nature research

Peer Review Information

Journal: Nature Human Behaviour
Manuscript Title: A Highly Replicable Decline in Mood During Rest and Simple Tasks
Corresponding author name(s): David C. Jangraw

Reviewer Comments & Decisions:

Decision Letter, initial version:

13th August 2021

Dear Dr Jangraw,

Thank you once again for your manuscript, entitled "Passage-of-Time Dysphoria: A Highly Replicable Decline in Mood During Rest and Simple Tasks that is Moderated by Depression", and for your patience during the peer review process.

Your manuscript has now been evaluated by 2 reviewers, whose comments are included at the end of this letter. In the light of their advice, I regret that we cannot offer to publish your manuscript in Nature Human Behaviour.

While the reviewers find your work of interest, they each raise concerns about the strength of the conclusions that can be drawn at this stage. Both reviewers state that the results presented here are not sufficient to support a conclusion that passage of time dysphoria is a generalizable phenomenon. We feel that these reservations are sufficiently important as to preclude publication of this work in Nature Human Behaviour.

I am sorry that we cannot be more positive on this occasion but hope that you will find our reviewers' comments helpful when preparing your paper for submission elsewhere.

Sincerely,
Jamie

Dr Jamie Horder
Senior Editor
Nature Human Behaviour

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:

Passage-of-Time Dysphoria
This paper argues that there is an unrecognized, general phenomenon: mood declines over time. The paper reaches this condition after analyzing data from several studies involving different procedures and samples and methods of measuring mood. Most of the studies do not seem to have been designed with the goal of testing for passage-of-mood dysphoria.

My general appraisal of the paper is that it failed to convince me that Passage-of-Time Dysphoria is a robust, systematic phenomenon. If it were, would not expect the population's mood to decline systematically over time? If it's just a phenomenon that researchers need to take account of in extended studies, it doesn't seem all that interesting and important.

The paper is a bit strange. First, "mood" is measured in a variety of different ways in the different studies. One of the measures I found especially suspect was the one in which a participant had to click repeatedly to move an indicator from very "very unhappy" to "very happy." What if participants were energetic and motivated early in the study, but got lazy later? Expressing happiness took more work (more key presses), so people who were unmotivated would be likely to press fewer times.

The construct of mood, moreover, is awfully vague. In some of the studies it seems to be picking up boredom (e.g., the condition in which participants stared at a fixation cross for an usually long period of time). In others, it might just be annoyance with the repetitiveness of the experiment.

All of the studies are between-subject, probably for reasons of statistical power. The same participant is asked the same question repeatedly over time. The problem with this is that the early questions are likely to cue the participants to the researcher's interest in mood/happiness. But, research has shown that when people pay attention to how happy they are, they tend to become less happy; so may it's not a passage of time effect, but an asking repeatedly effect. The paper claims to have created a modified version of the task to rule out that the decline in mood is due to "the aversive nature of rating one's

mood," but I didn't see such a version. Even if there was such a version, I'd want to know that there weren't other factors, such as an increase in boredom, leading to a decline in reported well-being.

The paper makes quite a big deal of two results: First, people with clinically depressed affect did not show the effect as strongly. Second, the drop in affect coincided with a decrease in gambling. The problem with both results is that they were not predicted ahead of time; as the paper acknowledges, there are good reasons that both results could have gone one way or the other way. So it isn't clear what one should make of either. Is it possible that there was a floor effect for participants who were clinically depressed?

In the Mturk study, it concerned me that a fairly large fraction of participants' data seems to have been lost – those who closed the task window before being asked to do so. Were they different in some way from those who stuck it out (admittedly, this selection effect would probably bias the results in the opposite of the mood dysphoria effect).

The paper, which follows participants up in a set of researcher-selected studies in which they were asked some kind of mood-related question repeatedly over time, raises a broad methodological issue. How should we sample the kinds of situations in which participants are waiting, to feel confident in reaching the broad conclusion that there is a Passage-of-Time Dysphoria effect. It seems to me that a clever experimenter who was motivated to demonstrate a Passage-of-Time Euphoria effect could probably devise experiments that would obtain such a result – e.g., by making an experiment really interesting and engaging. So, what are we to make of the fact that the current set of experimenters found increases in dysphoria over time? My bottom line is that I'm simply not convinced that this is an important, generalizable, result.

Reviewer #2:
Remarks to the Author:

Jangraw and colleagues present an interesting analysis of a large cohort of multiple studies and demonstrate that mood ratings decrease as a function of time on task. They coin a term, 'Passage-of-Time Dysphoria' for this phenomenon. They provide evidence that the rate of decline is shallower for individuals with higher depression levels and that there was a link between a reward sensitivity parameter and the slope of the mood decline as a function of time on task. While a highly stable effect with some nice controls (in particular varying the frequency of mood ratings), the authors briefly note in the discussion this finding could be considered or accounted for by many things, previously described and demonstrated in the literature, including mind-wandering or boredom.

I agree with the overall statement that affective scientists should be aware of this effect, insofar that, as psychologists, most of our tasks are incredibly boring and do not simulate the real-world at all. However, apart from a narrow correlational analysis examining choices to gamble after rest-blocks, the authors do not present any mechanistic data to show how this effect can be pushed around to help the reader understand what is causing the phenomenon (and whether or not it is boredom / mind-wandering / task agency). I would find this submission more compelling for this outlet were there to be experimental exploration of these effects.

I also have questions in terms of generalizability. For instance, I would also find this piece more compelling if other real-world scenarios yielded similar effects. What about so-called flow experiences? Or do gamers playing Warcraft for hours on end experience similar decrements in mood? I'm not suggesting that the authors *have* to do a study in this domain, but some consideration or empirical data speaking to other arenas and the relevance (or irrelevance) of this finding to those domains would be helpful.

A few additional comments below:

The authors note that the effects for the mobile task were significantly weaker than the other 'computer-based' tasks. What is different between the mobile app and being in a lab (or on MTurk)? One hypothesis is that you can simply exit the task more easily when playing the game on your phone. How long did people tend to play via their phone? That curve is not displayed in Fig 1 (only the slope parameter in Fig 2), and not displayed in the supplement. Does agency / ease of ability to exit impact this effect? Can the authors rule this out or address this?

At the outset of the discussion the authors write, "Our findings show that subjects incorporate information about the passage of time into their ratings of mood." How do authors know that it is this specifically and not simply boredom or some other process as they allude to later? Some of the stronger claims of this paper could be walked back.

Could passage-of-time dysphoria x depression effects be due to floor effects? I recognize that individuals 'could' rate themselves below 0.3 , 0.4, but it is rare to have individuals rate themselves at an absolutely horrible mood while in psychology studies and, given that depression was linked to lower initial mood ratings, a plausible hypothesis is simply that they had lower to drop and therefore the slope was more shallow.

This may not matter for model parameters (particularly given the cellphone subsample using large sample size), but M0 and other parameters from the computational model are not normally distributed. BT is highly kurtotic. Moreover, it is clear that individuals rated their mood in a nonnormal fashion (some ratings at 0, some at 1 and a lot right in the middle at .5). Using these types of data, I've found superior

fitting models using cumulative families than assuming the outcome space is gaussian. Further, because the outcome space is bounded (0,1) a gaussian model will yield 'posterior' predictions well outside the possible response space. Just something to consider.

For Fig 1, dotted line at mood = 0.5 is sort of meaningless because, while it is the middle of the scale (and assumed by authors to be a 'neutral mood'), it isn't the mean of subject's mood ratings. It's more likely that a 'neutral' mood is actually the mean across these ratings. As most individuals generally display a positivity bias in affect.

Author Rebuttal to Initial comments

Overview: In this document, we reproduce the comments of the two reviewers in grey. After each paragraph or section, we present a detailed response (in a numbered list for easy reference). Where appropriate, we reproduce or reference specific changes that have been made to the manuscript text.

Reviewer #1:
Remarks to the Author:
Passage-of-Time Dysphoria
This paper argues that there is an unrecognized, general phenomenon: mood declines over time. The paper reaches this condition after analyzing data from several studies involving different procedures and samples and methods of measuring mood. Most of the studies do not seem to have been designed with the goal of testing for passage-of-mood dysphoria.

My general appraisal of the paper is that it failed to convince me that Passage-of-Time Dysphoria is a robust, systematic phenomenon. If it were, would not expect the population's mood to decline systematically over time? If it's just a phenomenon that researchers need to take account of in extended studies, it doesn't seem all that interesting and important.

1. Both reviewers expressed concerns about the real-world generalizability of the passage-of-time dysphoria effect. Reviewer 1 (paragraph 2) suggested that If passage-of-time dysphoria were a robust, systematic phenomenon, the population's mood would decline systematically over time. Reviewer 2 (paragraph 3) noted that the manuscript did not explore the impact of real-world scenarios, where task agency is increased.
    a. We share the reviewers' important questions about the boundaries of this phenomenon in life outside of psychological studies. This part of life is, of course, challenging to explore in a psychological study. Nevertheless, these comments helped us realise that our manuscript's absence of any results *lacking* passage-of-time dysphoria could imply to readers the suggestion that it was universally occurring in all situations. (This, as Reviewer 1 pointed out,

would lead to a perpetual mood decline.)

b. We therefore preregistered (https://osf.io/gt7a8), collected, and analysed new data that gave (n=450) participants task agency and the freedom to choose their own activity. In this task, participants rated their mood immediately before and after a 7-minute break in which they were able to leave the task and do anything they chose. Results suggest that people do not experience a mood decline when they are freely able to choose their own real-life activities. This sharpens the boundaries of the passage-of-time dysphoria phenomenon, and it helps explain why the population's mood does not decline systematically over time.

c. The contrast between the scenarios that induced stability and declines in mood is interesting. In an environment where people are able to choose the type and degree of stimulation they receive, their mood (on average) does not significantly increase or decrease. Conversely, in situations where there are restrictions such as in psychological experiments with relatively low stimulation (keeping in mind that this "low stimulation" includes engagement with a mobile app), a mood decline is evident with substantial effect sizes. On average, 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.

d. This new result is described in the manuscript in the new results section labelled "Passage-of-Time Dysphoria Is Not Present in Freely Chosen Activities." The new text reads:

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

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

   iii. *This group was the first to not exhibit passage-of-time dysphoria. The mood ratings just after the free period were not statistically different from the mood ratings before the free period (Mean pre-break mood: (65.7%), post-break mood: (66.6%), change in mood: (0.13%/min), difference t_{449}=-1.33, p_{H0:decrease}=0.0918, p_{H0:increase}=0.908). The change in mood was significantly lower (i.e., more negative) for a cohort who received the standard rest period with interspersed mood ratings (cohort BoredomAfterOnly, n=150) (t_{598}=6.28, p=3.23e-10). This shows that, perhaps unsurprisingly, passage-of-time dysphoria is not*

universal to all activities. However, the nominal increase in mood during this period (0.130% mood/min) was much smaller than the decrease in mood observed during a typical rest period in our task (-1.89% mood/min). Put another way, each minute in which participants could choose their activity raised their collective mood less than 10% of the mood decline experienced during a minute of rest.

2. Reviewer 1 (paragraph 2) suggested that a phenomenon most relevant to extended psychology studies would be of limited interest to readers.
    a. We believe that an effect observed during rest and simple tasks can be quite relevant to real-life experiences. (These tasks need not be extended: the decline in mood begins right away, as seen in Figure 1.) Psychological research relies heavily on the premise that phenomena observed during psychology studies are also present in daily life, and we believe this phenomenon to be no exception. If this is true, then we believe readers would be interested in the phenomenon as an underappreciated part of human experience. If, conversely, this effect were observed only in psychology studies, this would be of great concern to psychologists who have been unwittingly including an unnatural stressor in their tasks without accounting for it in their analyses or interpretations.
    b. This study's rest-related findings alone will be of interest to a large group of readers given the recent prominence of resting-state neuroimaging, which draws trait-level inferences from brain activity recorded from participants during extended periods of rest (like the ones presented in this study). Many imaging protocols now include resting-state scans, but the duration and timing of this rest period (and the resulting decline in mood) are not typically standardised or even reported. To highlight the potential implications of passage-of-time dysphoria for this common element of research, we have modified the examples we use in the discussion. The new text reads:

        i. *We found that mood declined during rest and multiple tasks (including a mobile app more engaging than most paradigms) but not freely chosen activities. This suggests that researchers are subjecting their participants to a somewhat unnatural stressor in their experiments without accounting for it in their analyses or interpretations. Our study's compelling evidence against the notion of a constant affective background suggests that certain methodological challenges have been overlooked in affective neuroscience. First, the inter- and intra-subject variability in the rate of mood changes over time will add to error variance if it is not accounted for in studies of mood or its correlates. Secondly, group differences in the rate of mood changes could lead to apparent group differences in later mood or behaviour. Perhaps most broadly applicable, changes in mood on the scale of tens of minutes prevent these longer blocks of time from being truly interchangeable. This means that changes to experimental procedures that might seem inconsequential could still introduce confounds. We will illustrate these ideas through some examples.*
        ii. *For example, let's consider a large collaborative study that is based on multisite imaging data collection, such as ENIGMA (Thompson, 2014). In this dataset, there is much variability across*

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

The paper is a bit strange. First, "mood" is measured in a variety of different ways in the different studies. One of the measures I found especially suspect was the one in which a participant had to click repeatedly to move an indicator from very "very unhappy" to "very happy." What if participants were energetic and motivated early in the study, but got lazy later? Expressing happiness took more work (more key presses), so people who were unmotivated would be likely to press fewer times.

3. Reviewer 1 (paragraph 3) suggested that unmotivated participants may have pressed the button fewer times during mood ratings.
    a. We share the reviewer's concern about the effect of fatigue or laziness on mood ratings. This concern motivated the control analysis in which a participant could press a single key to rate their mood as any number 1-10. This analysis was described in the supplementary materials of our original manuscript labelled "Passage-of-Time Dysphoria Is Not an Artefact of the Rating Method." To help other readers find it, we have highlighted it more prominently in the main text. This and other methodological artefacts are now examined in a new subsection labelled "Passage-of-Time Dysphoria Is Robust to Methodological Choices," which includes some existing text and some new text. The new text reads:
        i. *The impact of ratings on mood was investigated by systematically varying the frequency of mood ratings. More frequent ratings did not lead to a more rapidly declining mood. The impact of fatigue on mood ratings was investigated by making every mood rating require an equally easy single keypress. This did not change the decline in mood. The results of these and other control experiments and analyses suggested that the observed dysphoria cannot be explained by the above list of factors (see Supplementary section titled "Eliminating Methodological Confounds").*

The construct of mood, moreover, is awfully vague. In some of the studies it seems to be picking up boredom (e.g., the condition in which participants stared at a fixation cross for an usually long period of time). In others, it might just be annoyance with the repetitiveness of the experiment.

4. Reviewer 1 (paragraph 4) and Reviewer 2 (paragraph 2) suggested that the decline in mood may be a trivial extension of the established psychological constructs of boredom or mind-wandering valence.
    a. We thank the reviewers for raising this important point. We have conducted a series of new experiments and believe that we can now confidently exclude this trivial explanation.
    b. We collected two new datasets to investigate passage-of-time dysphoria's relation to boredom and valence of ongoing thought (used as a standard in the study of mind-wandering). One new cohort of participants received a state boredom scale before and/or after a rest period with subjective momentary mood ratings. Another cohort completed experience sampling surveys before and/or after rest to quantify the emotional valence of their ongoing thoughts.
    c. The new data collections and analyses were preregistered on the Open Science Framework (https://osf.io/gt7a8). We took care to power our samples strongly and to investigate the effects of repeated administration of these state boredom and mind-wandering valence instruments. Participants were randomised to one of the four conditions described below, or to the freely chosen activities group described in point 1, at the time of participation.
    d. By including these groups' questionnaire scores in models of their mood, we learned that passage-of- time dysphoria is only weakly related to the existing constructs of boredom and mind-wandering valence. State boredom after the rest period, change in state boredom over the rest period, and trait boredom each explained a small fraction of the remaining variance in the subject-level mood slopes predicted by the model (F^2<0.07). Similarly, the emotional valence of ongoing thought observed after the rest period, its change over the rest period, and trait-level propensity for mind-wandering each explained a small fraction of the remaining variance (F^2<0.12). These results are described in a new section of the results called "Relationship to Boredom and Mind-Wandering". This result means that most variance in individual mood is not explained by the existing constructs of boredom, mind- wandering, and the emotional valence of ongoing thought.
    e. The new results text reads:
        i. ***Relationship to Boredom and Mind-Wandering***
        ii. *It is possible that existing terms, such as boredom or mind-wandering (MW), could readily explain the phenomenon we describe in this study, making the introduction of the term "passage-of-time dysphori"' redundant. We conducted a series of preregistered (https://osf.io/gt7a8) experiments to investigate this possibility (see Supplementary Materials for the concise results of all preregistered hypotheses). Boredom is typically defined as a state of "low arousal and dissatisfaction"' (Mikulas, 1993). Mind-wandering, often defined as task-unrelated or spontaneous thought (Mrazek, 2013, Christoff, 2016a), tends to be*

unpleasant, particularly because the emotional content of that thought is disproportionately negative (Killingsworth, 2010a, Poerio, 2013). In a preregistered data collection and analysis, we collected four new cohorts totaling n=600 participants to examine the relationship between passage-of-time dysphoria and these more established constructs at the state level, state change level, or trait level. Participants were randomised to one of these cohorts (or to the Activities cohort described in a previous section) at the time of participation.

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

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

v. We used the BoredomAfterOnly cohort to examine our first boredom-related hypothesis: that final state boredom reported after the rest period explains variance in subject-level passage-of-time dysphoria slope. Results showed that a model including state boredom explained additional variance beyond one excluding it (chi^2(2,N=16)=8.769, p=0.0125). But the effect of final state boredom's inclusion on model fit was quite small: the variance explained by the fixed effects in the model increased from R^2=0.341 to R^2=0.359 (F^2=0.0283).

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

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

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

ix. Two new cohorts were collected to quantify the degree to which passage-of-time dysphoria

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

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

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

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

f. We also added a paragraph to the discussion detailing our reasoning for introducing a new term rather than using an existing one. The new text reads:

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

5. Reviewer 1 (paragraph 4) suggested that the construct of mood is vague.
    a. Subjective mood is central to human experience and clinical disorders. Indeed, any measure in affective neuroscience is being measured against subjective reports of mood when assessing their validity. Moreover, subjective mood ratings are the cornerstone of clinical practice and modern neuroscience technology applied via Ecological Momentary Assessment. To underscore this reasoning in the manuscript, we have modified paragraph 10 of the discussion to read as follows.

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

All of the studies are between-subject, probably for reasons of statistical power. The same participant is asked the same question repeatedly over time. The problem with this is that the early questions are likely to cue the participants to the researcher's interest in mood/happiness. But, research has shown that when people pay attention to how happy they are, they tend to become less happy; so may it's not a passage of time effect, but an asking

repeatedly effect. The paper claims to have created a modified version of the task to rule out that the decline in mood is due to "the aversive nature of rating one's mood," but I didn't see such a version. Even if there was such a version, I'd want to know that there weren't other factors, such as an increase in boredom, leading to a decline in reported well-being.

6. Reviewer 1 (paragraphs 4 and 5) suggested that the decline in mood with time may be a result of annoyance with repetitive tasks, or with being repeatedly asked about mood.
    a. Apart from the mood ratings, our tasks were not unusually repetitive, but rather reflect common experimental setups such as the "resting state" that are fundamental to neuroscience. We too worried about annoyance with the frequency of mood ratings and had therefore included a control experiment in our original submission, an experiment in which participants rated their mood at varying frequencies. If mood ratings were aversive, we would expect those with more frequent mood ratings to have more rapid declines in mood. This, however, was not the case. This analysis was described in the supplementary materials of our original manuscript labelled "Passage-of-Time Dysphoria Is Not a Product of Aversive Mood Ratings." To help other readers find it, we have highlighted it more prominently in the main text. This and other methodological artefacts are now examined in a new subsection labelled "Passage-of-Time Dysphoria Is Robust to Methodological Choices," which includes some existing text and some new text. The new text reads:
        i. *The impact of ratings on mood was investigated by systematically varying the frequency of mood ratings. More frequent ratings did not lead to a more rapidly declining mood. The impact of fatigue on mood ratings was investigated by making every mood rating require an equally easy single keypress. This did not change the decline in mood. The results of these and other control experiments and analyses suggested that the observed dysphoria cannot be explained by the above list of factors (see Supplementary section titled "Eliminating Methodological Confounds").*

The paper makes quite a big deal of two results: First, people with clinically depressed affect did not show the effect as strongly. Second, the drop in affect coincided with a decrease in gambling. The problem with both results is that they were not predicted ahead of time; as the paper acknowledges, there are good reasons that both results could have gone one way or the other way. So it isn't clear what one should make of either. Is it possible that there was a floor effect for participants who were clinically depressed?

7. Reviewer 1 (paragraph 6) noted that the finding that increased depression risk was associated with reduced passage-of-time dysphoria was not predicted ahead of time.
    a. Whilst this finding was exploratory in the original cohort, it was supported by the correlation between life happiness ratings and time sensitivity parameter β_T in the mobile app cohort (Fig. 4, left panel). We have now replicated this effect more directly in the new datasets we collected. A linear mixed effects model of mood shows that the combined boredom and mind-wandering cohorts (n=450) exhibited a significant impact of the interaction between depression risk score

and time on mood *(T_{594}=3.64, p=2.93e-4).* New text highlighting this replication has been added to the section titled "Passage-of-Time Dysphoria Is Inversely Related to Depression Risk," where the findings in the original The new text reads:

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

8. Reviewer 1 (paragraph 6) and Reviewer 2 (paragraph 6) noted that the finding that increased depression risk reduced passage-of-time dysphoria could be attributable to floor effects in depressed participants.
    a. We thank the reviewers for this comment. We agree that this is a serious matter and we have done an extensive set of data collection and analysis to address it. We believe that we can confidently say that this is not the case.
    b. The relationship between depression risk and passage-of-time dysphoria was replicated in our new boredom and mind-wandering datasets (see above). But both reviewers pointed out that because depressed participants began with lower initial mood on average, they were more likely to reach the floor of the mood scale. In a sensitivity analysis, we excluded the 27/600 participants in the newly collected dataset who reached the absolute floor of the mood scale (i.e., mood = 0) at any time during the rest period. The effect persisted in this analysis. *(T_{566}=4.06, p=5.65e-5).*
    c. Reviewer 2 noted that many participants are reluctant to reach the floor of the scale but can still reach a

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

    d. During exploration of individuals' mood slopes during rest, we noted that depressed participants appeared more likely to show upwards trends in their mood rating during rest, compared to the rest of the participants. To convey this fact to readers, we took a categorical approach to the question of floor effects. We found that participants at risk of depression (i.e., those with a

clinical depression questionnaire score above a standard clinical threshold) are more likely to exhibit a positive mood slope during the rest period. (chi^2 = 14.57, p = 6.9x10^{-4}). These analyses demonstrate that the anticorrelation between depression risk and passage-of-time dysphoria is not driven by participants reaching their subjective mood floors, since an outsized proportion of depressed participants actually showed increases in mood during the rest period. The longitudinal analyses presented in the original Supplementary Materials suggest that these individual differences in mood slope have moderate stability and are unlikely to be the result of random fluctuations and regression to the mean. This result, combined with the modelling findings linking passage-of-time dysphoria to reward sensitivity, supports the interpretation that depressed individuals are more likely to prefer rest as an alternative to the daily activities whose rewards they undervalue.

e. These findings have been highlighted in the results section titled "Passage-of-Time Dysphoria Is Inversely Related to Depression Risk." This section has been modified to highlight the categorical analysis in which depressed participants are more likely to exhibit positive mood slopes, including two new panels in a modified Figure 3. It also references substantial new text in the supplementary materials that details the analyses described above, in a section titled "Examining Floor Effects in the Depression-Time Interaction." The new text in this section reads:

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

   ii. *…(existing text)...*

   iii. *We took care to rule out the possibility of floor effects driving these results. Individuals reporting greater depressive symptoms on average reported lower initial mood at the onset of the task. If their mood declined further, they therefore had less of the mood scale available to them to express it. This could lead to "floor effects" where the mood of depressed individuals appears to decline more slowly with time simply because they have reached the bottom of the scale and are forced to level out. To address this possibility, we performed sensitivity analyses in which participants reaching either an absolute or individual mood floor were excluded. In both cases, the interaction effect of depression risk and time remained significant (see Supplementary Materials).*

f. We have also added additional text to the discussion outlining our interpretation of these results. (This paragraph also includes a response to Reviewer 2's concern about non-gaussianity, which is similarly related to the bounded mood scale.)

   i. *Our use of a bounded mood scale has consequences. First, we must consider the possibility that our depression-related findings were driven by floor effects. The effect persisted in categorical analyses (an outsized proportion of depressed participants showed positive mood slopes) and*

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

In the Mturk study, it concerned me that a fairly large fraction of participants' data seems to have been lost – those who closed the task window before being asked to do so. Were they different in some way from those who stuck it out (admittedly, this selection effect would probably bias the results in the opposite of the mood dysphoria effect).

9. Reviewer 1 (paragraph 7) asserted that a large proportion of MTurk participants' data were lost, raising concerns

of selection bias.

   a. Just under 5% of the original MTurk participants had data that did not save due to technical difficulties or or the participant closing the task window before being asked to do so. This fraction is not markedly different from other studies. While it is possible that this introduces some selection bias, we are unable to control it and do not believe that this is an unusually large risk in our study. As the reviewer suggests, any selection bias resulting from the dropout of dissatisfied participants is probably more likely to reduce the group-level passage-of-time dysphoria effect than it is to inflate it.

The paper, which follows participants up in a set of researcher-selected studies in which they were asked some kind of mood-related question repeatedly over time, raises a broad methodological issue. How should we sample the kinds of situations in which participants are waiting, to feel confident in reaching the broad conclusion that there is a Passage-of- Time Dysphoria effect. It seems to me that a clever experimenter who was motivated to demonstrate a Passage-of-Time Euphoria effect could probably devise experiments that would obtain such a result – e.g., by making an experiment really interesting and engaging. So, what are we to make of the fact that the current set of experimenters found increases in dysphoria over time? My bottom line is that I'm simply not convinced that this is an important, generalizable, result.

10. Reviewer 1 (paragraph 8) suggests that the effect lacks generalizability, in part because a motivated researcher could have designed a task to prove a passage-of-time euphoria effect.
    a. The question of how to design a task that could reliably produce increases in subjective mood is an interesting one, and one that could have clinical importance as a potential depression treatment. Past and present research from our group suggests that positive mood induction is

more challenging than negative mood induction. In a gambling game, it required greater positive reward prediction errors (RPEs) to raise mood than it did negative RPEs to lower it (Keren, 2021). Findings from one of our new datasets supports this negative bias: even when participants could select any activity they wanted, their mood did not improve significantly. Conversely, mood declined as people played an engaging mobile app, and the often-presumed-neutral act of rest caused marked dysphoria.

b. Exploring the mood impact of the full space of possible tasks and situations, of course, is not a tractable problem. We have chosen to focus our attention on a class of paradigms that is extremely common in neuroscience: long, neutral, low-stimulation tasks. The presence of passage-of-time dysphoria in the rest and simple tasks we have administered suggests that many psychological studies are at risk of previously unknown confounds. Most researchers would see these qualities as unobjectionable or even desirable for brain or behaviour studies, even those with clear relationships to mood. We hope that the results of our study will lead researchers to reexamine this idea in their own research.

c. We have added text to the discussion that outlines this reasoning:

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

Reviewer #2:

Remarks to the Author:

Jangraw and colleagues present an interesting analysis of a large cohort of multiple studies and demonstrate that mood ratings decrease as a function of time on task. They coin a term, 'Passage-of-Time Dysphoria' for this phenomenon. They provide evidence that the rate of decline is shallower for individuals with higher depression levels and that there was a link between a reward sensitivity parameter and the slope of the mood decline as a function of time on task. While a highly stable effect with some nice controls (in particular varying the frequency of mood ratings), the authors briefly note in the discussion this finding could be considered or accounted for by many things, previously described and demonstrated in the literature, including mind-wandering or boredom.

I agree with the overall statement that affective scientists should be aware of this effect, insofar that, as psychologists, most of our tasks are incredibly boring and do not simulate the real-world at all. However, apart from a narrow correlational analysis examining choices to gamble after rest-blocks, the authors do not present any mechanistic data to show how this effect can be pushed around to help the reader understand what is causing the

phenomenon (and whether or not it is boredom / mind-wandering / task agency). I would find this submission more compelling for this outlet were there to be experimental exploration of these effects.

11. We thank the reviewer for their encouraging comments and constructive suggestions. Please see our third response to Reviewer 1 above for a discussion of the impacts of boredom, mind-wandering, and the valence of ongoing thought on this effect. For a discussion of task agency, see our first response to Reviewer 1 above.

I also have questions in terms of generalizability. For instance, I would also find this piece more compelling if other real- world scenarios yielded similar effects. What about so-called flow experiences? Or do gamers playing Warcraft for hours on end experience similar decrements in mood? I'm not suggesting that the authors *have* to do a study in this domain, but some consideration or empirical data speaking to other arenas and the relevance (or irrelevance) of this finding to those domains would be helpful.

A few additional comments below:

The authors note that the effects for the mobile task were significantly weaker than the other 'computer-based' tasks. What is different between the mobile app and being in a lab (or on MTurk)? One hypothesis is that you can simply exit the task more easily when playing the game on your phone. How long did people tend to play via their phone? That curve is not displayed in Fig 1 (only the slope parameter in Fig 2), and not displayed in the supplement. Does agency / ease of ability to exit impact this effect? Can the authors rule this out or address this?

12. We have attempted to address the questions about real-world activities and task agency with a targeted experiment, outlined in point 1 above. Participants fully free to choose their own activity did not exhibit passage- of-time dysphoria. But the fact that a mood decline was observed in the mobile app task, where participants received no monetary reward and could easily exit at any time, suggests that task agency alone is not enough to neutralise the dysphoria. Regarding the specific question about the duration of the mobile app game, we have calculated that the median participant played the game for 303 seconds. This information has been added to the methods section.

At the outset of the discussion the authors write, "Our findings show that subjects incorporate information about the passage of time into their ratings of mood." How do authors know that it is this specifically and not simply boredom or some other process as they allude to later? Some of the stronger claims of this paper could be walked back.

13. We have now used targeted experiments to demonstrate that boredom and the valence of ongoing thought explain only a small portion of the observed changes in mood; we have added new text to the results, discussion, methods, and supplement detailing these analyses (see point 4 above).

Could passage-of-time dysphoria x depression effects be due to floor effects? I recognize that individuals 'could' rate themselves below 0.3 , 0.4, but it is rare to have individuals rate themselves at an absolutely horrible mood

while in psychology studies and, given that depression was linked to lower initial mood ratings, a plausible hypothesis is simply that they had lower to drop and therefore the slope was more shallow.

14. We have performed a number of new analyses to address this important concern about floor effects (see our fifth response to Reviewer 1 above).

This may not matter for model parameters (particularly given the cellphone subsample using large sample size), but M0 and other parameters from the computational model are not normally distributed. BT is highly kurtotic. Moreover, it is clear that individuals rated their mood in a nonnormal fashion (some ratings at 0, some at 1 and a lot right in the middle at .5). Using these types of data, I've found superior fitting models using cumulative families than assuming the outcome space is gaussian. Further, because the outcome space is bounded (0,1) a gaussian model will yield 'posterior' predictions well outside the possible response space. Just something to consider.

15. Reviewer 2 (Paragraph 7) points out that the computational model's parameters often violate assumptions of

Gaussianity, and that other models may provide a better fit to the data.

a. We believe that the reviewer's suggestion is a good one, but one that should be beyond the scope of the current project. Because we are modelling subjective mood ratings bounded between 0 and 1, the errors are indeed not Gaussian: Gaussian errors would lead to the mood ratings being unbounded. The common i.i.d. Gaussianity assumption on error terms is often a poor fit to behavioural data, which may make non-Gaussian distributions (such as logistic-transformed Gaussian) a better fit to the errors in the data. We did our best to mitigate the effect of this mismatch by capping the model predictions to the allowable range. We also initialised many parameters to non-normal distributions and restricted several parameters to feasible ranges on every iteration.

b. We cannot rule out the possibility that using a model with a non-Gaussian error term would improve performance. However, implementing such a model would lead to substantial novel methodology. A new error term distribution could require non-linear parameter fitting that would exclude existing GLM models and require a new approach. Because very little is known about the true error distribution, exploring the space of possible error distributions would require substantial effort that should be tested, presented, and evaluated in a targeted paper.

c. In the absence of a well-established alternative model, we believe that the models currently in the manuscript reflect the current state of the art and provide a reasonable fit to the data.

d. We have added a section to the limitations portion of the discussion that addresses the consequences of our use of a bounded mood scale. It reads:

   i. *Our use of a bounded mood scale has consequences. First, we must consider the possibility that our depression-related findings were driven by floor effects. The effect persisted in categorical analyses (an outsized proportion of depressed participants showed positive mood slopes) and after excluding participants who reached an absolute or individual mood floor. Second, the bounded mood scale prevents the error term of our mood models from being truly Gaussian. We chose to maintain the Gaussian assumption because it is well established in existing models,*

but it is likely that a different assumption would better fit the data. Because very little is known about the true error distribution, exploring alternative models is beyond the scope of this study. We attempted to mitigate the effect of any mismatch by capping the model predictions to the allowable range. We also initialised many parameters to non-normal distributions and restricted several parameters to feasible ranges on every iteration.

For Fig 1, dotted line at mood = 0.5 is sort of meaningless because, while it is the middle of the scale (and assumed by authors to be a 'neutral mood'), it isn't the mean of subject's mood ratings. It's more likely that a 'neutral' mood is actually the mean across these ratings. As most individuals generally display a positivity bias in affect.

16. Reviewer 2 (Paragraph 8) noted that the dotted line at 0.5 is not really a "neutral" mood because it is not the mean of subjects' ratings.
    a. We thank the reviewer for this observation. We have changed the dotted horizontal lines in Figures 1, 3A, and S3 to reflect the mean across initial mood ratings. We hope that this provides a more data- driven and meaningful reference point, one that will help readers appreciate the size of the decline in mood at each time point.

Decision Letter, first revision:

8th August 2022

Dear Dr Jangraw,

Thank you once again for your manuscript, entitled "Passage-of-Time Dysphoria: A Highly Replicable Decline in Mood During Rest and Simple Tasks that is Moderated by Depression," and for your patience during the lengthier than usual peer review process.

Your manuscript has now been evaluated by 3 reviewers, whose comments are included at the end of this letter. Reviewer 2 has seen your manuscript before, while Reviewers 3 and 4 are new in this round of review. Original Reviewer 1 was unable to provide their comments this time.

Although the reviewers find your work to be of interest, they also raise some important concerns. We remain interested in the possibility of publishing your study in Nature Human Behaviour, but would like to consider your response to these concerns in the form of a revised manuscript before we make a decision on publication.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to (1) identifying key priorities that should be addressed in revision and (2) overruling referee requests that are deemed beyond the scope of the current study. We hope that you will find the prioritised set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

1) The phenomenon you have identified should be described more precisely and accurately, given that the term 'dysphoria' does not appear to be suitable (as Reviewer 3 points out) and the effect is very specific (Reviewers 2 and 3).

2) Reviewer 4 raises important concerns regarding the definition and measurement of boredom and the extent to which it is ruled out as an alternative explanation. You will need to address this thoroughly - the limitations of the measure of boredom need to be entirely clear and alternative explanations have to be presented.

3) Full statistics, including effect sizes and confidence intervals, should accompany the reporting of all results.

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

In sum, we invite you to revise your manuscript taking into account all reviewer and editor comments. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

We hope to receive your revised manuscript within two months. I would be grateful if you could contact us as soon as possible if you foresee difficulties with meeting this target resubmission date.

With your revision, please:

• Include a "Response to the editors and reviewers" document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. When formatting this document, please respond to each reviewer comment individually, including the full text of the reviewer comment verbatim followed by your response to the individual point. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.

• Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your

work. Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Sincerely,

Arunas Radzvilavicius, PhD
Editor
Nature Human Behaviour

Reviewer expertise:

Reviewer #2: original reviewer, affective neuroscience

Reviewer #3: new reviewer, cognitive psychology, mind wandering and boredom

Reviewer #4: new reviewer, social psychology, boredom

REVIEWER COMMENTS:

Reviewer #2:
Remarks to the Author:

The authors have done significant work to this submission. I appreciate their hard work on this.

I think the title should change as the authors now demonstrate that the "Passage-of-Time Dysphoria" is highly contextually and experimentally specific

In both their response letter and in the submission, the authors make a big deal about resting state data and the issues of the passage of time-dysphoria on resting state data. But isn't this suggestion contradicted by evidence that longer resting state scans (e.g., 25 minutes +) lead to more stable connectivity estimates? How do the authors make sense of these data with how they are characterizing the implication of these effects?

Similarly, is it really that during *all kinds of rest* that mood declines? This also seems, on the face of it totally counterintuitive, if not impossible. For instance, I can rest / relax on a beach somewhere for hours on end and my mood doesn't decline during that vacation (at least I don't think it does!). So, I would say that indeed it does seem like in simple tasks and / or where one's behavior is externally constrained when they're not doing anything that mood declines. I suppose the authors are providing evidence that this might also apply to the jargon term of 'rest' from the functional MRI studies. But that distinction is not really clear in the title or in the paragraphs describing the 'rest' findings. My takeaway of the main effects from the work by the researchers in this paper is that when you constrain a human's behavior and action to boring and menial tasks, their mood declines over time.

"This group was the first to not exhibit passage-of-time dysphoria" I would clarify that this was the first sample presented in this study to not exhibit the effect.

Reviewer #3:
Remarks to the Author:

Overall, this paper is very interesting. The findings may be important for many different fields and I think the paper should eventually be published either in NHB or elsewhere. At the same time, I think the paper should like undergo major revisions before publication to address concerns about the theoretical explanations, methodological choices, and general claims being made about dysphoria. I offer a set of critiques and concerns that I hope will be helpful for a revision.

1.
Perhaps my biggest concern of the current study is the interpretation of the general finding, which can be seen starting with the title. Dysphoria and negative mood have a specific meaning — namely, they are inherently and by definition negative. From the key figures in the paper (Fig 1: A, B, C) it is clear that the "negative" feelings over time are not actually negative. In A, 5 out of 6 lines never dip below .5 (the midpoint of the mood scale), and the single line that does is still above .45 — at best, this is neutral. Similar trends can be seen in B and C as well.

In light of this, I think the authors should strongly reconsider their claims and choice of wording. The implications of using such an emotionally-laden term to coin this phenomenon will have real consequences and I think this should be considered thoughtfully in a revision. Indeed, the drop in mood is extremely interesting and worthwhile publishing, but perhaps not framed as dysphoria without more evidence people actually felt negative — not simply less positive than when they started. Note that similar arguments have been made in the mind wandering and boredom literature as well (see, e.g., Fox, Andrews-Hanna, et al., 2018; Fox, Thompson, et al., 2014, etc.) as well. I do not think this can be dealt with by simply adding a caveat in the text, but rather should be deeply considered in the phrasing throughout.

2.
One of the other major suggestions is to strengthen the Intro, especially in terms of theory. The main argument seems to be that time-invariance is a bad assumption. I agree with this intuitively, but why would we expect to see more negative mood over time? I appreciate the authors admit it was initially a serendipitous finding — this is clearly a very interesting observation. At the same time, I still think there is likely enough literature to draw on to attempt to explain why time matters for mood theoretically.

Related, authors may want to make clear before introducing the two paradigms on Page 2 that the idea is time invariance regardless of rest vs task. On the surface, it seems a bit difficult to think that time itself is importance given that resting states and demanding tasks are quite different — more explanation for how each may display the same pattern would help readers.

The authors then suggest some possible reasons that we should expect to see that affective changes

vary systematically. Literature on negative mind wandering is then cited. First, I do not think this literature has been reviewed very thoroughly. Second, a few paragraphs later the authors then say that they findings are not attributable to mind wandering which seems counterintuitive.

3.
This comment is somewhat related to #2. The results detail a large set of analyses, all of which are very impressive. One issue, however, is that some (though not all) do not seem particularly well motivated. They are included perhaps because they offer some caveat to the phenomenon. For example, for the impact on behaviour section - why was this the behaviour chosen to assess given the many possible behaviors that may have been analyzed?

4.
There are also some incorrect and confusing claims made in the results section. Here are just a few examples (not an exhaustive list):
1) mind wandering is not typically unpleasant and the content is not disproportionately negative (Killingsworth is cited for this claim, but I don't think this is even true in this paper);
2) it seems like the relationship with boredom is downplayed — i.e., "explaining a small amount of variance" — what was the standardized slope here? An R2 above .3 is non-trivial for most psychological phenomenon. To this end, was this the R2 for the slope models reported elsewhere in the paper?
3)why did the authors used the Turnbull MDES method if they were interested in mind wandering specifically? The PCA approach with the emotion component is not actually representative of the MW-related phenomenon that is discussed elsewhere throughout the paper (see e.g, how it is used in Turnbull et al., which is not the same as traditional MW). I did not find this analyses meaningful in terms of answering the research question related to MW, as emotion was the primary analyses. As well, this PCA approach is not described in detail in the methods section.

These and other examples make the results section a bit difficult to follow in terms of impact, particularly because the methodological choices are not clear even in the results section.

5.
Is it more accurate to describe this phenomenon as "passage of time DURING TASKS"? I think it's important to note that resting states are still very much a "task" compared to daily life. Indeed, in many EMA studies assessing mood, the mood ratings to do not change systematically with short time gaps, so it seems to be specific to having a predefined task.

Minor:

In general, many of the citations used did not seem like the best examples for the claims being made. I suggest the authors more carefully examine who they are citing for what claim — i.e., is it the original source.

Please also add a citation for the reward valuation claim. Related: the authors go in to a bit of theorizing in the discussion about the depression results, but I think more can be developed on this by incorporating more literature.

I understand space limitations and journal specific formatting. However, I found the amount of results that were simply state without much statistical reference a bit difficult to follow and evaluate. It would be helpful if more results and statistics were included in the main paper rather than an overview claim. However, I defer to the journal Editors on this particular comment.

The abstract was full of results which is nice in some ways, but many of them were difficult to understand without being contextualized.

Please provide descriptive statistics for the key variables in the supplementary.

Reviewer #4:
Remarks to the Author:

Across a range of samples and tasks, people reported lower positive affect/higher negative affect when asked to report mood across multiple occasions, leading to what the authors call "passage-of-time dysphoria". This appears to be an important and highly replicable methodological artifact that will be of interest to a great many researchers, and has substantial implications for the interpretation of common pre-post designs (including those involving random assignment to condition). The impact of the present work thus seems sizeable, and the authors have done a thorough job of addressing earlier reviewer concerns. The comments here are intended in the spirit of strengthening what is already a sizeable contribution.

Although the paper is framed in terms of affective neuroscience, the implications for the field of affective science more generally as well as many subfields of psychology (e.g, well-being, positive psychology) would also seem substantial. Especially given the generalist nature of an outlet like Nature Human Behavior, reframing portions of the paper to be less discipline-specific would likely broaden its potential impact.

In doing so, the authors may wish to draw on existing work in justifying their use of mood as the primary construct of interest. Mood is a well-defined construct in affective science, and one that can be distinguished from other affective components (such as discrete emotions, or dimensions of affect such as valence and arousal); making that distinction explicit would be helpful in clarifying the contribution of the present work. For instance, it may be more useful to speak of declines in positive affect/increases in negative affect over time.

Likewise, self-report is the gold standard for the measurement of emotion (and mood), and the key criterion variable against which all physiological and neural "markers" are benchmarked. Furthermore, the usefulness of such affective markers is itself dubious, given evidence that discrete emotions cannot be distinguished by their physiological signatures (Siegel et al., 2018), facial expressions (Gendron et al., 2015; Barrett et al., 2019), or neural activity (Lindquist et al., 2015).

For this reason, the definition and measurement of boredom on page 10 is also however inappropriate. While boredom is defined here as "a state of low arousal and dissatisfaction," more recent evidence shows that this definition is not empirically supported (see e.g., Westgate & Wilson, 2018), and that boredom is associated with both high arousal states as often as with low arousal, or mixed states. For

this reason, the MSBS is not an appropriate measure of boredom and, indeed, manipulations that equally increase boredom (when measured by direct self-report) yield differing patterns of responses on the various MSBS subscales (e.g., dysphoria, agitated affect, inattention, disengagement, time distortions). This, unfortunately, makes the studies reported on pp. 10-12 difficult to interpret and does not rule out boredom as a potential mechanism to explains these results. A better measure of boredom would be face-valid items asking participants to directly report how bored they feel (see e.g., Barrett, 2004; Diener, 2000; Nisbett & Wilson, 1977; Robinson & Clore, 2002; Westgate & Steidle, 2021). This is particularly important in light of work showing that perceived autonomy can also potentially reduce boredom (see work on constraint and boredom by Thackray et al and Struk et al), and that people readily feel bored when asked to sit and "do nothing" (e.g., Wilson et al., 2014), which strongly suggests that boredom may be one contributing mechanism.

In sum, the authors have provided a substantial body of evidence for the effect encompassing a wide range of designs, datasets, and participant demographics, which would tend to point heavily in favor of generalizability. I particularly liked the experiments ruling out other potential mechanisms (e.g., regression to the mean), many of which were run in response to earlier reviewer comments and are detailed in the supplemental materials, and found the majority of them highly persuasive. No doubt, like most phenomenon, mood decline will also be subject to moderating effects. Unpacking what those circumstances and theoretical moderators are, however, seems more an avenue for future work than key questions necessary to resolve in this initial paper. I have no doubt that such future research will be fruitful and informative.

Author Rebuttal, first revision:

==========

REVIEWER COMMENTS:
Reviewer #2:
Remarks to the Author:

1. REVIEWER 2: The authors have done significant work to this submission. I appreciate their hard work on this.

   I think the title should change as the authors now demonstrate that the "Passage-of-Time Dysphoria" is highly contextually and experimentally specific.

   a. RESPONSE: Thank you for this suggestion, which was shared by Reviewer 3. We have removed the term from the title.
      i. The title now reads:
      "A Highly Replicable Decline in Mood During Rest and Simple Tasks"

b. We have now adopted the term “Mood Drift Over Time” (or “mood drift” for short) to describe the phenomenon. We believe that this term has several strengths: (a) it directly references mood, making it more precise than “dysphoria;” (b) the word “drift” concisely implies temporality; and (c) it is not restricted to negative emotional valence or negative mood slopes, thus capturing both group and individual results. This new term has been adopted throughout the paper.
    i. In the introduction (paragraph 7), we introduce the term Mood Drift Over Time by saying:
    “*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).”*

2. REVIEWER 2: In both their response letter and in the submission, the authors make a big deal about resting state data and the issues of the passage of time-dysphoria on resting state data. But isn’t this suggestion contradicted by evidence that longer resting state scans (e.g., 25 minutes +) lead to more stable connectivity estimates? How do the authors make sense of these data with how they are characterizing the implication of these effects?
    a. RESPONSE: We thank the reviewer for asking us to clarify this important point. We do not have evidence that mood drift contributes to variability and low reliability in resting state fMRI (rs-fMRI), and we do not want to imply this. We do use resting-state fMRI in the discussion as an example of how, if mood drift could lead to variability during rest, that acts as a mood-induction manipulation that could result in problematic confounds in the experiment that follows it. At the end of this example, we have added a more explicit caveat that we cannot conclude that these effects are causing low reproducibility in rs-fMRI.
        i. In the discussion, the end of paragraph 5 now reads:
        “(We do not imply that mood drift lowers reliability in resting-state MRI (Birn, 2013; Noble, 2017; Noble, 2019); we simply point out its role as a potential confound when drawing inferences about mood and brain states during/after rest.)”

    b. If mood drift did strongly affect rs-fMRI, the reviewer notes a perceived contradiction with the finding that longer scans give more stable estimates. Given that the mood scale is finite, a nonzero mood drift effect cannot be linear forever. At some point, the group-level mood will presumably either saturate or level off (or reverse direction, which seems unlikely). This levelling off could result in mood estimates being more stable at long scan durations (such as the 24 minutes suggested by

(Noble, 2017)), just as resting-state connectivity is.

Since readers may share this concern about extended task durations, we have added text to the limitations section of the manuscript.

i. In the discussion, paragraph 8 now reads:
"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 time scale of most experiments."

3. REVIEWER 2: Similarly, is it really that during *all kinds of rest* that mood declines? This also seems, on the face of it totally counterintuitive, if not impossible. For instance, I can rest / relax on a beach somewhere for hours on end and my mood doesn't decline during that vacation (at least I don't think it does!). So, I would say that indeed it does seem like in simple tasks and / or where one's behavior is externally constrained when they're not doing anything that mood declines. I suppose the authors are providing evidence that this might also apply to the jargon term of 'rest' from the functional MRI studies. But that distinction is not really clear in the title or in the paragraphs describing the 'rest' findings. My takeaway of the main effects from the work by the researchers in this paper is that when you constrain a human's behavior and action to boring and menial tasks, their mood declines over time.
    a. RESPONSE: We agree that not all types of rest induce a mood decline. This intuition is supported by the finding that people able to choose their own activity did not experience a decline in mood on average (see results section "Mood Drift Over Time Is Not Present in Freely Chosen Activities"), and by the finding that some rest and simple-task participants experienced near-zero or positive mood drifts (Supplementary Fig. S5C).

    We have now highlighted the phenomenon's apparent specificity to experimental tasks and rest in several places in the text. In the introduction, we now frame the behavioural constraint of a psychological task or rest period as a negative affective stimulus that could gradually draw mood downward, and we mention that the effect was not observed when participants chose their own activities. In the discussion, we reiterate this specificity and give a more prominent position to the results of the freely chosen activities experiment.

i. In the introduction, paragraph 5 now reads:
"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 affect (Eastwood, 2012). We might therefore conceive of a psychological task's behavioural constraint as a sort of negative affective stimulus that could gradually draw mood downward."

ii. The final paragraph of the introduction, sentences 3-4, now reads:
"This downward mood drift was replicated in 19 large and varied cohorts, totaling 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."

iii. In the discussion, paragraph 1 now begins:
"In this study, we describe the discovery of a highly replicable and relatively large effect which we call Mood Drift Over Time: the average participant's mood gradually declined with time as they completed simple tasks or rest periods."

iv. In the discussion, paragraph 4 now reads:
"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."

b. We thank the reviewer for their helpful summary of the manuscript's findings. While we cannot be certain that constraints on behaviour are the primary driver of mood drift, the freely-chosen-activities experiment provides some evidence in favour of this explanation. To frame this idea in readers' minds early in the paper, we have now referenced this finding in the introduction.

i. In the introduction, paragraph 5 now reads:
"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 affect (Eastwood, 2012). We might therefore conceive of a psychological task's behavioural constraint as a sort of negative affective stimulus that might gradually draw mood downward."

4. REVIEWER 2: "This group was the first to not exhibit passage-of-time dysphoria" I would clarify that this was the first sample presented in this study to not exhibit the effect.
    a. RESPONSE: Thank you for pointing this out. We have clarified this language.
        i. The results section titled "Freely Chosen Activities", paragraph 2, now begins: *"This group was the first sample investigated in this study that did not exhibit mood drift."*

Reviewer #3:

Remarks to the Author:

1. REVIEWER 3: Overall, this paper is very interesting. The findings may be important for many different fields and I think the paper should eventually be published either in NHB or elsewhere. At the same time, I think the paper should like undergo major revisions before publication to address concerns about the theoretical explanations, methodological choices, and general claims being made about dysphoria. I offer a set of critiques and concerns that I hope will be helpful for a revision.

    Perhaps my biggest concern of the current study is the interpretation of the general finding, which can be seen starting with the title. Dysphoria and negative mood have a specific meaning — namely, they are inherently and by definition negative. From the key figures in the paper (Fig 1: A, B, C) it is clear that the "negative" feelings over time are not actually negative. In A, 5 out of 6 lines never dip below .5 (the midpoint of the mood scale), and the single line that does is still above .45 — at best, this is neutral. Similar trends can be seen in B and C as well.

    In light of this, I think the authors should strongly reconsider their claims and choice of wording. The implications of using such an emotionally-laden term to coin this phenomenon will have real consequences and I think this should be considered thoughtfully in a revision. Indeed, the drop in mood is extremely interesting and worthwhile publishing, but perhaps not framed as dysphoria without more evidence people actually felt negative — not simply less positive than when they started. Note that similar arguments have been made in the mind wandering and boredom literature as well (see, e.g., Fox, Andrews-Hanna, et al., 2018;

Fox, Thompson, et al., 2014, etc.) as well. I do not think this can be dealt with by simply adding a caveat in the text, but rather should be deeply considered in the phrasing throughout.

a. RESPONSE: Thank you for this thoughtful and thorough critique of the term used to describe our main finding. This concern was shared by Reviewer 2. Our manuscript had been using language that conflated negative mood *slopes* with negative *mood*. We have now adopted the term "Mood Drift Over Time" ("mood drift" for short). We believe that this term has several strengths: (a) it directly references mood, making it more precise than "dysphoria;" (b) the word "drift" concisely implies temporality; and (c) it is not restricted to negative emotional valence or negative mood slopes, thus capturing both group and individual results. This new term has been adopted throughout the paper, including the title.
    i. In the introduction (paragraph 7), we introduce the term by saying: *"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)."*
    ii. Throughout the text, we have replaced references to "dysphoria" with references to "mood drift" or "decline in mood".

2. REVIEWER 3: One of the other major suggestions is to strengthen the Intro, especially in terms of theory. The main argument seems to be that time-invariance is a bad assumption. I agree with this intuitively, but why would we expect to see more negative mood over time? I appreciate the authors admit it was initially a serendipitous finding — this is clearly a very interesting observation. At the same time, I still think there is likely enough literature to draw on to attempt to explain why time matters for mood theoretically.

   Related, authors may want to make clear before introducing the two paradigms on Page 2 that the idea is time invariance regardless of rest vs task. On the surface, it seems a bit difficult to think that time itself is importance given that resting states and demanding tasks are quite different — more explanation for how each may display the same pattern would help readers.

   The authors then suggest some possible reasons that we should expect to see that affective changes vary systematically. Literature on negative mind wandering is then cited. First, I do not think this literature has been reviewed very thoroughly. Second, a few paragraphs later the authors then say that the findings are not attributable to mind wandering which seems counterintuitive.

a. RESPONSE: We share your goal to place this manuscript's findings in the context of existing theory. Thank you for lending your knowledge of the mind-wandering literature to this effort. We have attempted to incorporate and cite this literature more thoroughly in the current draft. We have now strengthened the introduction in a number of ways. First, we have expanded our description of current theory to connect the following important phenomena that we think may be interconnected: affective chronometry, anhedonia, delay aversion, opportunity cost, and explore/exploit tradeoffs. Second, we have used the idea of constrained behaviour (present in both psychological tasks and experimenter-mandated rest periods) to connect this theoretical description to violations of the methodological assumptions described in the introduction. Finally, the introduction now references the research cited by the reviewer (Fox, Andrews-Hanna, et al., 2018; Fox, Thompson, et al., 2014) when discussing existing mind-wandering literature.

   i. The introduction, paragraphs 3-6 now read:
   "Whilst 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 (Frijda, 1991; Scherer, 1994; Davidson, 1998; Davidson, 2015), and that individuals vary in the rates at which positive or negative affect decays after an event (Gilboa, 1994; Hemenover, 2003). Such individual differences may be linked to mental health. For instance, psychopathologists theorise that anhedonia, a symptom of both depression and schizophrenia, arises from a failure to sustain reward responses for a normative period of time (Kring, 2014). And studies of ADHD suggest that hyperactivity's impulsive behaviour results from delay aversion, the idea that a delay is itself unpleasant and impulsivity is simply a rational choice to avoid it (Sonuga-Barke, 1992; Solanto, 2001; Sonuga-Barke, 2016).

   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) (McRae, 1970; Hoskin, 1983; Palmer, 1999). This idea is fundamental to the explore/exploit question that has recently preoccupied neuroscientists (Cohen, 2007; Constantino, 2015; Addicott, 2017). 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 (Geana, 2016; Agrawal, 2020).

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 affect (Eastwood, 2012). 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 (Robison, 2020; Killingsworth, 2010; Fox, 2014; Fox, 2018). 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 (Van Hooff, 2014; Miner, 2010). 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. But the size, stability, and clinical correlates of this variability remain unexplored."

b. Thank you for pointing out the introduction's seemingly conflicting statements about boredom and mind-wandering. We believe that many readers will intuitively hypothesise that mood drift is simply boredom or negatively valenced mind-wandering viewed through the lens of mood, and we wish to use the introduction to encourage readers that we share this concern and will address it in our analyses. We have modified the introduction to be more precise about the extent to which our results support this hypothesis.
    i. The introduction's final paragraph, sentence 5, now reads:
    "We show that mood drift is related to, but not a trivial extension of, the existing constructs of boredom and mind-wandering."

3. REVIEWER 3: This comment is somewhat related to #2. The results detail a large set of analyses, all of which are very impressive. One issue, however, is that some (though not all) do not seem particularly well motivated. They are included perhaps because they offer some caveat to the phenomenon. For example, for the impact on behaviour section - why was this the behaviour chosen to assess given the many possible behaviors that may have been analyzed?
    a. RESPONSE: The reviewer is correct that caveats motivated some follow-up experiments. Many were included to confirm or deny confounded and trivial explanations for mood drift, such as mood ratings' aversiveness or regression to the mean. We have attempted to give readers a greater sense of the motivation for these analyses by adding an introductory sentence to any sections that did not yet have them.
        i. The results section titled "Mood Drift Over Time Is Sizeable" now begins: *"Our first objective was to estimate the size of the effect."*
        ii. The results section titled "Inter-Individual Differences" now begins: *"Having characterised 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."*
        iii. The results section titled "Mood Drift Over Time Is Associated with Sensitivity to Rewards" now begins:
        "Mood is central to depression, which is thought to relate etiologically to reward responsiveness (Pizzagalli, 2008; Halahakoon, 2020). 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."
        iv. The results section titled "Participants Receiving Rest Periods Are Less Likely to Gamble" now begins:
        "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 (Isen, 1983; Arkes, 1988), music (Schulreich, 2014), and feedback (Vinckier, 2018). We asked whether mood drift would influence this behaviour in a similar way."

b. The behaviour analysed was convenient, since it was part of our probabilistic reward task when the original serendipitous finding was made. However, this task is a standard one commonly used to examine mood (Rutledge, 2014; Camille, 2004; Eldar2016; Vinckier2018), and one that our group has written about substantially (Keren, 2021; Liuzzi, 2022). 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 (Isen, 1983; Arkes, 1988), music (Schulreich, 2014), and feedback (Vinckier, 2018). Given our finding that periods of low stimulation are associated with downward drifts in mood, we asked whether this affective manipulation would influence the same behaviour in a similar way.

   To make this rationale clearer to readers, we have added text to the results sections on behaviour.

   i. The results section titled "Characterising the Effect", paragraph 1 has been modified. Beginning at sentence 7, it now reads:
   "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 (Rutledge, 2014; Camille, 2004; Eldar, 2016; Vinckier, 2018). It was included to observe the effects of rest on rational behaviour, to maintain links with previous studies of mood and reward (Rutledge, 2014; Keren, 2021; Liuzzi, 2022), and to enable related analyses on a large cohort of participants (n=26,896) playing a similar game on their smartphones (Bedder, 2020)."

   ii. The results section titled "Participants Receiving Rest Periods Are Less Likely to Gamble" now begins:
   "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 (Isen, 1983; Arkes, 1988), music (Schulreich, 2014), and feedback (Vinckier, 2018). We asked whether mood drift would influence this behaviour in a similar way."

4. REVIEWER 3: There are also some incorrect and confusing claims made in the results section. Here are just a few examples (not an exhaustive list):
   1) mind wandering is not typically unpleasant and the content is not disproportionately

negative (Killingsworth is cited for this claim, but I don't think this is even true in this paper); 2) it seems like the relationship with boredom is downplayed — i.e., "explaining a small amount of variance" — what was the standardized slope here? An R2 above .3 is non-trivial for most psychological phenomenon. To this end, was this the R2 for the slope models reported elsewhere in the paper?
3) why did the authors used the Turnbull MDES method if they were interested in mind wandering specifically? The PCA approach with the emotion component is not actually representative of the MW-related phenomenon that is discussed elsewhere throughout the paper (see e.g, how it is used in Turnbull et al., which is not the same as traditional MW). I did not find this analyses meaningful in terms of answering the research question related to MW, as emotion was the primary analyses. As well, this PCA approach is not described in detail in the methods section.

These and other examples make the results section a bit difficult to follow in terms of impact, particularly because the methodological choices are not clear even in the results section.

a. RESPONSE: Thank you for identifying these important elements of the MW and boredom analyses as opportunities for improvement. We have done our best to improve our analyses and clarify the statements identified above. We will address them in order in the items below.
b. (1) the reviewer is correct that (Killingsworth, 2010) did not identify mind-wandering as unpleasant in that it was associated with mood reports that were, on average, above 0.5. However, the average MW episode was associated with lower mood ratings than other activities where the participant was "in the moment." Later research (Poerio, 2013) clarified that only affectively negative mind-wandering tended to have a dampening effect on mood 15 minutes later. This motivated our decision in the previous draft to focus on the emotional content of MW rather than its presence alone (which we have since reconsidered, see response 4.d below). We have removed the statement identified by the reviewer from the results section. A more precise and correct description of the motivation for our preregistered MW-related analyses is provided in the supplement.
    i. The supplementary section titled "Results of Preregistration on Boredom, Mind-Wandering, and Freely Chosen Activities", paragraph 4, now reads: *"Past research has found that it is not mind-wandering in the general or "traditional" sense (i.e., any task-unrelated thought) that decreases mood, it is mind-wandering with negative affective content (Poerio, 2013). This notion is*

supported by current theories of mind-wandering not as a monolith, but as a collection of thoughts whose content shapes brain activity and behaviour (Smallwood, 2021). Research has linked thought probe responses about the affective content of this ongoing thought to brain activity patterns in the mOFC (Tusche, 2014). The method described in (Turnbull, 2019) provides a way to quantify the negative affective content of this ongoing thought that more robustly separates affective tone from the mere presence of task-unrelated thought (see Methods).”

c. (2) Boredom did have a statistically significant effect on the fit of the LME model of mood (as assessed by a chi^2 test, $p<0.05$). We characterised this effect as small for the following reason. The $R^2>0.3$ previously cited in the boredom section is the variance explained by all variables in the LME model. Most of this variance is explained by factors other than boredom (such as depression risk score’s ability to capture each subject’s initial mood). Including final boredom in the model, for example, increased the variance explained from $R^2=0.341$ (without boredom) to $R^2=0.359$ (with boredom), i.e. a $\text{delta}\_R^2 = 0.018$. We use a Cohen’s f^2 statistic (Cohen, 1988), to summarise this effect size. In Cohen’s guidelines, $0.02 \geq f^2 < 0.15$ is considered a small effect. Because we observed $f^2=0.0283$, we described the effect of final boredom’s inclusion in the model as statistically significant but small.

In response to this and other reviewer feedback, we have now reexamined our method of quantifying the ability of boredom, MW, and depression risk to quantify mood drift. Previously, we focused on the difference in variance explained between a model without the new factor in it and one with both the new factor and its interaction with time. This method was giving each factor credit for explaining initial mood as well as mood slope (i.e., mood drift), when our question was specifically about mood drift. In our new draft, we compare two models whose only difference is the addition of the new factor’s *interaction with time* (both models contain the new factor alone). We have also switched from a general residual sum-of-squares R^2 to the more specific R_1^2 (Snijders, 1994; Nakagawa, 2013) to capture the ability of the new factor’s interaction with time to explain *within-individual* variance. We use the difference in R_1^2 values between the expanded model (with the new factor’s interaction with time) and the reduced model (without it) to calculate an f^2 value to describe the effect size. This approach more specifically addresses the question of how well the new factor can capture each participant’s

mood drift. By this measure, state boredom has a very small effect on variance explained (f^2<0.02).

We believe that this new analysis answers our scientific question more completely than the analysis in our previous draft. We have therefore changed the main text to describe this new analysis. Because the previous analysis was the one described in our preregistration, we feel obliged to still report its results. But we have moved this report to the supplement.

To make this logic clearer to readers, we have modified text in the Results, Methods, and Supplementary Notes. To stay within the journal's word limits, we moved much of the explanatory text to the Methods section and Supplementary Notes. The Supplementary Note explaining the preregistered results is now followed by another that explains our revised approach and details the results (see next section for that new text).

d. (3) Our reasoning for focusing on the emotional dimension of MDES in the previous draft was as follows. (Poerio, 2013) found that it is not mind-wandering in the general or "traditional" sense (i.e., any task-unrelated thought) that decreases mood. Instead, it is mind-wandering with negative affective content. This notion is supported by current theories of mind-wandering not as a monolith, but as a collection of thoughts whose content shapes brain activity and behaviour (Smallwood, 2021). Research has linked thought probe responses about the affective content of this ongoing thought to brain activity patterns in the mOFC (Tusche, 2014). The PCA method described in (Turnbull 2019) provides a way to quantify the negative affective content of this ongoing thought that more robustly separates affective tone from the mere presence of task-unrelated thought.

However, the reviewer astutely points out that although our preregistered hypothesis was about the emotional content of thought, readers would be interested in any thought content that could explain mood drift. To evaluate this possibility, we used the same method described above to measure boredom's effect, but with time's interaction with the collection of all 13 MDES principal components (rather than just the emotion component) used as inputs to the model. (PCA was still used because it enforces orthogonality, which is desirable for our LME analysis (Schielzeth, 2020)). Using this method, we find that the full collection of MDES components have

a small effect on the within-individual variance explained by the model.

To make this reasoning clearer to readers, we have modified text in the Results, Methods, and Supplementary Notes. Again, because the previous analyses were preregistered, we have continued to use them and explain their motivation in a Supplementary Note. But this note is now followed by another that explains our revised approach and details the results.

i. The results section titled "Relationship to Boredom and Mind-Wandering" now reads:
"*Relationship to Boredom and Mind-Wandering*

We next examined whether the existing construct of boredom or mind-wandering (MW) could trivially explain mood drift. In a preregistered (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.-M.). Participants were randomised 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) (Hunter, 2016) afterwards (cohort BoredomAfterOnly, n=150), or before and afterwards (cohort BoredomBeforeAndAfter, n=150), and a trait-boredom questionnaire (the short boredom proneness scale, SBPS) (Struk, 2017).

In our LME model of mood, we added a factor for final state boredom (i.e., at the end of the rest block). We then compared this baseline model to one that further added the interaction between final-boredom and time. The difference represents the ability of boredom to account for mood drift. Whilst the model fit improved, the added within-individual variance explained by the addition of this new interaction term was very small (f^2=0.00578). The change in state boredom across the rest block produced similar results (f^2=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 Mind-Wandering

We also assessed whether mood drift could be explained by mind-wandering. New participants completed a rest block with interspersed mood ratings, plus an MDES questionnaire (Turnbull, 2019) afterwards (cohort MwAfterOnly, n=150), or before and afterwards (cohort MwBeforeAndAfter, n=150), and a trait-MW questionnaire (the mind-wandering questionnaire (MWQ) (Mrazek, 2013)). 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 (i.e., at the end of the rest block). We then compared this baseline model to 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. Whilst the model fit improved, the within-individual variance explained by the addition of these new interaction terms was small (f^2=0.0227). The change in MDES components across the rest block produced similar results (f^2=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). "

ii. In the methods section titled "Thought Probes and Activities Questions", the fourth paragraph now reads:
"As described by Ho et al. (Ho, 2020), 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 orthogonalises the MDES responses, which is desirable for their use as explanatory variables in an LME (Schielzeth, 2020).

For a preregistered 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.-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 prelearned 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. "

iii. A new methods section called "LME Model Comparisons" has been added to detail the new analyses that evaluate individual factors' interactions with time. It reads:
"To compare the ability of additional terms like depression risk and state boredom to explain variance in our model of mood, we employed an 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 (Snijders, 1994; Nakagawa, 2013). This calculation required a null model including only an intercept and random effects, which we defined as:

Mood ~ 1 + (1 + Time | Subject)

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

$$R_1^2 = 1 - \frac{\sigma_\varepsilon^2 + \sigma_\alpha^2}{\sigma_{\varepsilon 0}^2 + \sigma_{\alpha 0}^2}$$

where σ_ε^2 is the variance of the residuals of the model, σ_α^2 is the variance of the random effects, $\sigma_{\varepsilon 0}^2$ is the variance of the residuals of the null model, and $\sigma_{\alpha 0}^2$ 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 library (Barton, 2009), taking into account the correlation between model factors.

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

$$R_2^2 = 1 - \frac{\sigma_\varepsilon^2 + \sigma_\alpha^2/k}{\sigma_{\varepsilon 0}^2 + \sigma_{\alpha 0}^2/k}$$

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

Because the depression risk, boredom, and mind-wandering 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 statistic (Cohen, 1988; Selya, 2012), defined as:

$$f^2 = \frac{R_{AB}^2 - R_A^2}{1 - R_{AB}^2}$$

Where R_{A B}^2 is the variance explained by the expanded model and R_A^2 is the variance explained by the reduced model. Separate f^2 values can be calculated using the within-individual or between-individual variances. Using Cohen's guidelines (Cohen, 1988), f^2 ≥ 0.02 is considered a small effect, f^2 ≥

0.15 is considered a medium effect, and f^2 ≥ 0.35 is considered a large effect. "

iv. The supplementary note on preregistration ("Results of Preregistration on Boredom, Mind-Wandering, and Freely Chosen Activities") now begins with a disclaimer introducing the post-registration change. A new supplementary note that follows it ("Amended Analyses on Boredom and Mind-Wandering") describes the results in detail. We have omitted these changes in this response to save space, but they can be found at the end of the supplement.

e. Upon reflection, we concluded that this revised method could also be used to more accurately quantify the impact of depression risk on our study. The results of this analysis showed that depression risk score's interaction with time also had a very small effect on the within-individual variance explained by the model, R_1^2=0.291 (without this interaction term in the model) to R_1^2=0.293 (with it) (f^2=0.00289). This result has led us to reconsider depression risk's prominence in our results and discussion. We have removed its mention from the title, abstract, and Introduction. We have placed the results section and figure previously dedicated to it into a supplementary note ("Mood Drift Over Time Is Inversely Related to Depression Risk"). This note is now summarised in a new paragraph as part of the Results section dedicated to reward sensitivity. We hope that this change will more appropriately weight the relative attention given to depression risk, boredom, mind-wandering and their ability to explain mood drift.

   i. In the results section entitled "Mood Drift Over Time Is Associated with Sensitivity to Rewards", paragraph 2 now reads:
   *"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 * time interaction, Mean ± SE = 0.515 ± 0.109 % mood/min, t_{869}=4.75, p<10^{-6}). Whilst the model fit improved, the within-individual variance explained by the addition of this interaction term was very small (f^2 = 0.00289) (Cohen, 1988; Selya, 2012). 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, Figure 3, left) and was robust to methodological artefacts such as floor effects (Supplementary Notes E.-G.)."*

ii. A new Supplementary Note (E) has been added to detail the depression risk findings previously found in the main text. Another (G) has been added to address the impact of floor effects on these results. We have omitted these changes in this response to save space, but they can be found in the supplement.

f. In addition, we have read the updated draft closely and attempted to correct statements that might seem conflicting or misleading to the reader. These changes are detailed in the response to other reviewer comments.

5. REVIEWER 3: Is it more accurate to describe this phenomenon as "passage of time DURING TASKS"? I think it's important to note that resting states are still very much a "task" compared to daily life. Indeed, in many EMA studies assessing mood, the mood ratings to do not change systematically with short time gaps, so it seems to be specific to having a predefined task.
    a. RESPONSE: Thank you for noting this concern, which was shared by Reviewer 2. We have now highlighted the phenomenon's apparent specificity to experimental tasks and rest in several places in the text. In the introduction, we now frame the behavioural constraint of a psychological task or rest period as a negative affective stimulus that could gradually draw mood downward, and we specifically mention that the effect was not observed when participants chose their own activities. In the discussion, we reiterate this specificity and give a more prominent position to the results of the freely chosen activities experiment.
        i. In the introduction, paragraph 5 now reads:
        "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 affect (Eastwood, 2012). We might therefore conceive of a psychological task's behavioural constraint as a sort of negative affective stimulus that could gradually draw mood downward."
        ii. The final paragraph of the introduction, sentences 3-4, now reads:
        "This downward mood drift was replicated in 19 large and varied cohorts, totaling 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."

iii. In the discussion, paragraph 1 now begins:
"In this study, we describe the discovery of a highly replicable and relatively large effect which we call Mood Drift Over Time: the average participant's mood gradually declined with time as they completed simple tasks or rest periods."

iv. In the discussion, paragraph 4 now reads:
"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."

6. REVIEWER 3: Minor: In general, many of the citations used did not seem like the best examples for the claims being made. I suggest the authors more carefully examine who they are citing for what claim — i.e., is it the original source.
    a. RESPONSE: We have done our best to investigate each citation and ensure that it is the original source of the statement being made. These updated citations can be seen in the modified manuscript that has changes tracked. Some of the insufficiently cited claims may also have been removed to address reviewer comments and journal word limits.
        i. We realise that our first citation "Statistical Parametric Mapping: The analysis of Functional Brain Images" is a somewhat oblique reference for the claim that "An important but typically implicit notion amongst behavioural and affective scientists is that each participant has a baseline mood or affective state that will remain constant during an experiment or only vary with emotionally salient events." However, this assumption is so rarely explicitly stated, that we were unable to find any better citations. In order to reflect this ubiquity, we have removed the modifier "typically" from this opening sentence.
        ii. For the claim "Affective chronometry research has demonstrated that affect changes systematically with time after an affective stimulus…", we have added:
        - Frijda, N., Mesquita, B., Sonnemans, J., and Goozen, S. The duration of affective

phenomena or emotions, sentiments and passions. In International Review of Studies on Emotion, vol. 1. Jan. 1991, pp. 187–225.
- Scherer, K. R., and Wallbott, H. G. Evidence for universality and cultural variation of differential emotion response patterning. Journal of Personality and Social Psychology 66, 2 (Feb. 1994), 310–328.

iii. For the claim "that individuals vary in the rates at which positive or negative affect decays after an event", we have added:
- Gilboa, E., and Revelle, W. Personality and the Structure of Affective Responses. In Emotions. Psychology Press, 1994.

iv. For the claim "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)," we have added:
- McRae, T. W. Opportunity and Incremental Cost: An Attempt to Define in Systems Terms. The Accounting Review 45, 2 (1970), 315–321.

v. For the claim "Mood is central to depression, which is thought to relate etiologically to reward responsiveness," we have added:
- Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G., and Fava, M. Reduced hedonic capacity in major depressive disorder: Evidence from a probabilistic reward task. Journal of Psychiatric Research 43, 1 (nov 2008), 76–87.
- Halahakoon, D. C., Kieslich, K., O'Driscoll, C., Nair, A., Lewis, G., and Roiser, J. P. Reward-processing behavior in depressed participants relative to healthy volunteers: A Systematic Review and Meta-analysis. JAMA psychiatry (2020).

7. REVIEWER 3: Please also add a citation for the reward valuation claim. Related: the authors go in to a bit of theorizing in the discussion about the depression results, but I think more can be developed on this by incorporating more literature.
    a. RESPONSE: Thank you for pointing out this specific area for improvement. We have incorporated more literature in the areas of the manuscript relating to reward valuation. We have also removed some text in an effort to stay within the journal's word limits.
        i. In the results subsection called "Mood Drift Over Time Is Associated with Sensitivity to Rewards" (first paragraph, first sentence), we now cite an empirical paper and a meta-analysis:
        "Mood is central to depression, which is thought to relate etiologically to reward

responsiveness (Pizzagalli, 2008; Halahakoon, 2020). 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."

ii. In the discussion, paragraphs 2-3 now read:
"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 behaviour (Agrawal, 2020). Lower 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) (Geana, 2016).

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 (e.g., by Schopenhaur (Schopenhaur, 1851) and Kierkergaard (Kierkergaard, 1992)). Yet it has been argued cogently (Elpidorou, 2014) that such a view conflates negative affect as a trait (e.g., proneness to boredom) with negative affect as a state (a momentary experience). Since valuation of reward is thought to be reduced in depression (Pizzagalli, 2008; Halahakoon, 2020), 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 (Figure 3). It is tempting to speculate that reduced mood drift could contribute to reduced motivation for action or environmental change in those with depression."

8. REVIEWER 3: I understand space limitations and journal specific formatting. However, I found the amount of results that were simply state without much statistical reference a bit difficult to follow and evaluate. It would be helpful if more results and statistics were included in the

main paper rather than an overview claim. However, I defer to the journal Editors on this particular comment.

a. RESPONSE: We have cut text to make room to include more results and statistics in the main text. The updated draft now includes previously-supplemental statistics in the main text's results section. We are willing to copy more results from the supplement to the main text, but (as the reviewer notes) we are somewhat constrained in our ability to do so by the journal's word limit.
    i. In the Results section titled "Mood Drift Over Time is Robust to Methodological Choices", the numbered list now includes statistics for each item:
    "*1. The aversive nature of rating one's mood: more frequent ratings did not significantly change mood drift (inter-rating-interval x time interaction = -0.0103 %mood, 95%CI = (-0.0267, 0.0061), t_{810} = -1.23, p=0.219.*
    2. The method of rating mood and its susceptibility to fatigue: making every mood rating require an equally easy single keypress did not significantly change mood drift (-2.22 vs. -2.45 %mood/min, 95%CI = (-0.772, 1.23), t_{70} = 0.427, p=0.671).
    3. The expected duration of the rest period: groups expecting different rest durations did not have different mood drift (-1.47 vs. -1.53%mood/min, 95%CI = (-0.613, 0.743), t_{104} = 0.185, p=0.854).
    4. Multitasking or task switching: participants moved their mood rating slider on 97.7% of trials."
    ii. In the Results section titled "Inter-individual differences", sentence 4 now reads:
    *"Using cohorts that completed the task more than once, we found that these individual differences had moderate, statistically significant stability across blocks (ICC(2,1) = 0.465, p=2.8*10^{-6}), days (ICC(2,1) = 0.343, p=3.1*10^{-3}), and weeks (ICC(2,1) = 0.411, p=1.9*10^{-8}) (Supplementary Note D.). "*

9. REVIEWER 3: The abstract was full of results which is nice in some ways, but many of them were difficult to understand without being contextualized.
    a. RESPONSE: Thank you for the suggestion to revisit the abstract, which is key to readers' experience of the manuscript. We have interpreted "without being contextualized" as referring to the lack of detail about what participants did in the study. To rectify this, we have added a sentence early in the abstract stating that subjective momentary mood ratings were interspersed into repetitive psychological

paradigms. To stay within word limits, we have removed less central details about the tasks investigated and depression risk.

i. The abstract now reads:
"*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. 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 totaling 28,482 adult and adolescent participants. The drift was relatively large (-13.8% after 7.3 minutes 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.*"

10. REVIEWER 3: Please provide descriptive statistics for the key variables in the supplementary.
    a. RESPONSE: We have done our best to add descriptive statistics and confidence intervals to each case where an inferential statistic was referenced in both the main text and the supplement. If we have missed any, we will be happy to add them.
        i. In the Results and Supplement, each test now includes the mean or median of each group and the std error, IQR, or 95% confidence interval for the difference in addition to the statistic with the degrees of freedom and the p value.
    b. We also identified areas that were lacking inferential statistics, which were each referencing statistics already listed elsewhere in the text. We have now revised the supplement to explicitly reference those sections.
        i. In the supplementary section titled "Mood Drift Over Time Is Not a Product of Aversive Mood Ratings", the second-to-last sentence now reads:
        *"This finding was later confirmed by our multi-cohort LME model, in which a participant's mean inter-rating interval did not have a significant relationship with their slope parameter (impact of meanIRIOver20*Time interaction = -0.0103 ± 0.0084 % mood, t_{810} = -1.23, p=0.219, Supplementary Table 2). "*
        ii. In the supplementary section titled "Mood Drift Over Time Is Not Driven by Multitasking," sentence 3 now reads:

"Cohorts with short rest periods between mood ratings likely had to make responses too frequently to multitask, but the time between ratings did not change participants' level of mood drift (see section titled "Mood Drift Over Time Is Not a Product of Aversive Mood Ratings" above)."

Reviewer #4:

Remarks to the Author:

1. REVIEWER 4: Across a range of samples and tasks, people reported lower positive affect/higher negative affect when asked to report mood across multiple occasions, leading to what the authors call "passage-of-time dysphoria". This appears to be an important and highly replicable methodological artifact that will be of interest to a great many researchers, and has substantial implications for the interpretation of common pre-post designs (including those involving random assignment to condition). The impact of the present work thus seems sizeable, and the authors have done a thorough job of addressing earlier reviewer concerns. The comments here are intended in the spirit of strengthening what is already a sizeable contribution.

   Although the paper is framed in terms of affective neuroscience, the implications for the field of affective science more generally as well as many subfields of psychology (e.g, well-being, positive psychology) would also seem substantial. Especially given the generalist nature of an outlet like Nature Human Behavior, reframing portions of the paper to be less discipline-specific would likely broaden its potential impact.

   In doing so, the authors may wish to draw on existing work in justifying their use of mood as the primary construct of interest. Mood is a well-defined construct in affective science, and one that can be distinguished from other affective components (such as discrete emotions, or dimensions of affect such as valence and arousal); making that distinction explicit would be helpful in clarifying the contribution of the present work. For instance, it may be more useful to speak of declines in positive affect/increases in negative affect over time.

   Likewise, self-report is the gold standard for the measurement of emotion (and mood), and the key criterion variable against which all physiological and neural "markers" are benchmarked. Furthermore, the usefulness of such affective markers is itself dubious, given evidence that discrete emotions cannot be distinguished by their physiological signatures

(Siegel et al., 2018), facial expressions (Gendron et al., 2015; Barrett et al., 2019), or neural activity (Lindquist et al., 2015).

a. RESPONSE: We thank the reviewer for their positive words about the manuscript, and we share their hope that a wide readership will find these findings relevant to their work. In response to concerns from other reviewers, we have now adopted the term "Mood Drift Over Time" ("mood drift" for short) to refer to the phenomenon described in this manuscript. We hope that this broader term will help readers outside of affective neuroscience to see the potential applicability of this finding to their own work. To more specifically address the concern expressed by Reviewer 4 and others from Reviewer 3, we have incorporated additional theoretical bases into our introduction, and some of this language links our work to concepts in additional fields, including behavioural economics, affective chronometry, and psychopathology.
    i. We have replaced each instance of "*affective neuroscience*" with "*affective science*," including text in the abstract, introduction, and discussion.
    ii. The introduction, paragraphs 3-6 now read:
    "Whilst 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 (Frijda, 1991; Scherer, 1994; Davidson, 1998; Davidson, 2015), and that individuals vary in the rates at which positive or negative affect decays after an event (Gilboa, 1994; Hemenover, 2003). Such individual differences may be linked to mental health. For instance, psychopathologists theorise that anhedonia, a symptom of both depression and schizophrenia, arises from a failure to sustain reward responses for a normative period of time (Kring, 2014). And studies of ADHD suggest that hyperactivity's impulsive behaviour results from delay aversion, the idea that a delay is itself unpleasant and impulsivity is simply a rational choice to avoid it (Sonuga-Barke, 1992; Solanto, 2001; Sonuga-Barke, 2016).

    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) (McRae, 1970; Hoskin, 1983; Palmer, 1999). This idea is fundamental to the explore/exploit question that has recently preoccupied neuroscientists (Cohen, 2007; Constantino, 2015; Addicott, 2017). 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 (Geana, 2016; Agrawal, 2020).

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 affect (Eastwood, 2012). 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 (Robison, 2020; Killingsworth, 2010; Fox, 2014; Fox, 2018). 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 (Van Hooff, 2014; Miner, 2010). 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. But the size, stability, and clinical correlates of this variability remain unexplored."

b. The reviewer also points out the potential utility of speaking of positive/negative affect. As Reviewers 2 and 3 pointed out, the term "passage-of-time dysphoria" was inviting readers to infer that negative affect was increasing, but our results showed average mood ratings remaining above 0.5. To be more precise in our terminology, we have now adopted the term "Mood Drift Over Time" ("mood drift" for short) to refer to the phenomenon described in this manuscript. By removing the reference to "dysphoria", we are no longer implying an increase in negative affect, but rather a gradual decline in overall mood. "Mood" and "affect" may both be used to describe affective states, but "mood" has been used before with this experimental setup (Rutledge, 2014, Keren, 2021). To remain consistent, we have chosen to use the term "mood" throughout the manuscript.

c. We also thank the reviewer for their constructive and specific suggestions to help readers understand the choice to use mood as the primary construct of interest. We have done our best to incorporate them into the manuscript. We have added text to the discussion to help convince readers that mood (and, specifically, self-reported mood) is a well established and widely relevant construct worthy of our study and their attention. This includes several of the references suggested in comment #2 to justify self-reported boredom (Diener, 2000; Robinson, 2002), which appear to similarly support the use of self-reported affect of any kind.

   i. The discussion, paragraphs 11-12 now read:
   "*First, this study uses self-reported momentary mood ratings as in previous studies with similar methodology (Rutledge, 2014; Keren, 2021). Such ratings can be criticised 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 (Frijda,2009; Ekkekakis, 2013) makes it central to the study of mood disorders defined by long-term affect (Rottenberg, 2005). Mood is distinct from emotion, in part, by being less temporally responsive (Nowlis, 1956; Ekman, 1992; Watson, 2000). 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 (Diener, 2000; Watson, 2000; Robinson, 2002). It is widely used in clinical (Costello, 1988), epidemiological (Pavot, 1993), and psychological research (including ecological momentary assessment (Ebner-Priemer, 2009). 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 signatures (Siegel, 2018), facial expressions (Gendron, 2015; Barrett, 2019), or neural activity (Lindquist, 2012). In our experiments, initial mood ratings showed strong association with trait mood ratings, underscoring their psychometric validity (Supplementary Figure 12)."

2. REVIEWER 4: For this reason, the definition and measurement of boredom on page 10 is also however inappropriate. While boredom is defined here as "a state of low arousal and dissatisfaction," more recent evidence shows that this definition is not empirically supported (see e.g., Westgate & Wilson, 2018), and that boredom is associated with both high arousal

states as often as with low arousal, or mixed states. For this reason, the MSBS is not an appropriate measure of boredom and, indeed, manipulations that equally increase boredom (when measured by direct self-report) yield differing patterns of responses on the various MSBS subscales (e.g., dysphoria, agitated affect, inattention, disengagement, time distortions). This, unfortunately, makes the studies reported on pp. 10-12 difficult to interpret and does not rule out boredom as a potential mechanism to explains these results. A better measure of boredom would be face-valid items asking participants to directly report how bored they feel (see e.g., Barrett, 2004; Diener, 2000; Nisbett & Wilson, 1977; Robinson & Clore, 2002; Westgate & Steidle, 2021). This is particularly important in light of work showing that perceived autonomy can also potentially reduce boredom (see work on constraint and boredom by Thackray et al and Struk et al), and that people readily feel bored when asked to sit and "do nothing" (e.g., Wilson et al., 2014), which strongly suggests that boredom may be one contributing mechanism.

a. RESPONSE: We agree that MSBS is an incomplete measure for boredom, and we cannot conclude purely from this measure that boredom is not driving this effect. We are now alerting the reader to this and are more circumspect in our inference – please see below. As the reviewer suggests in their final comment, unpacking complex relationships like that between our finding and boredom is an important avenue for future work. In the interest of making the current uncertainty about boredom clearer to the reader, we have added boredom-specific text to the discussion.
    i. In the discussion, a new paragraph 7 now reads:
    "*The distinction between mood drift and boredom requires special consideration due to their apparent similarities. State boredom assessed using the MSBS-SF (Hunter, 2016) 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 conceptualisation of boredom and its heterogeneity (Westgate, 2018; Elpidorou, 2014; Eastwood, 2012). 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 boredom (Barrett, 2004), enabling more frequent assessment of boredom as an emotion (Westgate, 2020).*"

3. In sum, the authors have provided a substantial body of evidence for the effect encompassing a wide range of designs, datasets, and participant demographics, which would tend to point heavily in favor of generalizability. I particularly liked the experiments ruling out other potential mechanisms (e.g., regression to the mean), many of which were run in response to earlier reviewer comments and are detailed in the supplemental materials, and found the majority of them highly persuasive. No doubt, like most phenomenon, mood decline will also be subject to moderating effects. Unpacking what those circumstances and theoretical moderators are, however, seems more an avenue for future work than key questions necessary to resolve in this initial paper. I have no doubt that such future research will be fruitful and informative.
    a. RESPONSE: We thank the reviewer for their kind words about the manuscript. We hope that the changes detailed above have further strengthened the paper and advanced its suitability for publication in Nature Human Behaviour.

Decision Letter, second revision:

15th November 2022

Dear Dr. Jangraw,

Thank you for submitting your revised manuscript "A Highly Replicable Decline in Mood During Rest and Simple Tasks" (NATHUMBEHAV-210515366B). It has now been seen by the original referees and their comments are below. As you can see, the reviewers find that the paper has improved in revision. We will therefore be happy in principle to publish it in Nature Human Behaviour, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements within a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Please do not hesitate to contact me if you have any questions.

Sincerely,

Arunas Radzvilavicius, PhD
Editor, Nature Human Behaviour
Nature Research

Reviewer #3 (Remarks to the Author):

I would like to thank the authors for their diligent responses to the reviewer concerns. After reading through their response letter and the revised manuscript, I am happy to report that I think the contribution and interpretation are much improved. Most of my main concerns have been adequately addressed, and I think the findings have important contributions for many fields.

My remaining concerns are as follows:

I do not think the mind wandering section --namely that it is weakly correlated — is entirely accurate. The MDES questionnaire does not capture mind wandering; it captures a host of different qualities of ongoing thought, that may or may not have any number of characteristics (and these characteristics are then dependent on the PCA, which will vary across studies). I believe the authors are measuring general qualities of thought, not mind wandering specifically. If mind wandering is what the authors wish to comment on, please add how it is defined and operationalized in light of the broader literature. As well, the trait level variable did not explain any additional variance. Given this, I'm not sure the weakly related interpretation is warranted either (even when a traditional measure of mind wandering is used).

A general comment on effect sizes and interpretations: A correlation of .016 was interpreted as a small but weak correlation elsewhere in the manuscript. I find it very hard to endorse this as even a weak correlation, as the sample size is essentially influencing the p value but a correlation of .016 is not even in line with a small effect size according to conventions. Such interpretations should likely be made more carefully throughout the paper. With large sample sizes (a good thing), effect sizes are much more important, especially for a general outlet such as this.

Author Rebuttal, second revision:

Response to Reviewers

Reviewer #3 (Remarks to the Author):

I would like to thank the authors for their diligent responses to the reviewer concerns. After reading through their response letter and the revised manuscript, I am happy to report that I think the contribution and interpretation are much improved. Most of my main concerns have been adequately addressed, and I think the findings have important contributions for many fields.

My remaining concerns are as follows:

I do not think the mind wandering section --namely that it is weakly correlated — is entirely accurate. The MDES questionnaire does not capture mind wandering; it captures a host of different qualities of ongoing thought, that may or may not have any number of

characteristics (and these characteristics are then dependent on the PCA, which will vary across studies). I believe the authors are measuring general qualities of thought, not mind wandering specifically. If mind wandering is what the authors wish to comment on, please add how it is defined and operationalized in light of the broader literature. As well, the trait level variable did not explain any additional variance. Given this, I'm not sure the weakly related interpretation is warranted either (even when a traditional measure of mind wandering is used).

A general comment on effect sizes and interpretations: A correlation of .016 was interpreted as a small but weak correlation elsewhere in the manuscript. I find it very hard to endorse this as even a weak correlation, as the sample size is essentially influencing the p value but a correlation of .016 is not even in line with a small effect size according to conventions. Such interpretations should likely be made more carefully throughout the paper. With large sample sizes (a good thing), effect sizes are much more important, especially for a general outlet such as this.

Our Response:

Thank you for your constructive comments, both in previous drafts and this one. We have removed the mention of the very weak correlation you identified, as it is not central to our findings. We hope that this will remove a source of potential confusion for readers.

We agree that the term mind-wandering (MW) presents a challenge in our study. Readers commonly ask about its influence on our results, but disagreements about its definition make it difficult to address comprehensively. We use MDES responses as an umbrella for thought content that includes, but is not limited to, the task unrelatedness sometimes used as a definition of MW (Seli, 2018). Some researchers strongly disagree with this definition (Christoff, 2018). The text now cites 3 papers that articulate this difference of opinion to help readers draw their own conclusions.

To be circumspect about our findings and their interpretation, we now refer to the MDES responses as "thought content" in the introduction and section titles. We link this term to MW when first address the MDES results. We use MW in other places as shorthand because it is brief, intuitive, and consistent with our preregistrations.

We wish to alert the reviewers and editors that we have also made a change to a figure. In the original draft of the manuscript, the left panel of Figure 3 was referenced before the other two panels in the main text as a replication of mood drift's relation to depression risk. However, in the process of responding to reviewer comments, this depression discussion was

moved to the supplement, and this panel of Fig. 3 was changed to being mentioned last (and only tangentially). To meet editorial guidelines about referencing figures in order, we have removed this panel from Figure 3. It is still present in the supplement (Fig. S10, bottom right), and we now reference this supplementary figure when referring to the replication of the depression risk finding.

Final Decision Letter:

Dear Professor Jangraw,

We are pleased to inform you that your Article "A Highly Replicable Decline in Mood During Rest and Simple Tasks", has now been accepted for publication in Nature Human Behaviour.

Please note that *Nature Human Behaviour* is a Transformative Journal (TJ). Authors whose manuscript was submitted on or after January 1st, 2021, may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. IMPORTANT NOTE: Articles submitted before January 1st, 2021, are not eligible for Open Access publication. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including self-archiving policies. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see http://www.nature.com/nathumbehav/info/gta). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication

date (the day on which it is uploaded onto our web site).

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Human Behaviour as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Human Behaviour logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

You will not receive your proofs until the publishing agreement has been received through our system.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

We look forward to publishing your paper.

With best regards,

Arunas Radzvilavicius, PhD
Editor, Nature Human Behaviour
Nature Research