167**, 99–103 (1995).
93. Angold, A., Costello, E. J., Messer, S. C. & Pickles, A. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *Int. J. Methods Psychiatr. Res.* **5**, 237–249 (1995).
94. Birmaher, B. et al. Psychometric properties of the screen for child anxiety related emotional disorders (SCARED): a replication study. *J. Am. Acad. Child Adolesc. Psychiatry* **38**, 1230–1236 (1999).
95. Jolly, E. Pymer4: connecting R and Python for linear mixed modeling. *J. Open Source Softw.* **3**, 862 (2018).
96. Snijders, T. A. B. & Bosker, R. J. Modeled variance in two-level models. *Sociol. Methods Res.* **22**, 342–363 (1994).
97. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R2 from generalized linear mixed effects models. *Methods Ecol. Evol.* **4**, 133–142 (2013).
98. Barton, K. MuMIn: multi-model inference. *R Project* <http://r-forge.r-project.org/projects/mumin/> (2009).
99. Paszke, A. et al. Pytorch: an imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32**, 8026–8037 (2019).
100. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. *arXiv* <https://doi.org/10.48550/arXiv.1412.6980> (2014).
101. Rutledge, R. B. Risky decision and happiness task: The Great Brain Experiment smartphone app. *Dryad* <https://doi.org/10.5061/dryad.prr4xgxxk> (2021).

Acknowledgements

This research was supported in part by the Intramural Research Program of the National Institute of Mental Health, part of the NIH (grant nos. ZIA MH002957 (to A.S.), ZIC MH002968 (to F.P.), ZIA MH002871 (to D.S.P.), ZIA MH002872 (to D.S.P.) and ZIC MH002960 (to A.G.T.)). This work used the computational resources of the NIH high-performance computing (HPC) Biowulf cluster (<http://hpc.nih.gov>). Data collection for the mobile app dataset was supported by the Wellcome Trust (grant no. 101252/Z/13/Z). The online adolescent sample in this study was collected under NIH IRB protocol number 18-M-0037, registered on clinicaltrials.gov as NCT03388606. The online adult sample was collected under NIH Office of Human Subjects Research Protection protocol P194594. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The views expressed in this article do not necessarily represent the views of the NIH, the Department of Health and Human Services or the United States Government.

Author contributions

D.C.J., H.K., D.M.N. and A.S. devised the task. D.C.J. wrote the online experiments. D.C.J. and H.S. collected the online data. R.L.B. and R.B.R. provided data and information from the mobile app experiments. C.Z. and F.P. devised the computational model. D.C.J., C.Z. and D.M.N. wrote analysis code. D.C.J. and D.M.N. ran the analyses. D.C.J., D.M.N. and A.S. wrote the manuscript. All authors provided revisions and finalized the text.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41562-023-01519-7>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41562-023-01519-7>.

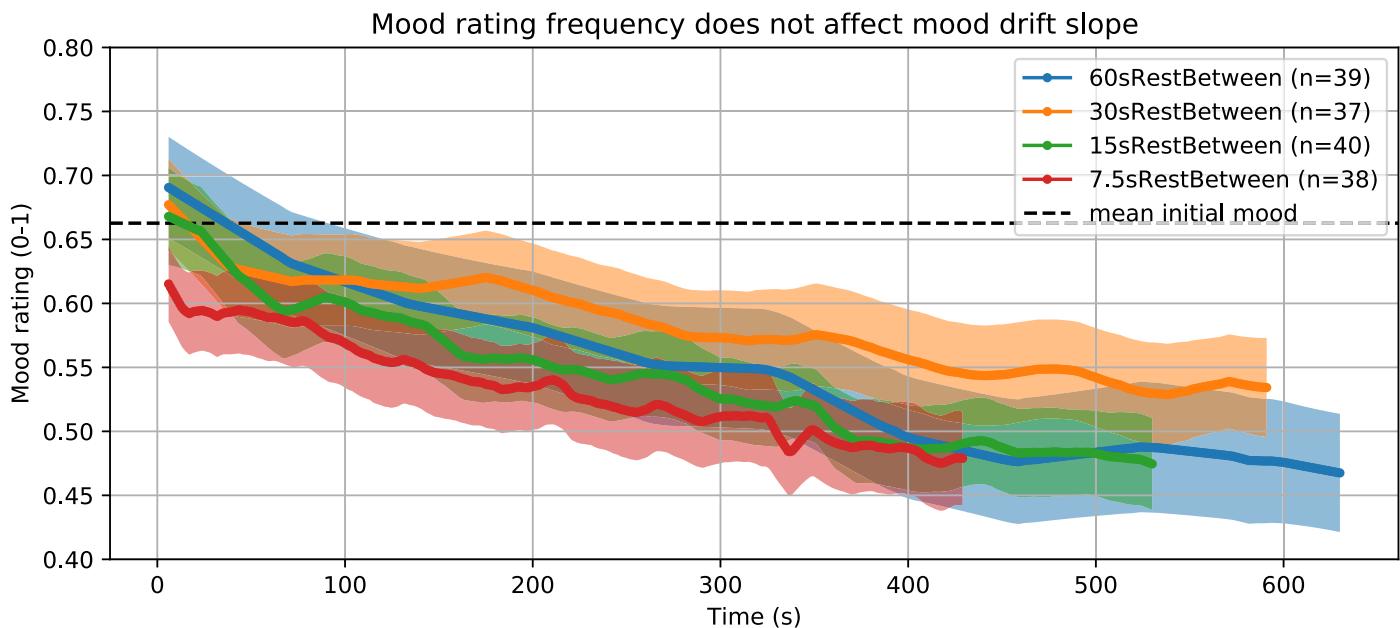
Correspondence and requests for materials should be addressed to David C. Jangraw.

Peer review information *Nature Human Behaviour* thanks Aaron Heller, Caitlin Mills and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

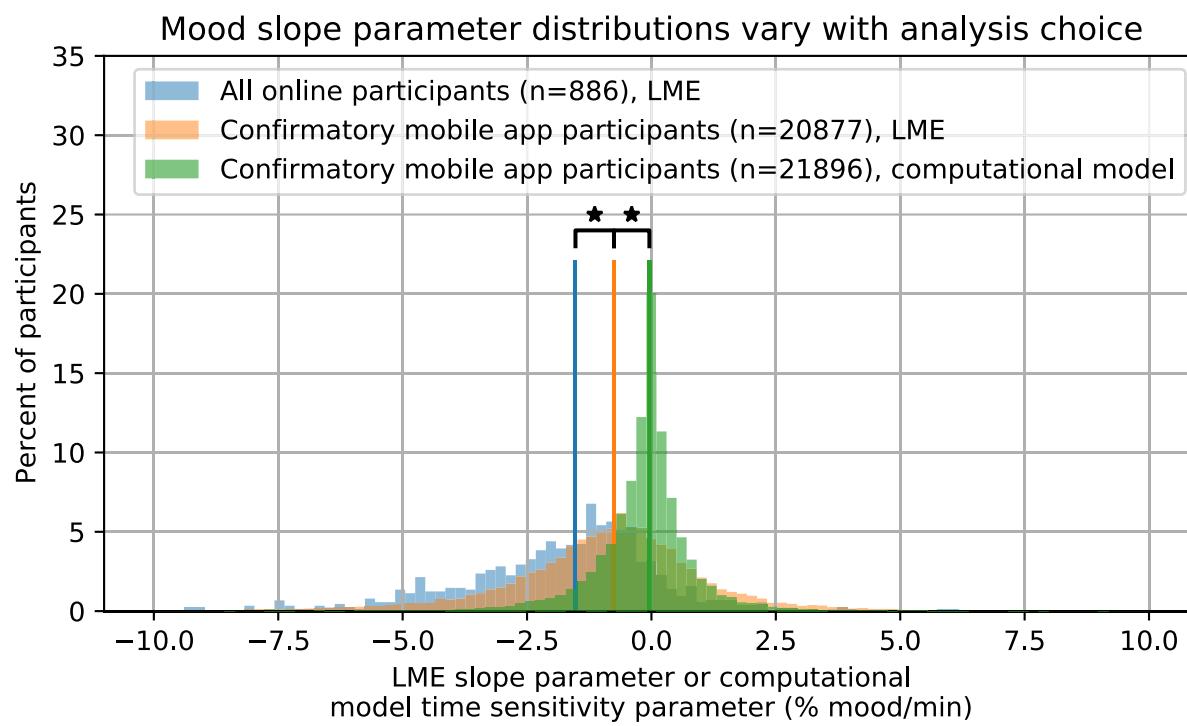
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023

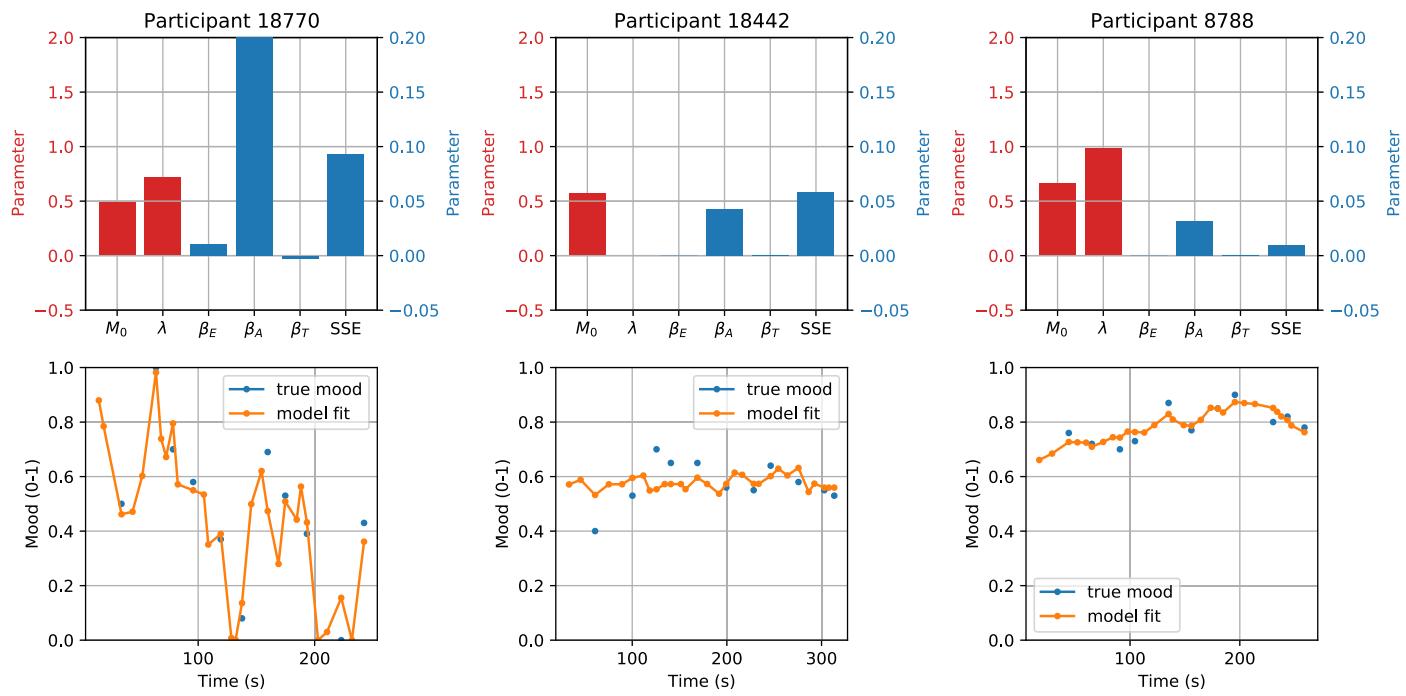


Extended Data Fig. 1 | Mood rating frequency does not affect mood drift slope. Mean \pm SE mood rating at each time in the 4 cohorts with 60 s, 30 s, 15 s, and 7.5 s of rest between mood ratings (cohorts 60sRestBetween, 30sRestBetween, 15sRestBetween, and 7.5sRestBetween, respectively). The magnitude of mood drift did not vary with the frequency of mood ratings.

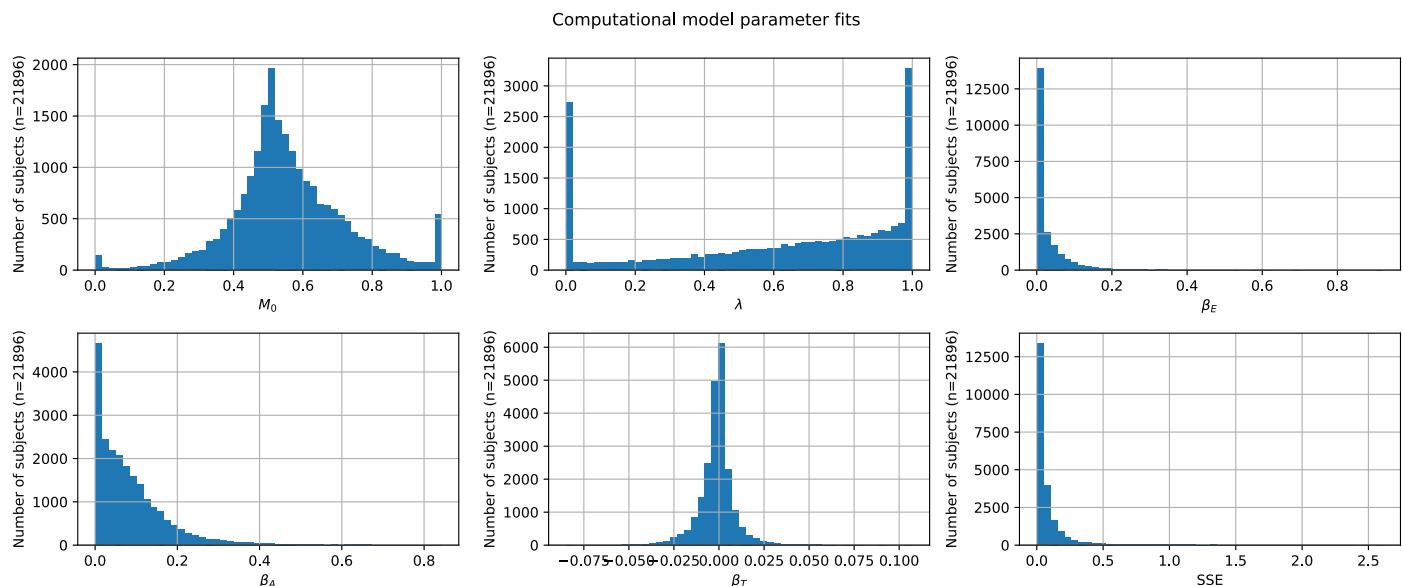


Extended Data Fig. 2 | Mood slope parameter distributions vary with analysis choice. Histogram of the LME mood slope parameters for the online cohort (blue) and the confirmatory mobile app cohort (orange), along with the computational model time sensitivity parameter for the confirmatory mobile app cohort (green). Mobile app participants with outlier task completion times were excluded from the LME analysis (see Methods). Note that the use of LME modeling to analyze the mobile app data significantly lowered the distribution of

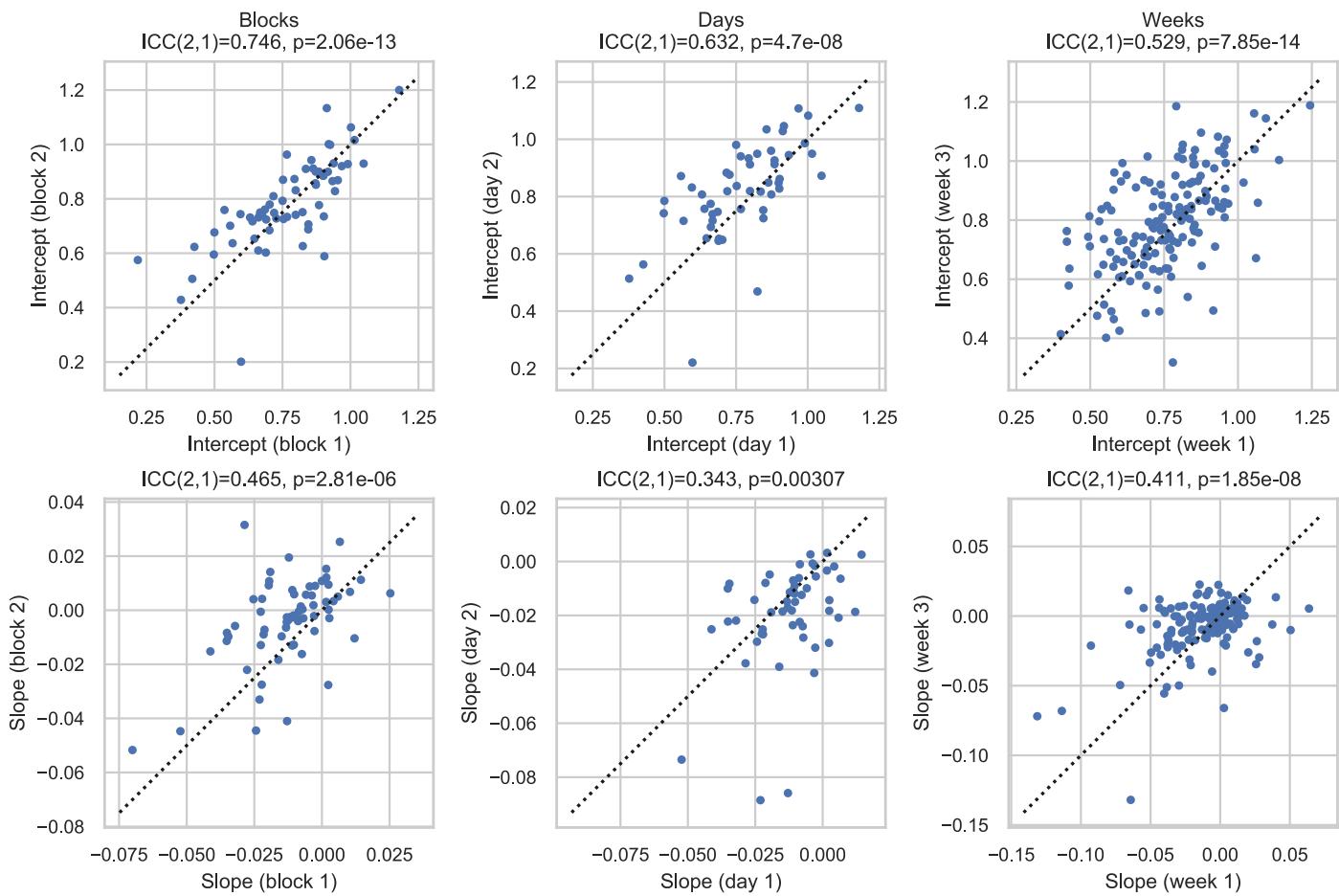
slopes compared to when the computational model was used (median = -0.752 vs. -0.0408, IQR = 2.10 vs. 0.764 %mood min⁻¹, 2-sided Wilcoxon rank-sum test, $W_{42771} = -54.2$, $p < 0.001$), but the LME slopes from the mobile app were still significantly greater than those of the online cohort (median = -1.53 vs. -0.752, IQR = 2.34 vs. 2.1 %mood min⁻¹, 2-sided Wilcoxon rank-sum test, $W_{21761} = 14.5$, $p < 0.001$). Vertical lines represent group medians. Stars indicate $p < 0.05$. P values were not corrected for multiple comparisons.



Extended Data Fig. 3 | Sample fits of the computational model. 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.

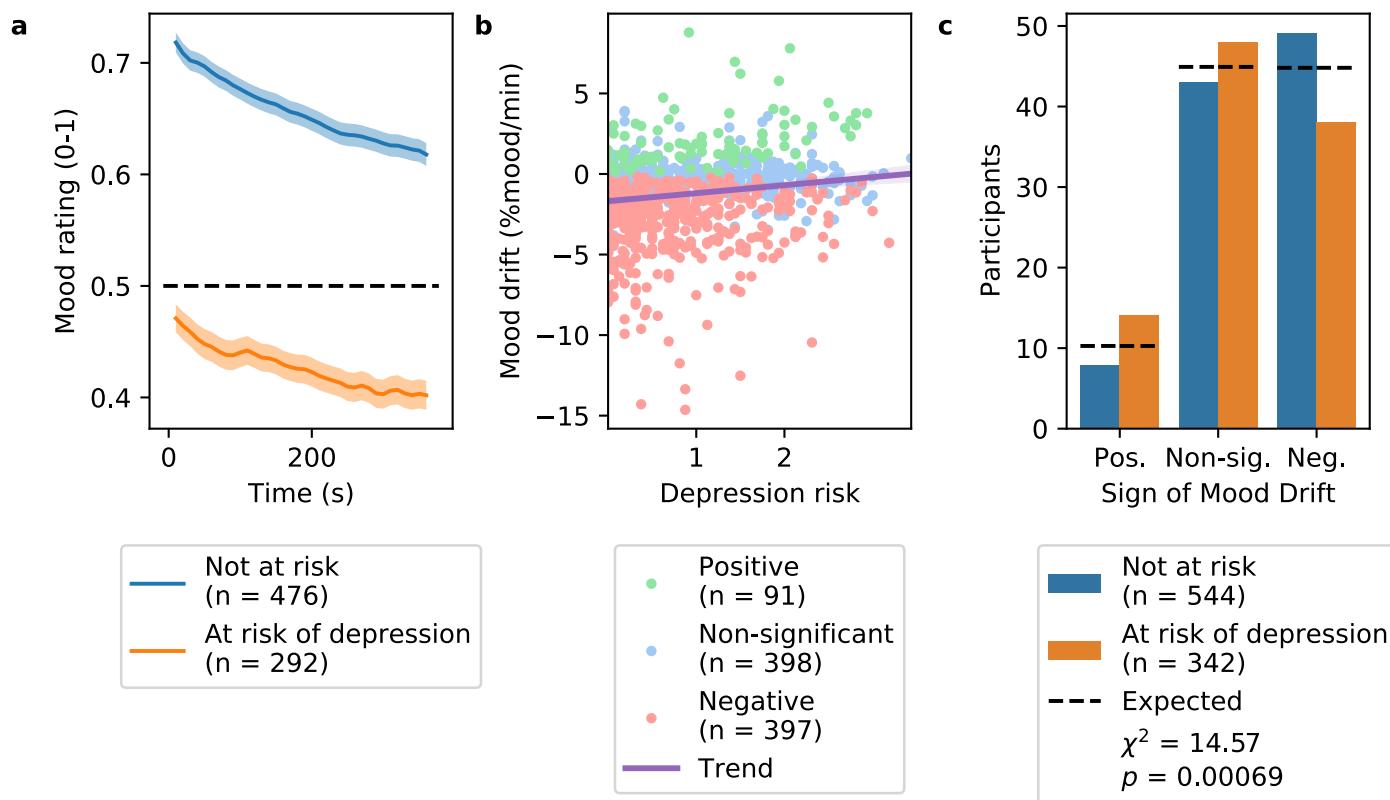


Extended Data Fig. 4 | Histogram of computational model parameters. Histogram of computational model parameters across the 21,896 confirmatory mobile app subjects.


Extended Data Fig. 5 | Mood drift stability over blocks, days, and weeks.

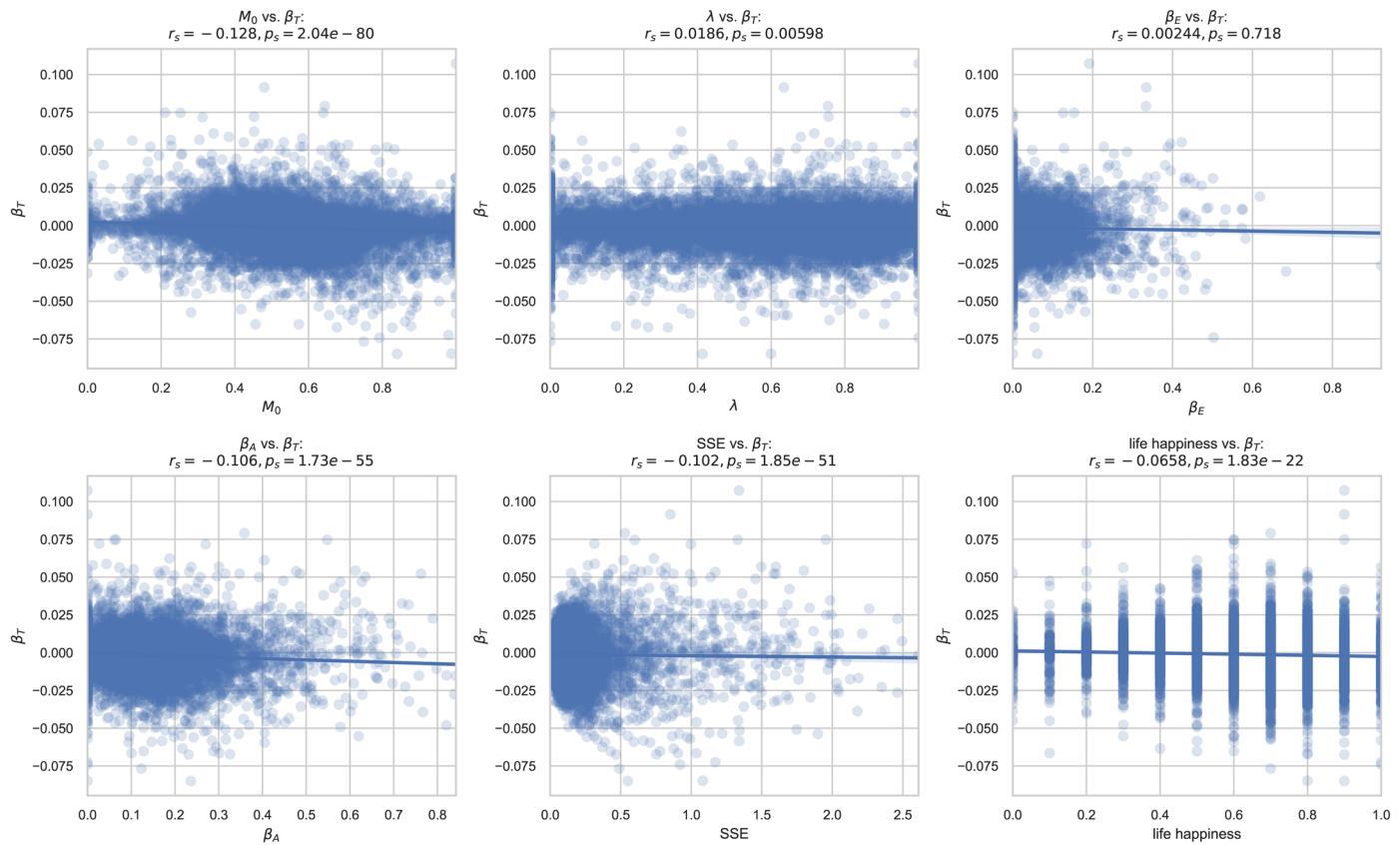
Stability of LME coefficients estimating the initial mood (top) and slope of mood over time (bottom) for each participant across rest periods one block apart (left), 1 day apart (middle), and 2 weeks apart (right). ICC denotes the intra-

class correlation coefficient for each comparison. P values shown are one-sided (since ICC values are expected to be positive) with no correction for multiple comparisons.


Extended Data Fig. 6 | Relationship between mood drift and depression risk.

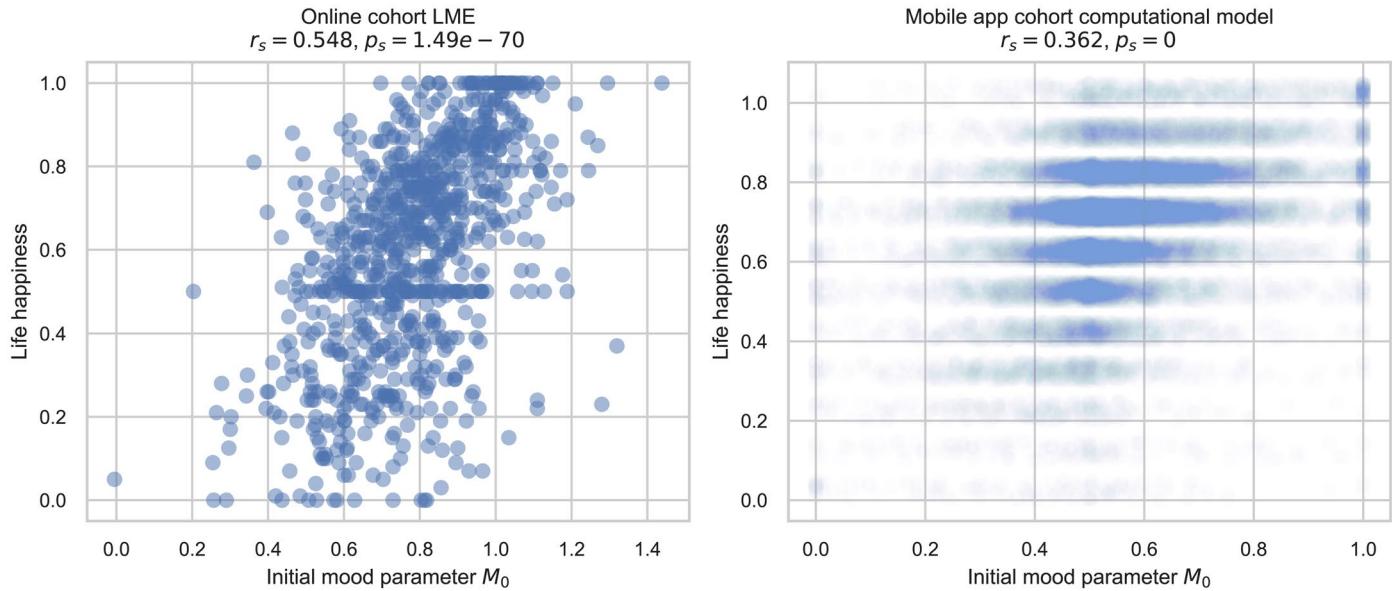
Relationship between mood drift and depression risk. (a) Mood ratings over time of online participants at risk of depression (defined as MFQ>12 or CES-D>16) vs. those not at risk for the 768 participants with at least 6 minutes of resting mood data (error bars are SEM). The dotted line represents the 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 (2-sided $\alpha = 0.5$, $p < 0.05$) 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). The line is a linear best fit, and the patch shows the 95% confidence interval of this fit. (c) Proportion of individuals with or without risk of depression (that is, depression risk >1 or <1) with positive (significantly greater than zero), non-significant (no evidence of a significant difference from zero), and negative (significantly less than zero) 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, $\chi^2(1, N=886)=14.57$, $p < 0.001$ (2-sided Pearson's chi-squared statistic with no correction for multiple comparisons).



Extended Data Fig. 7 | Mood drift's relation to other computational model parameters. Time sensitivity parameter β_T vs. other parameters in the confirmatory mobile app cohort. Each dot is a participant ($n=21,896$). Each line

is a linear best fit, and patches show the 95% confidence interval of this fit. r_s denotes Spearman correlation coefficient. P values shown are 2-sided with no correction for multiple comparisons.


Extended Data Fig. 8 | Initial mood parameter's relation to life happiness.

Initial mood parameter vs. life happiness rating in the online cohort (left) and the confirmatory mobile app cohort (right). Life happiness ratings were always multiples of 0.1; small positive random values were added during plotting to

reduce overlap between data points. Each dot is a participant (left: n=886, right: n=21,896). r_s denotes Spearman correlation coefficient. P values shown are 2-sided with no correction for multiple comparisons.

Extended Data Table 1 | List and description of cohorts collected

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		
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	

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 cycles used in that block. "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 cycles 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 (for example, 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.

Extended Data Table 2 | Linear mixed effects (LME) model results

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	

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 mood drift. 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 2-sided p value. All values are rounded to 3 decimal places. The Sig (significance) column contains * if p<0.05.

Corresponding author(s): David Jangraw

Last updated by author(s): Dec 9, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

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. The code for the data collection task and survey are now available at <https://gitlab.pavlovia.org/mooddrift>.

Data analysis

Data were analysed using custom Python 3 (version 3.8.8) software depending on several Python packages, including the Pymer4 package (version 0.7.3). This software, as well as the means to create a Python environment that automatically installs it on a user's machine, has been made available at <https://github.com/djangraw/MoodDrift>. As stated on that site, the dependencies and their versions are: python=3.8.8 numpy=1.19.2 pandas=1.1.5 pytest=6.2.2 joblib=1.0.1 rpy2=3.4.3 matplotlib=3.3.4 seaborn=0.11.1 scikit-learn=0.24.1 numexpr=2.7.3 patsy=0.5.1 statsmodels=0.12.2 openpyxl=3.0.7 pymer4=0.7.3 r-psych=2.1.3 xlrd=2.0.1 r-mumin=1.43.17.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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/km69z>. Mobile App Participants' data can be found on Dryad at <https://doi.org/10.5061/dryad.prr4gxkk>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

The study consists primarily of quantitative cross-sectional data about the subjective mood of participants and how patterns in these mood ratings relate to their depression risk as assessed by clinical questionnaires. Some cohorts of participants completed slightly modified versions of the task to assess these patterns' sensitivity to changes in the task. These cohorts always consisted of naive participants to reduce expectation effects, and they were, in general, recruited in series rather than randomized to a condition on the same day. Some participants, however, were asked to return multiple times to assess the stability of their mood rating behaviour.

Research sample

There are four distinct research samples in this study. 1) Original cohort of Online Adult Participants recruited using Amazon Mechanical Turk. Participants were required to be adults living in the United States who had completed over 5,000 jobs for other requesters with a 97% satisfactory completion rate. 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).

2) Online adolescent participants recruited in person at the National Institute of Mental Health. 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. 129 adolescents completed the task. 10 adolescents who had not completed all three questionnaires were excluded from the results as were 3 participants 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 (\pm 5.5 SD) and a mean anxiety score of SCARED = 2.2 (\pm 3.0 SD).

3) 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, largely from the US and UK in our analyses. The median life satisfaction of the included subjects, 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%).

4) Follow-Up Cohort: a cohort of online adult participants recruited identically to Sample #1 to answer specific preregistered hypotheses about boredom, mind-wandering, and freely chosen activities. 1143 participants completed the task online in this cohort. 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).

We recruited the adult (MTurk) and an adolescent sample so as to ensure that we cover both age ranges and can therefore draw inferences about any developmental differences. The adult (MTurk) sample was collected online and should be considered a convenience rather than a nationally representative sample. The adolescent sample was collected so as to be enriched for mental health difficulties, MDD in particular. This enrichment allows us also to draw inferences about the relationship between depression and passage of time dysphoria in that age range. Finally, the Mobile App Participant sample should also be considered a convenience rather than a nationally representative sample.

Sampling strategy

Cohorts 1-3 are convenience samples. No sample-size calculations were performed a priori. Sample sizes for each batch of the Online Adult Participants was ad hoc. All together, nearly 1,000 participants were recruited on MTurk and in person, which was sufficient to demonstrate the mood drift over time effect repeatedly both within cohorts and within individuals. The Online Adolescent data and a subset of the Online Adult Participants were asked to complete the task multiple times on different days or different weeks, all other data was collected cross-sectionally. The adolescent online participants were recruited from an ongoing study characterizing adolescent depression.

The Follow-Up Cohort sample sizes were selected using power calculations described in detail in the preregistration. For the scale validation experiments, a sample size of 150 in each group with an alpha of 0.01 gives 99.02 power to detect a medium effect ($d = 0.5$) and 83.04% power to detect an intermediate effect ($d = 0.3$) assuming the effect truly is null at a population level. Power for linear multiple regression tests were calculated in G*Power. In the boredom and MW cohorts, samples of 150 participants were selected to provide 80% power to detect a 7.99% increase in variance explained with the inclusion of a single parameter (alpha = 0.01, 20 total predictors) and a 95% power to detect a 12.18% change in variance explained. In analyses using a pair of cohorts, 300 participants gives 80% power to detect a 3.93% increase in variance explained and a 95% power to detect a 6.01% increase in variance explained. An Activities cohort of 450 participants was chosen to provide 80% power to detect a difference between the Activities and MTurk cohorts of Cohen's $d = 0.2$, and it also provides 80% power to detect a decrease in mood in the Activities cohort of Cohen's $d = 0.15$.

Data collection

All data were collected online. Data from Mobile App Participants was collected through a mobile app, all other data was collected online through a web browser. No members of the study team were present when participants were completed forms or tasks. Researchers were not blinded to experimental condition or study hypotheses.

Timing

Data for the original online adult sample was collected in batches from September 27th, 2019 to June 19th, 2020. Data from the online adolescent sample was collected from April 6th, 2020 to June 5th, 2020. Data from the mobile app participants was collected from January 4th, 2013 to December 9th 2015. Data from the follow-up cohorts (Activities, Mind-wandering, and Boredom) were collected from December 9, 2021 to January 8, 2022.

Data exclusions

Data from the original cohort of Online Adult Participants were excluded if data was missing or incomplete or they failed to answer catch questions correctly, which impacted 44 participants. In the Online Adolescent Participants, 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. In linear mixed effects models on the Mobile App Participants, we excluded participants whose average response times were more than 1.5 times the interquartile range below the first quartile or above the third quartile. In the 5000 exploratory participants, 217 had long average response times and 15 had short average reaction times for a total of 232 participants excluded. In the 21,896 confirmatory participants 981 had long average response times and 38 had short average reaction times for a total of 1019 participants excluded. In the Follow-Up Cohort participants, 93 participants whose data was missing or incomplete or who failed to answer catch questions correctly were excluded.

Non-participation

There were 230 adolescents enrolled in the NIMH depression characterization study who were offered to complete tasks for this study as Online Adolescent Participants. 129 agreed to complete tasks for this study, a participation rate of 56.1%. 82 (70.6%) individuals completed the task a week later and 4 (3.4%) completed the task a third time the following week. Due to the low participation rate in the third visit, only the first two were used. There were two longitudinal samples in the Adult Online cohort. One subset was asked to return 1 day later (Daily-Rest). 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. One subset was asked to return weekly for three weeks (Weekly-Rest). 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 on all three waves.

Randomization

Participants in the Expectation cohorts (a subset of the original Online Adult Participants) were randomised to the Expectation-7mRest or Expectation-12mRest cohort at the time of participation. Participants in the Follow-Up Cohort were randomised to one of 5 groups (BoredomBeforeAndAfter, BoredomAfterOnly, MwBeforeAndAfter, MwAfterOnly, or Activities) at the time of participation. There was no random allocation of participants between conditions in the other cohorts because they were collected in series to answer an evolving set of questions about mood drift over time.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	See above
Recruitment	See above
Ethics oversight	<p>Because we did not obtain information by direct intervention or interaction with the Adult Online 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.</p> <p>Adolescent participants recruited in person at the National Institute of Mental Health provided informed consent to a protocol approved by the NIH Institutional Review Board.</p> <p>Gambling behaviour and mood rating data were collected from the Mobile App Participants in a mobile app called "The Great Brain Experiment", described in (Rutledge, 2014). The Research Ethics Committee of University College London approved the original study.</p>

Note that full information on the approval of the study protocol must also be provided in the manuscript.