Neuroscience
Computational and Systems Biology

Both neuropsychiatric symptoms and circadian rhythm alter effort-based decision-making

Reviewed Preprint
v2 • August 23, 2024
Revised by authors

Reviewed Preprint
v1 • June 14, 2024

Sara Z Mehrhof ✉, **Camilla L Nord**

MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK • Department of Psychiatry, University of Cambridge, Cambridge, UK

Abstract

Motivational deficits are common in several brain disorders, and motivational syndromes like apathy and anhedonia predict worse outcomes. Disrupted effort-based decision-making may represent a neurobiological underpinning of motivational deficits, shared across neuropsychiatric disorders. We measured effort-based decision-making in 994 participants using a gamified online task, combined with computational modelling, and validated offline for test-retest reliability. In two pre-registered studies, we first replicated studies linking impaired effort-based decision-making to neuropsychiatric syndromes, taking both a transdiagnostic and a diagnostic-criteria approach. Next, testing participants with *early* and *late* circadian rhythms in the morning and evening, we find circadian rhythm interacts with time-of-testing to produce parallel effects on effort-based decision-making. Circadian rhythm may be an important variable in computational psychiatry, decreasing reliability or distorting results when left unaccounted for. Disentangling effects of neuropsychiatric syndromes and circadian rhythm on effort-based decision-making will be essential to understand motivational pathologies and to develop tailored clinical interventions.

eLife assessment

This **important** study provides **convincing** evidence that both psychiatric dimensions (e.g. anhedonia, apathy, or depression) and chronotype (i.e., being a morning or evening person) influence effort-based decision-making. This is of importance to researchers and clinicians alike, who may make inferences about behaviour and cognition without taking into account whether the individual may be tested or observed out-of-sync with their phenotype. The current study can serve as a starting point for more targeted investigation of the relationship between chronotype, altered decision making and psychiatric illness.

https://doi.org/10.7554/eLife.96803.2.sa4

1 Introduction

Our circadian rhythm aligns us with our environment, regulating physiological and behavioural processes to follow 24-hour rhythms[1]. Circadian integrity is pivotal to mental wellbeing and has been bidirectionally linked to numerous psychiatric disorders[2–8]. Yet little is known about the cognitive or computational mechanisms of circadian dysfunction—and their alignment or diversion from mechanisms driving neuropsychiatric symptoms.

Inter-individual differences in circadian timing and alignment manifest behaviourally as chronotypes (i.e., diurnal preference)[9,10], with individuals commonly categorized as early, late, or intermediate chronotypes[11]. A disproportionate number of psychiatric patients have a late chronotype, based on self-report[12] and genetic analysis[9]. Within clinical groups, late chronotype has been linked to depression severity and non-remission[13], higher rates of psychiatric and general medical comorbidities[14], more severe cognitive impairment, and higher symptoms of apathy[15,16]. Converging evidence on the importance of circadian alignment in psychiatric pathology has led to proposals of a circadian psychiatric phenotype, either within disorders[14,17] or cuding across diagnostic categories[18,19].

Syndromes of deficient motivational behaviour, such as apathy and anhedonia, are also observed across neuropsychiatric disorders[20–23], suggesting transdiagnostic relevance[24]. Anhedonia and apathy are associated with worse clinical outcomes[25,26] and are poorly targeted by current treatments[27–29]. Empirical work suggests a common underlying neurocognitive mechanism: the integration of costs and benefits during effortful decision-making[24]. Effort-based decision-making is commonly assessed using effort expenditure tasks: Subjects are asked to decide whether to pursue actions associated with varying levels of effort and reward levels[30]. Computational models applied to effort-based decision-making tasks provide a formal mathematical estimate of a subject's integration of costs and benefits into a subjective value[31,32]. Higher costs devalue associated rewards, an effect referred to as *effort-discounting*[33–37]. This computational approach enables measurement of inter-and intra-individual differences on distinct aspects of effort-based decision-making.

One key source of individual differences in motivational behaviour and effort-based decision-making is likely dopamine signalling, especially dopaminergic projections from the ventral tegmental area (VTA) to the ventral striatum[24]. Pre-clinical animal studies show dopamine depletion reduces engagement in effortful behaviour[38,39], while dopamine enhancement promotes motivational effort exertion[40,41]. In humans, dopamine depletion reduces willingness to exert effort for reward[42,43], while pharmacological dopamine enhancement increases motivation in effort-based decision—making[44,45]. Further, naturally occurring variations in dopamine responsivity are correlated with effort-based decision-making: a higher dopamine responsivity (as quantified with positron emission tomography following *d*-amphetamine administration) is associated with willingness to exert greater effort for larger rewards[46].

Bi-directional links between chronobiology and several neurotransmitter systems have been reported, including dopamine[47]. In animals, dopamine transmission and biosynthesis vary diurnally[48,49], and growing evidence suggests a bi-directional regulation between dopamine signalling and circadian rhythm[50–52]. In human studies, dopamine availability, dopamine transporter genes, and dopamine receptors have been linked to proxies of circadian rhythm[53–55] and circadian-regulating gene polymorphisms[56]. On a behavioural level, sleep deprivation, poor sleep quality, and insomnia were linked to low motivation in effort-based decision—making[57–59] and evening bright-light exposure enhanced effort willingness, possibly by enhancing dopamine through melatonin suppression[60]. Early chronotype predicted treatment

effect on motivational behaviour in a sample of depressed subjects with comorbid insomnia[61]. Chronotype effects are also reported for other reward decision—making tasks, with late chronotypes showing higher delay discounting[62], less rational decision-making[63], and lower willingness to take risks for rewards[64,65]. A circadian effect on decision-making under risk is reported, with the sensitivity to losses decreasing with time-of-day[66]. This suggests that chronobiology may contribute to individual differences in effort-based decision-making, potentially in parallel ways with neuropsychiatric syndromes.

Here, we tested the relationship between motivational decision-making and three key neuropsychiatric syndromes: anhedonia, apathy, and depression, taking both a transdiagnostic and categorical (diagnostic) approach. To do this, we validate a newly developed effort-expenditure task, designed for online testing, and gamified to increase engagement. Participants completed the effort-expenditure task online, followed by a series of self-report questionnaires.

Next, we pre-registered a follow-up experiment to directly investigate how circadian preference interacts with time-of-day on motivational decision-making, using the same task and computational modelling approach. While this allows us to test how circadian effects on motivational decision-making compare to neuropsychiatric effects, we do not test for possible interactions between neuropsychiatric symptoms and chronobiology. All analyses were pre-registered (except when labelled as exploratory): see hdps://osf.io/2x3au and hdps://osf.io/y4ke.

2 Results

2.1 Sample characteristics

Nine hundred and ninety-four participants completed all study components (i.e., demographic questions, effort-expenditure task, self-report questionnaires). After exclusion (see Methods 4.1.5), 958 participants were included in our analyses. We used a stratified recruitment approach to ensure our sample was representative of the UK population in age, sex, and history of psychiatric disorder[67–69]; mean questionnaire-based measures were comparable to previous general population studies (**Table 1**).

Questionnaire sum scores highly correlated within groupings of questionnaires targeting psychiatric symptoms, chronobiology, and metabolic health. We also found significant correlations between some, but not all, questionnaires (**Fig. 1**)

2.2 Effort-expenditure task

In this novel, online effort-expenditure task (**Fig. 2A-B**), subjects were given a series of challenges associated with varying levels of effort and reward. By weighing up efforts against rewards, they decide whether to accept or reject challenges. We first use model-agnostic analyses to replicate effects of effort-discounting (i.e., devaluation of reward with increasing effort). Next, we took a computational modelling approach to fit economic decision-making models to the task data (**Fig.3A-D**). The models posit efforts and rewards are joined into a subjective value (SV), weighed by individual effort (β_E) and reward sensitivity (β_R) parameters. The subjective value is then integrated with an individual motivational tendency (a) parameter to guide decision-making. Specifically, the motivational tendency parameter determines the range at which subjective values are translated to acceptance probabilities: the same subjective value will translate to a higher acceptance probability the higher the motivational tendency.

	Included	Excluded
Cohort size (%)	958 (96.4%)	36 (3.62%)
Demographics		
Age, mean (SD; range)	45.00 (15.01; 18-79)	47.90 (13.60; 20-70)
Gender, number (%)		
Male (%)	470 (49.06)	12 (33.33)
Female (%)	484 (50.52)	24 (66.67)
Non-binary (%)	4 (0.42)	0 (0.0)
Ethnicity, number (%)		
White (%)	852 (88.94)	28 (77.78)
Asian (%)	53 (5.53)	4 (11.1)
Black (%)	27 (2.82)	3 (8.33)
Mixed (%)	18 (1.88)	1 (2.78)
Other (%)	8 (0.84)	0 (0.0)
SES (/9), median (IQR)	5 (4-6)	5 (4-6)
Psychiatric comorbidities		
Current or past, number (%)		
Any (%)	264 (27.60)	5 (13.90)
Major depressive disorder (%)	94 (9.81)	1 (2.78)
Generalised or social anxiety disorder (%)	195 (20.35)	2 (5.56)
Current antidepressant use, number (%)	151 (15.80)	5 (13.9)
Task metrics		
Testing time, number (%)		
Morning testing (8:00-11:59; %)	492 (51.40)	19 (52.80)
Evening testing (18:00-21:59; %)	458 (47.80)	17 (47.2)
Time taken (minutes), mean (SD; range)	33.13 (9.63; 22-151)	37.06 (15.30; 26-105)
Mean clicking calibration, mean (SD; range)	60.6 (16.10, 8-206)	74.10 (123.00, 0-721)
Psychiatric questionnaire measures		
SHAPS, mean (SD; range)	9.15 (6.28; 0-36)	10.90 (6.97; 1-33)
DARS, mean (SD; range)	54.50 (9.18, 17-68)	53.70 (9.77, 36-68)
AES, mean (SD; range)	55.70 (9.42; 25-72)	55.10 (9.51; 37-71)
M.I.N.I., current MDD (%)	56 (5.85)	
Circadian questionnaire measures		
MEQ, mean (SD; range)	52.80 (10.6, 18-81)	52.08 (7.87, 34-71)
MCTQ, mean time in min. (SD; range)	03:56 (89min; 00:14-11:05)	04:03 (87min; 01:05-09:05)
Metabolic questionnaire measures		
BMI, mean (SD; range)	26.90 (6.29, 15.20-63.30)	27.02 (5.77, 19.10-46.90)
FINDRISC, mean (SD; range)	7.46 (5.09, 0-25)	8.56 (5.26, 0-22)

Table 1

Demographic characteristics and descriptive questionnaire measures in the included sample and excluded participants.

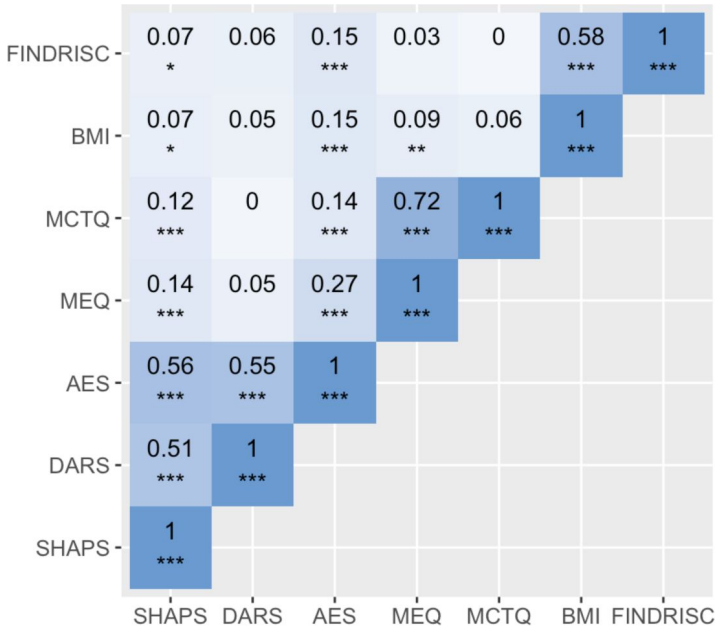


Figure 1.

Correlations between questionnaire scores

Correlations between questionnaire sum scores for the Snaith Hamilton Pleasure Scale (SHAPS), the Dimensional Anhedonia Rating Scale (DARS), the Apathy Evaluation Scale (AES), Morningness-Eveningness Questionnaire (MEQ), Munich Chronotype Questionnaire (MCTQ), Body mass index (BMI) and the Finish Diabetes Risk Score (FINDRISC). Asterisks indicate significance: * p<.05, ** p<.01. *** p<.001 (not accounting for multiple comparisons). Note that sum scores for the AES and the DARS have been transformed such that increasing scores can be interpreted as higher symptom severity, in line with the SHAPS. Sum scores of the MEQ have been transformed such that higher scores indicate higher eveningness, in line with the MCTQ.

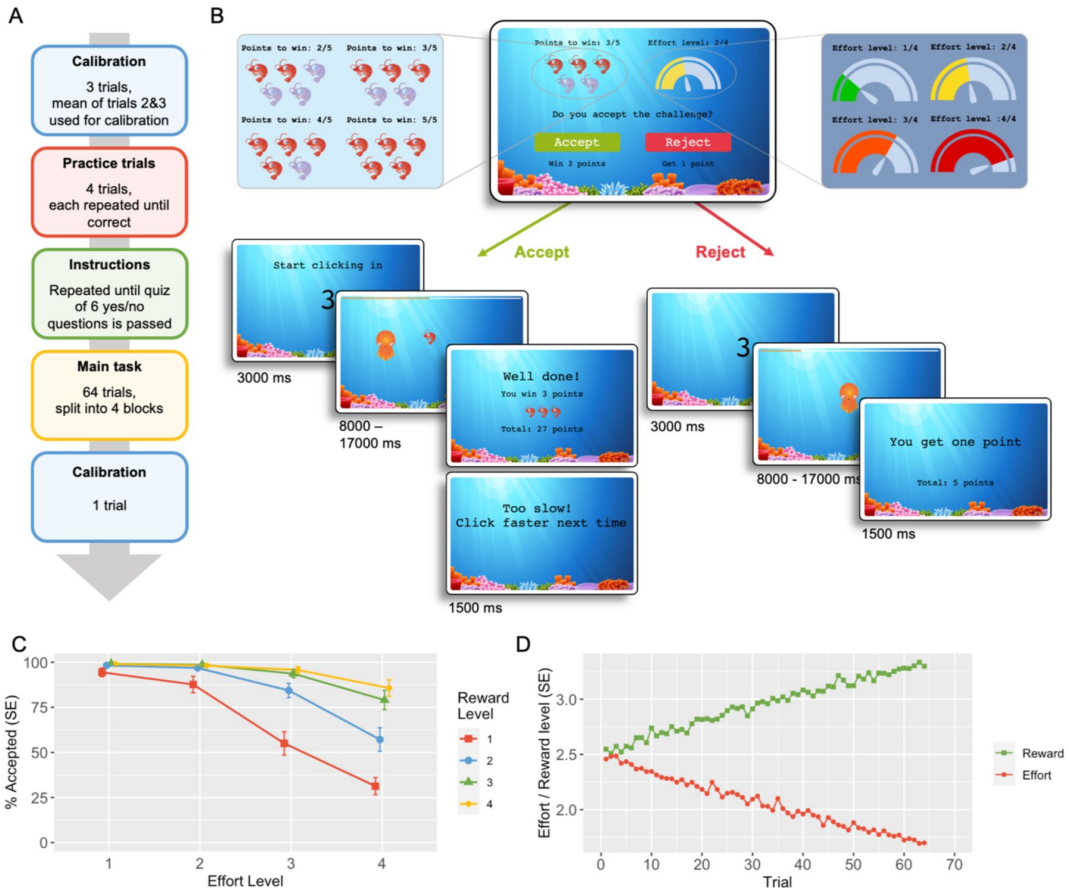


Figure 2.

Effort-based decision-making: task design and model-agnostic results.

A: The task can be divided into four phases: a calibration phase to determine individual clicking capacity to calibrate effort-levels, practice trials that participants practice until successful on every effort-level, instructions and a quiz that must be passed, and the main task, consisting of 64 trials split into 4 blocks. **B:** Each trial consists of an offer with a reward (2,3,4, or 5 points) and an effort level (1,2,3, or 4, scaled to the required clicking speed and time the clicking must be sustained for) that subjects accept or reject. If accepted, a challenge at the respective effort level must be fulfilled for the required time to win the points. If rejected, subjects wait for a matched amount of time and receive one point. **C:** Proportion of accepted trials, averaged across participants and effort-reward combinations. Error bars indicate standard errors. **D:** Staircasing development of offered effort and reward levels across the task, averaged across participants.

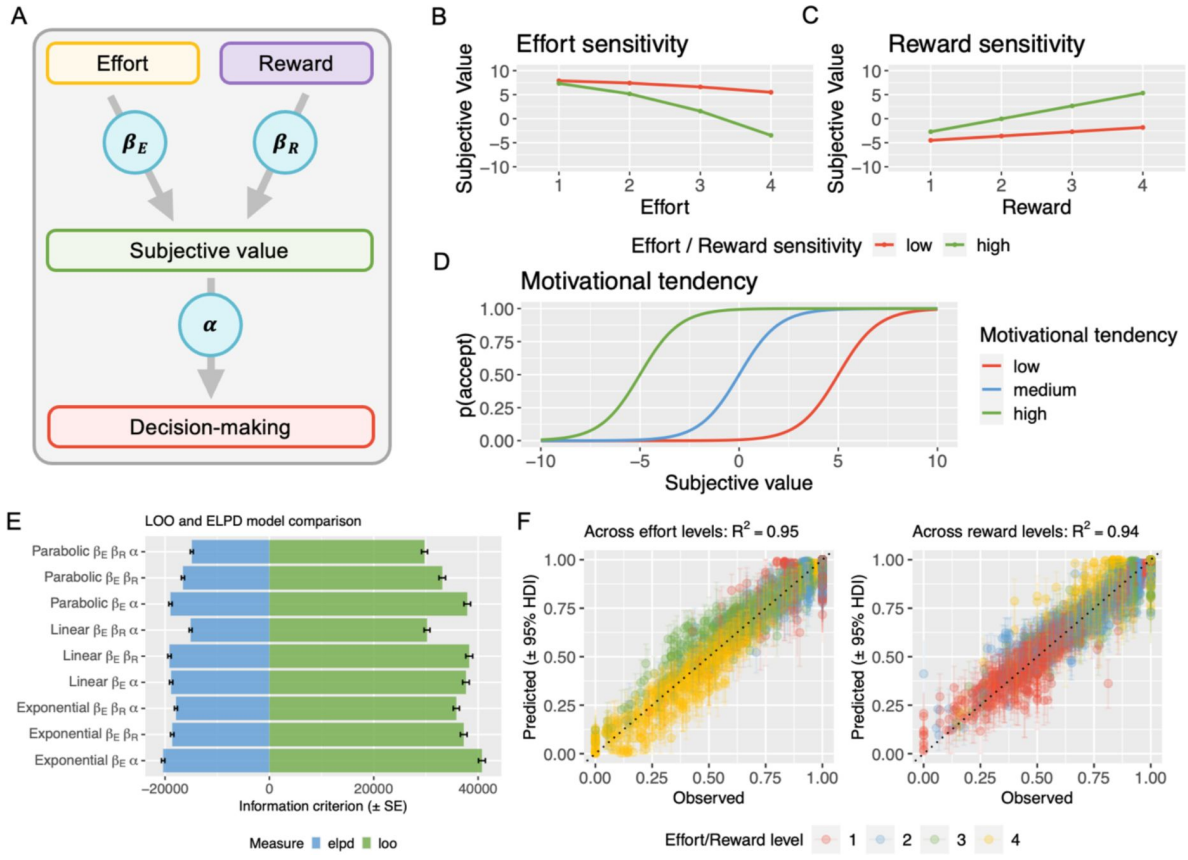


Figure 3.

Computational modelling: model visualization and model-based results.

A: Economic decision-making models posit that efforts and rewards are joined into a subjective value (SV), weighed by individual *effort* (β_E) and *reward sensitivity* (β_R) parameters. The SV is then integrated with a *motivational tendency* parameter and translated to decision-making. **B-C:** The model suggests that SV decreases as effort increases and increases as reward increases. The magnitude of this relationship depends on the individual effort and reward sensitivity parameters. **D:** The motivational tendency parameter acts as an intercept to the softmax function, thereby changing the relationship between SV and acceptance probability. **E:** Model comparison based on leave-out-out information criterion (LOOIC; lower is beger) and expected log posterior density (ELPD; higher is beger). **F:** Posterior predictive checks for the full parabolic model, comparing observed vs. model predicted subject-wise acceptance proportions across effort-levels (left) and reward-levels (right).

2.2.1 Replication of model-agnostic effects

The proportion of accepted trials for each effort-reward combination is ploded in **Figure 2C**. In line with our pre-registered hypotheses, we found significant main effects for effort ($F(1,14367)=4961.07$, $p<.0001$) and reward ($F(1,14367)=3037.91$, $p<.001$), and a significant interaction between the two ($F(1,14367)=1703.24$, $p<.001$). In post hoc ANOVAs, effort effects remained significant at all reward-levels (all $p<.001$) and reward effects remained significant at all effort-levels (all $p<.001$). The development of offered effort and reward levels across trials is shown in **Figure 2D**; this shows that as participants generally tend to accept challenges rather than reject them, the implemented staircasing procedure develops toward higher effort and lover reward challenges.

The mean success rate of accepted challenges across participants was high (M=98.7%) and varied lidle between participants (SD=3.50), indicating feasibility of all effort-levels across participants. Comparing clicking calibration results from pre-to post-task, the maximum clicking capacity decreased by 2.34 clicks on average (SD=14.5). Sixty-two (6.47%) participants reported having deviated from our instructions (i.e., changed the hand and/or finger used to make mouse clicks) throughout the game, but all effects could still be replicated in this subsample: main and interaction effects of effort and reward on the proportion of accepted trials could be replicated in this subsample (all p<.001) and there was no significant difference between participants that did or did not report finger switching in the mean percentage of accepted trials (switching: 79.51%, no switching: 76.62%; p=.149).

Subjects were engaged with the task, shown by a high rate of challenge acceptance (M=76.80%, SD=15.20, range=15.60-100%) and moderate-to-good enjoyment ratings (M=2.56, SD=0.92; on a 0-4 scale). Qualitative data of subjects describing their decision-making process during the task further confirmed high engagement (see Supplement 3).

2.2.2 Computational modelling

A model space of nine models was considered, varying in the implemented parameter and cost function (see Supplement 1.1 for mathematical definitions of all models). Prior to model fiding, parameter recovery confirmed all models yield meaningful parameter estimates (Supplement 1.2). All models showed good convergence (effective sample size (ESS)>4,223; R-hats<1.002 for all estimates). Model comparison by out-of-sample predictive accuracy identified the model implementing three parameters (motivational tendency α, reward sensitivity β_R, and effort sensitivity β_E), with a parabolic cost function (subsequently referred to as the *full parabolic model*) as the winning model (leave-one-out information criterion [LOOIC; lower is beder] = 29734.8; expected log posterior density [ELPD; higher is beder] = - 14867.4; **Fig. 3E**). This was in line with our pre-registered hypotheses. Predictive validity of the full parabolic model was validated with posterior predictive checks, showing excellent accordance between observed and model-predicted choice data (across effort-levels: R^2=.95, across reward-levels: R^2=.94; **Fig.3F**).

2.2.3 Test-retest reliability

We validated the task in a smaller in-person sample (N=30, tested twice ~7 days apart, holding time-of-day at testing constant) to assess test-retest reliability of parameter estimates, showing moderate to excellent reliability for all parameters (i.e., all intraclass correlation coefficients >0.4, all p<.01). Parameter estimates from modelling the data at one session predicted subjects' choices at the other session beder than chance and beder than group-level parameters predictions (all $p<0.01$)[72] (full details reported in Supplement 2).

2.3 Transdiagnostic analysis: Questionnaire measures predict effort-based decision-making

We used partial-least-squares (PLS) regression to relate individual-level mean posterior parameter values resulting from the model fiding of the full parabolic model to the questionnaire measures. To explore individual effects post-hoc, we followed up on effects found in the PLS regression using Bayesian generalised linear models (GLMs), controlling for age and gender.

2.3.1 Motivational tendency

Motivational tendency was best predicted by a model with one component, with its highest factor loadings from psychiatric measures (increasing values indicate symptom severity; SHAPS[71]: -0.665; AES[70]: -0.588; Dimensional Anhedonia Rating Scale [DARS][73]: -0.487). Weaker loadings were found for circadian measures (higher values indicate later chronotype; Mornigness-Eveningness Questionnaire [MEQ][11]: - 0.262; Munich Chronotype Questionnaire [MCTQ][74]: -0.117) and metabolic measures (higher values indicate higher metabolic risk; body mass index [BMI]: -0.115; Finnish Type-2 Diabetes Risk Score questionnaire [FINDRISC][75]: -0.068). Permutation testing indicated the predictive value of the resulting component (with factor loadings described above) was significant out-of-sample (root-mean-squared error [RMSE]=0.203, p=.001).

Bayesian GLMs confirmed evidence for psychiatric questionnaire measures predicting motivational tendency (SHAPS: M=-0.109; 95% highest density interval (*HDI*)=[-0.17,-0.04]; AES: M=-0.096; 95%HDI=[-0.15,-0.03]; DARS: M=-0.061; 95%HDI=[-0.13,-0.01]; **Fig. 4A**). Post-hoc GLMs on DARS sub-scales showed an effect for the sensory subscale (M=-0.050; 95%HDI=[-0.10,-0.01]). This result of neuropsychiatric symptoms predicting a lower motivational tendency is in line with our pre-registered hypothesis. For the MEQ (95%HDI=[-0.09,0.06]), MCTQ (95%HDI=[-0.17,0.05]), BMI (95%HDI=[-0.19,0.01]), and FINDRISC (95%HDI=[- 0.09,0.03]) no relationship with motivational tendency was found, consistent with the smaller magnitude of reported component loadings from the PLS regression. This null finding for dimensional measures of circadian rhythm and metabolic health was not in line with our pre-registered hypotheses.

2.3.2 Effort sensitivity

For effort sensitivity, the intercept-only model outperformed models incorporating questionnaire predictors based on RMSE.

2.3.3 Reward sensitivity

For reward sensitivity, the intercept-only model outperformed models incorporating questionnaire predictors based on RMSE. This result was not in line with our pre-registered expectations.

2.3.4 Questionnaire measures predict model agnostic task measures

Both SHAPS (M=-0.07; 95%HDI=[-0.12,-0.03]) and AES (M=-0.05; 95%HDI=[- 0.10,-0.002]) sum scores could predict the proportion of accepted trials averaged across effort and reward levels (**Fig. S4**).

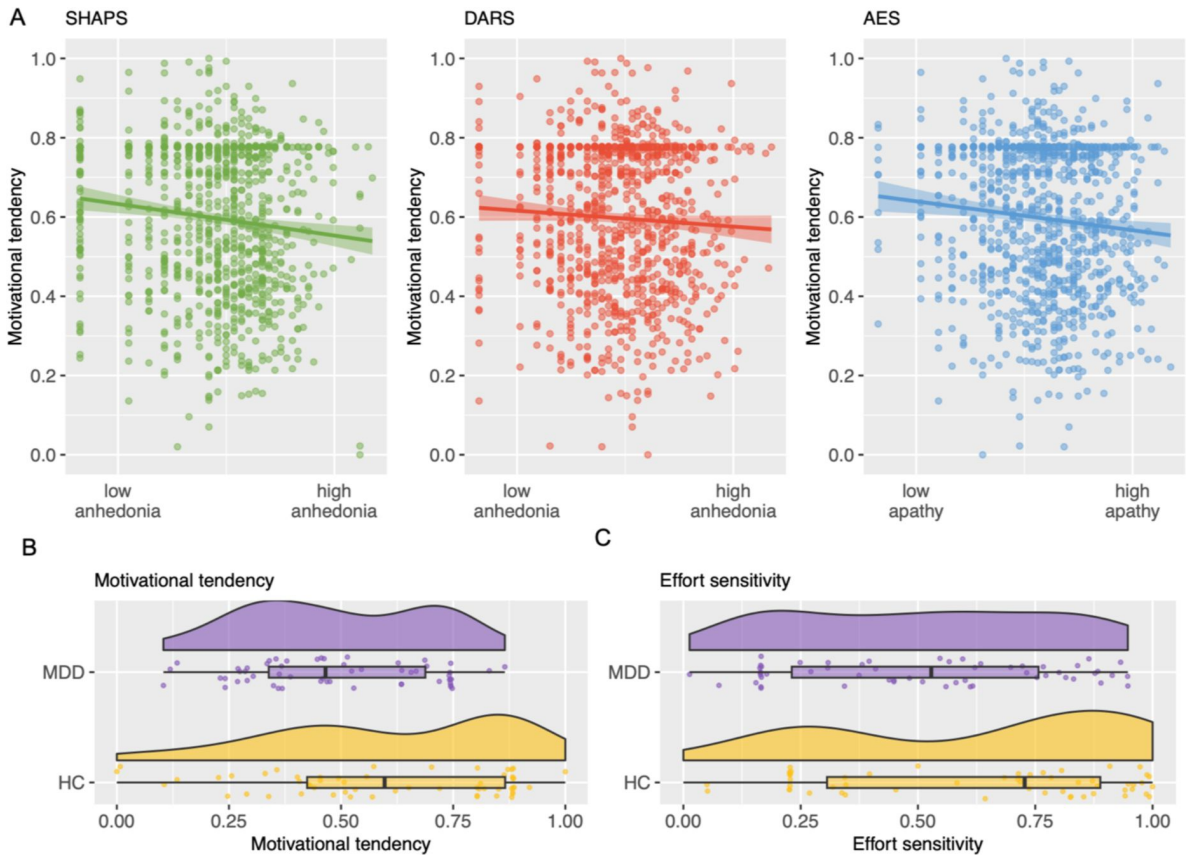


Figure 4.

Associations between task parameter estimates and psychiatric measures.

A: Visualizations of associations between the motivational tendency task parameter and the Snaith-Hamilton Pleasure Scale (SHAPS), the Dimensional Anhedonia Rating Scale (DARS)[73], and the Apathy Evaluation Scale (AES)[70]. **B-C:** Comparison of motivational tendency (left) and effort sensitivity (right) between a sample of participants meeting criteria for current major depressive disorder (MDD; purple, upper) on the the Mini-International Neuropsychiatric Interview 7.0.1 (M.I.N.I)[76] and age- and gender-matched controls (yellow, lower).

2.4 Diagnostic analysis: Depressed and healthy subjects differ in effort-based decision-making

In an exploratory analysis, we compared a sample of *N*=56 participants that met criteria for current major depressive disorder (MDD), to fifty-six healthy controls (HC), matched by age (MDD: *M*=37.07; HC: M=37.09, p=.99) and gender (MDD: 31 female, 23 male, 2 non-binary; HC: 32 female, 22 male, 2 non-binary; p=.98). Effort-discounting effects were confirmed in both groups. For both groups, model fiding and comparison identified the full parabolic model as the best-fiding model. We used age- and gender-controlled Bayesian GLMs to compare individual-level mean posterior parameter values between groups.

2.4.1 Motivational tendency

As in our transdiagnostic analyses of continuous neuropsychiatric measures (Results 2.3), we found evidence for a lower motivational tendency parameter in the MDD group compared to HCs (*M*=-0.111, 95% HDI=[-0.20,-0.03]) (**Fig. 4B**). This result confirmed our pre-registered hypothesis.

2.4.2 Effort sensitivity

Unlike our transdiagnostic analyses, we also found evidence for lower effort sensitivity in the MDD group compared to HCs (*M*=-0.111, 95% HDI=[-0.22,-0.01]) (**Fig. 4C**).

2.4.3 Reward sensitivity

There was no evidence for a group difference in reward sensitivity (95%HDI=[- 0.07,0.11]), as in our transdiagnostic analyses.

2.5 Circadian measures affect effort-based decision-making

Due to our hypothesised interaction between circadian preference and time-of-day, testing was conducted in two specified time windows: morning (08:00-11:59) and evening (18:00-21:59), resulting in a binary time-of-day measure (morning vs. evening testing). A total of 492 participants completed the study in the morning testing window and 458 in the evening testing window. We used the two chronotype questionnaires to identify two established circadian phenotypes: “early” or “late” chronotype (see Methods 4.5), behavioural categories indicating underlying chronobiological differences[9,11,74]. These classifications result in four sub-sample groups, with 89 early chronotypes (morning testing: *n*=63; evening testing: *n*=26) and 75 late chronotypes (morning testing: *n*=20; evening testing: *n*=55).

Bayesian GLMs, controlling for age and gender, predicting task parameters by time-of-day and chronotype showed effects of chronotype on reward sensitivity (i.e. those with a late chronotype had a higher reward sensitivity; *M*= 0.325, 95% HDI=[0.19,0.46]) and motivational tendency (higher motivational tendency in early chronotypes; *M*=-0.248, 95% HDI=[-0.37,-0.11]), as well as an interaction between chronotype and time-of-day on motivational tendency (*M*=0.309, 95% HDI=[0.15,0.48]).

2.5.1 Additional pre-registered data collection

As these analyses rely on unevenly distributed sub-samples, we conducted an additional, pre-registered data collection to replicate and extend these findings (hdps://osf.io/y4ke). We screened participants for their chronotype and then invited early chronotypes to take part in our study in the evening testing window, and late chronotypes in the morning testing window (Methods 4.5.1).

Using our pre-registered Bayesian stopping rule, we tested 13 early chronotype participants and 20 late chronotype participants. The data was then combined with the data from our main data collection, resulting in a full sample of *n*=197 participants that was used for subsequent chronotype analyses (see **Table 2** for sample characteristics and statistical significance of differences).

2.5.2 Motivational tendency

Late chronotypes showed a lower motivational tendency than early chronotypes (*M*=-0.11, *95% HDI*=[-0.22,-0.02])—comparable to effects of transdiagnostic measures of apathy and anhedonia, as well as diagnostic criteria for depression. Crucially, we found motivational tendency was modulated by an interaction between chronotype and time-of-day (*M*=0.19, *95% HDI*=[0.05,0.33]): post-hoc GLMs in each chronotype group showed this was driven by a time-of-day effect within late, rather than early, chronotype participants (*M*=0.12, *95% HDI*=[0.02,0.22], such that late chronotype participants showed a lower motivational tendency in the morning testing sessions, and a higher motivational tendency in the evening testing sessions; early chronotype: *95% HDI*= [-0.16,0.04]) (**Fig. 5A**). These results of a main effect and an interaction effect of chronotype on motivational tendency confirmed our pre-registered hypothesis.

2.5.2.1 Neuropsychiatric symptoms and circadian measures have separable effects on motivational tendency

Exploratory analyses testing for the effects of neuropsychiatric questionnaires on motivational tendency in the subsamples of early and late chronotypes confirmed the predictive value of the SHAPS (*M*=-0.24, *95% HDI*=[-0.42,-0.06]), the DARS (*M*=- 0.16, *95% HDI*=[-0.31,-0.01]), and the AES (*M*=-0.18, *95% HDI*=[-0.32,-0.02]) on motivational tendency.

For the SHAPS, we find that when adding the measures of chronotype and time-of-day back into the GLMs, the main effect of the SHAPS (*M*=-0.26, *95% HDI*=[- 0.43,-0.07]), the main effect of chronotype (*M*=-0.11, *95% HDI*=[-0.22,-0.01]), and the interaction effect of chronotype and time-of-day (*M*=0.20, *95% HDI*=[0.07,0.34]) on motivational tendency remain. Model comparison by LOOIC reveals motivational tendency is best predicted by the model including the SHAPS, chronotype and time- of-day as predictors, followed by the model including only the SHAPS. Note that this approach to model comparison penalizes models for increasing complexity.

Repeating these steps with the DARS, the main effect of the DARS is found numerically, but the 95% HDI just includes 0 (*M*=-0.15, *95% HDI*=[-0.30,0.002]). The main effect of chronotype (*M*=-0.11, *95% HDI*=[-0.21,-0.01]), and the interaction effect of chronotype and time-of-day (*M*=0.18, *95% HDI*= [0.05,0.33]) on motivational tendency remain. Model comparison identifies the model including the DARS and circadian measures as the best model, followed by the model including only the DARS.

For the AES, the main effect of the AES is found (*M*=-0.19, *95% HDI*=[-0.35,- 0.04]). For the main effect of chronotype, the 95% narrowly includes 0 (*M*=-0.10, *95% HDI*=[-0.21,0.002]), while the interaction effect of chronotype and time-of-day (*M*=0.20, *95% HDI*=[0.07,0.34]) on motivational tendency remains. Model comparison identifies the model including the AES and circadian measures as the best model, followed by the model including only the AES.

2.5.3 Effort sensitivity

We found no evidence for circadian or time-of-day effects on effort sensitivity (chronotype main effect: 95%HDI=[-0.06,0.18], time-of-day main effect: 95%HDI=[- 0.08,0.13]).

	Early chronotype	Late chronotype	Significance
Sample size (%)	102 (51.78%)	95 (48.22%)	
Demographics			
Age, mean (SD; range)	51.80 (14.10; 20-78)	35.80 (14.40; 19-68)	$p < .001$
Gender, number (%)			$p < .05$
Male	42 (41.18)	55 (57.89)	
Female	60 (58.82)	40 (42.11)	
Testing time			
Start testing time, number (%)			$p < .01$
Morning testing (8:00-11:59)	63 (31.98)	38 (19.29)	
Evening testing (18:00-21:59)	39 (19.80)	57 (28.93)	
Psychiatric comorbidities			
Current or past, number (%)			
Any	22 (21.60)	40 (42.10)	$p < .01$
Major depressive disorder	4 (3.92)	22 (23.16)	$p < .001$
Generalised or social anxiety disorder	18 (17.65)	24 (25.26)	$p = .258$
Current antidepressant use, number (%)	9 (8.82)	26 (27.40)	$p < 0.1$
Psychiatric questionnaire measures			
SHAPS, mean (SD; range)	9.65 (6.38)	11.80 (5.92)	$p < .05$
DARS, mean (SD; range)	54.00 (9.37)	52.70 (9.39)	$p = .322$
AES, mean (SD; range)	56.00 (9.72)	50.60 (10.10)	$p < .001$
M.I.N.I., current MDD (%)	3 (2.94)	15 (15.79)	$p < .01$

Table 2

Demographic characteristics and descriptive questionnaire measures in the early and late chronotype participants.

2.5.4 Reward sensitivity

Participants with an early chronotype had a lower reward sensitivity parameter than those with a late chronotype (M=0.27, 95% HDI=[0.16,0.38]). We found no effect of time-of-day on reward sensitivity (95%HDI=[-0.09,0.11]) (**Fig. 5B**). These results were in line with our pre-registered hypotheses.

3 Discussion

Various neuropsychiatric disorders are marked by disruptions in circadian rhythm, such as a late chronotype. However, research has rarely investigated how transdiagnostic mechanisms underlying neuropsychiatric conditions may relate to inter-individual differences in circadian rhythm. Here, combining a large-scale online study with computational modelling, we replicate and extend previous work linking anhedonia, apathy, and depression to a lower motivational tendency to exert effort for reward. Crucially, we found participants with a late compared to early chronotype show the same decrease in motivational tendency. Moreover, by testing participants at chronotype-compatible and -incompatible times of day, we discovered late chronotypes show a decreased motivational tendency to exert effort for reward when tested in the morning compared to evening. This reveals neuropsychiatric symptoms and chronotype (interacting with time-of-testing) show paralleling effects on effort-based decision-making. Our results demonstrate a crucial role for circadian rhythm in computational psychiatry, potentially affecting our assessment and treatment of neurocognitive mechanisms.

We replicate and extend effects of aberrant effort-based decision-making in neuropsychiatric syndromes in a large, broadly population-representative sample. Our finding that dimensional measures of apathy and anhedonia predict motivational tendency, a computational parameter describing someone's tendency to exert effort for reward, aligns with previous reports of impaired effort-based decision-making in psychiatric[31,77–85] and neurodegenerative populations[86–90], as well as studies linking effort-based decision-making with apathy and anhedonia specifically. The positive link between effort-based decision-making and apathy and anhedonia has been observed in both patients[77,84,91] and healthy controls[77,92,93], though some did not find this effect[85].

Our work supports previous theories that impaired effort-based decision-making represents a common, transdiagnostic mechanism across the psychiatric and neurological syndromes of anhedonia and apathy (respectively). We found corresponding effects of apathy and anhedonia on the same computational parameter — motivational tendency — reinforcing the suggestion of possible shared mechanistic underpinnings of the two motivational syndromes[24]. Aberrant effort-based decision-making may manifest behaviourally as deficient motivation, a symptom category that cuts across traditional disease boundaries of psychiatric, neurological, and neurodevelopmental disorders[32].

Our categorical (diagnostic-criteria based) analysis comparing depressed to healthy subjects likewise found depressed patients showed a lower motivational tendency, echoing our dimensional results in apathy and anhedonia. In addition, our categorical analysis revealed a distinct effect of group on effort sensitivity: depressed subjects had lower effort sensitivity, meaning their decisions were less influenced by effort changes. Possibly, this effect stems from decreased perceived differences in effort levels, as recently reported[94], indicating there are both dimensional (transdiagnostic) and potentially some diagnosis-specific effects of mental health on effort-based decision-making.

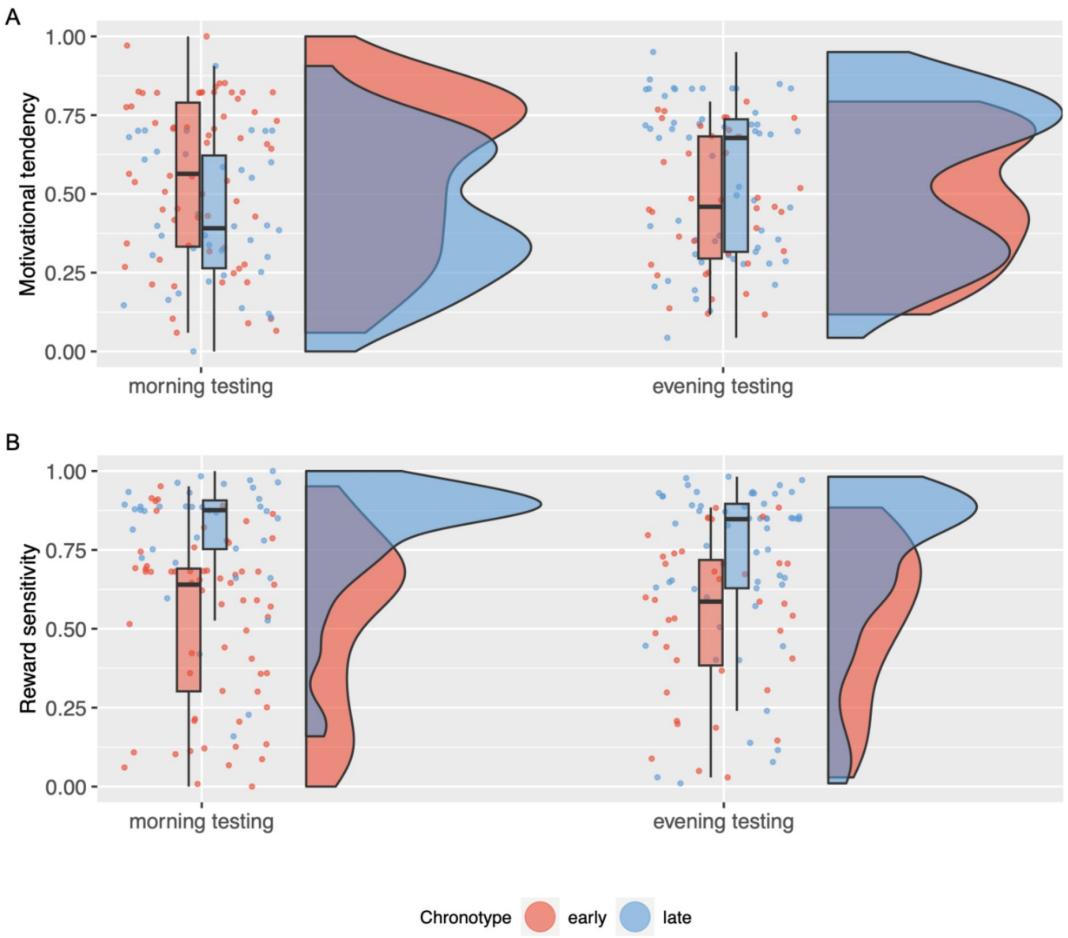


Figure 5.

Effects of chronotype and time-of-day on task parameter estimates.

A: Effect of chronotype and time-of-day on reward sensitivity parameter estimates. **B:** Effect of chronotype and time-of-day on motivational tendency parameter estimates.

It is possible that a higher motivational tendency reflects a more optimistic assessment of future task success, in line with work on the optimism bias[95]; however our task intentionally minimized unsuccessful trials by titrating effort and reward; future studies should explore this more directly.

We also found circadian effects on effort-based decision-making, paralleling those of apathy, anhedonia, and depression measures in both the affected neurocomputational parameter and the direction of effect. We observed a difference in motivational tendency between chronotypes, with late chronotypes showing a lower tendency to accept to exert effort for reward. Previous studies have suggested late chronotypes were also less accepting of delays[62] and risk for reward[64,65].

Most importantly, we found an interaction between chronotype and time-of-day in a synchrony effect manner: early and late chronotypes showed a *higher* motivational tendency towards accepting effort for reward *at their preferred time of day*. This effect was driven by the late chronotype group, who showed a markedly lower motivational tendency in the morning, but a higher tendency in the evening. This suggests that chronotype effects on neurocomputational parameters such as motivational tendency depend on time-of-testing. Synchrony effects have previously been observed in other cognitive domains including inhibitory control[96], adention[97], learning[98], and memory[99]. One interpretation of our cognitive synchrony effects may be that late chronotype participants show a diminished ability to adapt to suboptimal times-of-day due to reduced cognitive resources[100].

We also report a distinct effect of chronotype on effort-based decision-making that is not paralleled by effects of neuropsychiatric symptoms, nor dependent on time-of-day. Compared to early chronotypes, late chronotypes were more guided by differences in reward value, indicated by higher reward sensitivity parameters. Previous studies report altered reward functioning in late chronotypes, who show a reduced reactivity to reward in the medial prefrontal cortex, a key component of reward circuitry[101–103]. Note this is not incompatible with higher reward sensitivity due to our modelling approach, in which higher reward sensitivity does not imply higher reward valuation, but rather larger subjective value differences between reward levels. Therefore, reduced reactivity to reward could be compatible with late chronotypes devaluing low reward levels more, which in our models would emerge as a reduced reward sensitivity parameter.

It is striking that the effects of neuropsychiatric symptoms on effort-based decision-making largely are paralleled by circadian effects on the same neurocomputational parameter. Exploratory analyses predicting motivational tendency by neuropsychiatric symptoms and circadian measures simultaneously indicate the effects go beyond recapitulating each other, but rather explain separable parts of the variance in motivational tendency. Overall, our results raise the possibility of altered effort-reward processing as a critical mechanism linking neuropsychiatric conditions and circadian rhythm. Previous research demonstrated depressed patients with an evening chronotype show increased diurnal mood variation[104]. Our finding of time-of-day differences in motivational tendency among late chronotypes illustrates a potential cognitive underpinning for the observed diurnal characteristic within depression and late chronotype. Together, these findings support the idea of a circadian psychiatric phenotype[14,17], which should be considered in measurement (e.g., design of computational psychiatry studies) and potentially treatment (e.g., administration of motivation-based psychological interventions, which could be timed compatibly with chronotype).

To our surprise, we did not find statistical evidence for a relationship between effort-based decision-making and measures of metabolic health (BMI and risk for type-2 diabetes). Our analyses linking BMI to motivational tendency reveal a numeric effect in line with our hypothesis: a higher BMI relating to a lower motivational tendency. However, the 95% HDI for this effect

narrowly included zero (95%HDI=[-0.19,0.01]). Possibly, our sample did not have sufficient variance in metabolic health to detect dimensional metabolic effects in a current general population sample. A recent study by our group investigates the same neurocomputational parameters of effort-based decision-making in participants with type-2 diabetes and non-diabetic controls matched by age, gender, and physical activity[105]. We report a group effect on the motivational tendency parameter, with type-2 diabetic patients showing a lower tendency to exert effort for reward.

Our study results should be considered in light of a few limitations. First, we used online self-report measures of neuropsychiatric symptoms and depression status. There has been a large shift toward online data collection in psychiatric research, and while online data is undoubtedly noisier, results (including our own, presented in the supplemental material) usually show excellent accordance with labbased studies[106]. Similarly, we lack biological measures of circadian rhythm, the gold standard of chronotype assessment. However, this concern might be mitigated by previous reports of high covariance between biological- and questionnaire-based circadian measures[107,108], as well as significant chronobiological differences between the questionnaire-determined chronotypes[10,109] we use in our key findings. Nevertheless, future work should incorporate biological measures in adempts to replicate circadian effects on effort-based decision-making. This could take the form of identifying chronotypes by DNA analysis or dim-light melatonin onset, or continuous measurements of circadian proxies, such as core body temperature, heart rate, or actigraphy.

Note also that our time-of-day effects are limited by a between-subjects study design (i.e., the same participants were not tested in morning and evening sessions). It will be interesting to explore such diurnal variation in effort-based decision-making within individuals. The newly developed effort-expenditure task we present here may lend itself particularly well to such endeavours. First, it allows remote testing, meaning subjects can complete the task at different times of the day without in-person testing. Second, we demonstrated good test-retest reliability of task measures when time of testing was held constant within participants. This good test-retest reliability of our task contrasts with recent reports of poor test-retest reliability of other tasks and computational modelling parameters[110].

Our reported analyses investigating neuropsychiatric and circadian effects on effort-based decision-making simultaneously are exploratory, as our study design was not ideally set out to examine this. Further work is needed to disentangle *separable* effects of neuropsychiatric and circadian measures on effort-based decision-making. One approach could be a group-based study design enabling the dissociation of the two effects (e.g., examining high-anhedonia participants with early chronotypes and low-anhedonia patients with late chronotypes, as well as the respective other, more common groupings, and testing each group in the morning and evening to examine time-of-day interactions with both anhedonia and chronotype).

Finally, we would like to note that as our study is based on a general population sample, rather than a clinical one. Hence, we cannot speak to transdiagnosticity on the level of multiple *diagnostic categories.*

Taken together, our results implicate circadian rhythm as an important factor in effort-based decision-making and its relationship to neuropsychiatric conditions. These results have implications for research, clinical interventions, and policy. We demonstrate that neuropsychiatric effects on effort-based decision-making are paralleled by effects of circadian rhythm and time-of-day. Exploratory analyses suggest these effects account for separable parts of the variance in effort-based decision-making. It unlikely that effects of neuropsychiatric effects on effort-based decision-making reported here and in previous literature are a spurious result due to multicollinearity with chronotype. Yet, not accounting for chronotype and time of testing, which is the predominant practice in the field, could affect results. This could take the form of either

inflating or depressing results in the existing literature. On the one hand, reported neuropsychiatric effects may be inflated by systematic circadian differences between participants (i.e., overrepresentation of late chronotype in patient samples), which could be further amplified by time of testing (often the morning, incompatible with late chronotypes, and producing motivational impairments on neurocognitive measures). On the other hand, true effects may be masked by interactions between chronotype and time-of-day: Testing psychiatric subjects with a late chronotype in the evening (e.g., as a consequence of subject-selected testing times) may paint a false picture of group equivalence, as researchers are only observing part of a daily trajectory.

Our growing understanding of the relationship between circadian rhythm and neuropsychiatry may allow for critical advances in improving therapeutic outcomes from treatments[111,112]. Such advances are particularly called for in the case of symptoms of apathy and anhedonia, as current treatments often fail to improve motivational deficits[27–29], but could potentially be coupled with a patient's chronotype to increase efficacy. At minimum, clinical trials predicting change in motivational measures, such as effort-based decision-making, should assess patients at similar times of day, as this could reduce or inflate treatment effects.

Circadian rhythm and neuropsychiatric syndromes may affect motivation via parallel, as well as distinct, mechanisms—but crucially, this overlap is dependent on time of testing. Our work suggests that chronotype and time of testing are essential variables to consider in future effort-based decision-making experiments, particularly those measuring effort-based decision-making in patient groups, such as those with depression, high apathy, or high anhedonia. Beyond experimental work, future interventions should consider the role of chronotype in measurement and modulation of motivation.

4 Methods and materials

4.1 Study protocol

After providing demographics and basic medical history, subjects completed an effort-expenditure task, followed by a badery of self-report questionnaires. The study was coded in JavaScript, using Phaser v.3.50.0 for the task and jsPsych[113] for questionnaires.

4.1.1 Ethics

This study was approved by the University of Cambridge Human Biology Research Ethics Commidee (HBREC.2020.40). Participants provided informed consent through an online form, complying with the University of Cambridge Human Biology Research Ethics Commidee procedures for online studies.

4.1.2 Recruitment

We recruited participants using Prolific[114], in September 2022. Data was collected on weekdays, in specified daily time-windows (*morning testing*: 08:00-11:59; *evening testing*: 18:00- 21:59). To sample participants broadly representative of the UK population in age, sex, and history of psychiatric disorder, we implemented a previously-described procedure[67] using Prolific pre-screeners to obtain batches of participants aimed to match target numbers calculated based on UK population data.

Nine hundred and ninety-four participants completed all components and were paid a fixed rate of £6. A bonus of £10 was paid to ten participants. Subjects were told they could increase their chances of winning the bonus by engaging well with the study (e.g., reading questions carefully, following task instructions).

4.1.3 Effort-expenditure task

We developed a new effort-expenditure task that allowed us to assess effort-based decision-making in a remote seding; this task was also tested in-person to assess test-retest reliability. To increase engagement, we gamified the task to take place in an underwater seding and each challenge is framed as a race in which an octopus catches a shrimp. The task structure is shown in **Figure 2A** and the trial-level structure in **Figure 2B**.

The task began with an individual calibration phase to standardise maximum effort capacity, followed by the main task, which used a semi-adaptive staircase design to maximise the informative value of each choice.

For the calibration, subjects were prompted to collect points by clicking as fast as possible for ten seconds, repeated three times. The second and third repetitions were then averaged to serve as the maximum clicking capacity reference for the main task. One calibration trial was repeated at the end of the main task to monitor any notable changes in clicking capacity. Then, subjects were familiarized with their individually calibrated effort levels during a practice phase of the task. Effort levels were scaled to a given participant's mean clicking speed (based on the calibration phase), and the time clicking must be sustained for. We used four effort-levels, corresponding to a clicking speed at 30% of a participant's maximal capacity for 8 seconds (level 1), 50% for 11 seconds (level 2), 70% for 14 seconds (level 3), and 90% for 17 seconds (level 4). Therefore, in each trial, participants had to fulfil a certain number of mouse clicks (dependent on their capacity and the effort level) in a specific time (dependent on the effort level). Subjects were instructed to make mouse-clicks with the finger they normally use, and to not change fingers throughout the task (compliance was checked at the end of the main task). In the practice phase, all effort-levels were completed without reward associations, and failed levels were repeated until subjects succeed at each level. If a subject failed a level twice, the clicking capacity reference was adjusted to the speed reached in the practice. Finally, subjects needed to pass a six-question quiz to ensure task instructions were fully understood. If a subject failed any question on the quiz, they were returned to the instruction screens and re-took the quiz until all questions were answered correctly.

The main task took a binary-choice design: In each trial, participants accepted or rejected a challenge associated with one of the four specific effort levels and rewards. Reward was conceptualized as points (shrimp caught by the octopus) that could be collected in that trial. The points to win per challenge varied between four levels (2,3,4, or 5 points). If a subject accepted a given challenge, they needed to achieve the given effort-level to win the associated points. If a subject rejected a given challenge, they waited and received one point, with waiting times matched to the respective effort level to prevent confounding with delay discounting. Participants were able to infer their clicking progress from the distance between the octopus and the shrimp and the remaining time was indicated by a time-bar.

Subjects completed 64 trials, split into four blocks of 16 trials. For each subject, trial-by-trial presentation of effort-reward combinations were made semi-adaptively by 16 randomly interleaved staircases. Each of the 16 possible offers (4 effort-levels x 4 reward-levels) served as the starting point of one of the 16 staircase. Within each staircase, after a subject accepted a challenge, the next trial's offer on that staircase was adjusted (by increasing effort or decreasing reward). After a subject rejected a challenge, the next offer on that staircase was adjusted by decreasing effort or increasing reward. This ensured subjects received each effort-reward combination at least once (as each participant completed all 16 staircases), while individualizing trial presentation to maximize the trials' informative value. Therefore, in practice, even in the case of a subject rejecing all offers (and hence the staircasing procedures always adapting by

decreasing effort or increasing reward), the full range of effort-reward combinations will be represented in the task across the startingpoints of all staircases (and therefore before adaption takeplace).

4.1.4 Self-report questionnaires

Subjects completed a questionnaire badery assessing mental and physical health, presented in a randomised order. We assessed anhedonia using the SHAPS[71], as well as the DARS[73]. Apathy was assessed with the AES[70]. Additionally, we screened participants for meeting diagnostic criteria for current, past, or recurrent MDD using the M.I.N.I.[76]. Two questionnaires targeted circadian rhythm: the MEQ[11] and the MCTQ[74]. Metabolic health was assessed by collecting self-reported height and weight, used to calculate BMI. Additionally, the FINDRISC[75] was used to calculate individual risk scores for metabolic disease. Finally, the International Physical Activity Questionnaire (IPAQ)[116] was included for exploratory investigations of physical activity.

4.1.5 Compliance checks and exclusion criteria

All exclusion criteria were preregistered. Participants were excluded when reporting a severe neurological condition (*n=14*) or English proficiency below B2 (i.e., good command/working knowledge; *n=2*).

To check compliance with the questionnaires, four catch questions were presented during questionnaires, including two easy questions (e.g., “Please answer ‘Not at all’.”) and two harder questions (e.g., “In the past week, I (would have) wanted to eat mouldy food.”, expected answer “Disagree” or “Definitely disagree”). Participants failing at least one easy question, or both harder questions were excluded (*n=12*).

As task-based exclusion criteria, subjects rejecting all offers were excluded (*n*=0). Participants had to have a clicking-calibration score of at least seven, as values below would lead to challenges with just one mouse-click (*n=4*). Subjects showing a large difference between minimum and maximum clicking speed (i.e., >3 standard deviations (SD)) during calibration trials were excluded, as a misestimation of the calibration reference is likely (*n=3*). Finally, subjects showing a large change in their clicking capacity (i.e., >3 SD) pre- to post-task were excluded, as it can be assumed the applied calibration was not valid during the task (*n=1*). We also asked two open-answer questions after completion of the main task to monitor participants’ self-reported task strategies as a way of assessing rule adherence.

4.2 Analyses of effort-expenditure task data

4.2.1 Model-agnostic analyses

Using the proportion of accepted challenges as the dependent variable, we investigated main effects of effort- and reward-levels and their interaction, using a repeated measures analysis of variances (ANOVA) of repeated measures. This approach accommodates the unbalanced design resulting from the implemented staircasing procedure.

4.2.2 Model-based analyses

4.2.2.1 Model space

To model effort-based decision-making, we considered a model space of nine models. All models are variations of the economic decision-theory model, consisting of two basic equations. First, a cost function transforms costs and rewards associated with an action into a subjective value (SV):

$$SV = (\beta_R \cdot R) - (\beta_E \cdot E) \quad (1)$$

with β_R and β_E for reward and effort sensitivity, and) and + for reward and effort. Higher effort and reward sensitivity mean the SV is more strongly influenced by changes in effort and reward, respectively (**Fig. 3B-C**). Hence, low effort and reward sensitivity mean the SV, and with that decision-making, is less guided by effort and reward offers, as would be in random decision-making.

This SV is then transformed to an acceptance probability by a softmax function:

$$p(accept) = \frac{1}{1 + e^{-(\alpha + SV)}} \quad (2)$$

with $p(accept)$ for the predicted acceptance probability and 3 for the intercept representing motivational tendency. A high motivational tendency means a subjects has a tendency, or bias, to accept rather than reject offers (**Fig. 3D**).

The models differed in two aspects. First, inclusion or exclusion of the free parameters reward sensitivity (β_R) and motivational tendency (α). Second, the form of the cost function, which used either a linear function (proportional discounting at all effort-levels), a parabolic function (increases at higher effort-levels are discounted over-proportionally) or an exponential function (increases at lower effort-lower levels are discounted over-proportionally). See supplement 1.1 for mathematical definitions of all models.

4.2.2.2 Model fitting, checks, and comparisons

We took a hierarchical Bayesian approach to model fiding[117], implemented with the CmdStan R interface[118], with Stan code adapted from hBayesDM[119]. Prior to model fiding, effort- and reward-levels were standardized for computational ease. All models were fit using Markov-Chain Monte Carlo (MCMC), with 2000 warm-up iterations and 6000 sampling iterations, by four chains. Model convergence and chain mixing were checked using numerical diagnostics of ESS and split R-hats, and by visually inspecting trace plots. We conducted parameter recoveries for all models, confirming their ability to meaningfully recover known parameters. Model performance was compared based on out-of-sample predictive accuracy using the LOOIC (lower is beder) and ELPD (higher is beder). The winning model was validated using posterior predictive checks, comparing model predictions to subjectwise observed choices.

4.2.2.3 Test-retest reliability

We conducted an in-person study to validate the effort-expenditure task and assess the test-retest reliability of our computational modelling parameters. A sample of N=30 participants was recruited and tested in two sessions, about one week apart. Test-retest reliability of task parameters was assessed by intra-class correlation coefficients, Pearson's correlation coefficients (estimated both after model fiding and by embedding a correlation matrix into the model fiding procedure), and by testing the predictive accuracy of parameter estimates across sessions. See supplement 2 for full methods and results.

4.3 Linking model parameters to outcome measures

To aid interpretability and comparability of effects, task parameters and questionnaire outcome measures were standardized to be between zero and one. Questionnaire measures resulting from the DARS, AES, and MEQ were additionally transformed to be interpretable with the same directionality within questionnaire groupings (i.e., for all psychiatric measures higher values are interpreted as higher symptom severity, for all circadian measures higher values are interpreted as later chronotype).

To investigate associations between effort-based decision-making and self-report questionnaires, we ran partial least squares (PLS) regressions with questionnaire outcome measures predicting modelling parameters. PLS regression allows joint modelling of questionnaire measures, without issues due to expected multicollinearity between questionnaires. Following the best practice of model validation[120], data was split into a training (75%) and a testing (25%) subset. The training data was used to obtain the optimal number of components, based on ten-fold-cross validation, and to train the model. The winning model's predictive performance was tested out-of-sample using the held-out testing data. Statistical significance of obtained effects (i.e., the predictive accuracy of the identified component and factor loadings) was assessed by permutation tests, probing the proportion of root-mean-squared errors (RMSEs) indicating stronger or equally strong predictive accuracy under the null hypothesis.

To follow up on relationships suggested by the PLS regression, we performed Bayesian general linear models (GLMs), adjusting for age and gender (male or female, imputing natal sex for non-binary participants, given low numbers).

4.4 Comparing depressed and healthy subjects

We compared participants meeting criteria for a current MDD based on the M.I.N.I.[76], to a subset of age- and gender-matched healthy controls (HCs, participants that did not meet criteria for current MDD). For computational sparsity, we only fit the three best-fiding models from the full sample. Models were fit separately to the MDD and HC groups, using the same methods and parameters described above. Bayesian GLMs were used to quantify evidence for associations between individual-level modelling parameters and group status. As we could not be certain whether we would obtain a large enough sample size of subjects meeting criteria for MDD, these analyses were exploratory.

4.5 Investigating circadian effects

We used the two circadian rhythm questionnaires to determine participants' chronotypes. Early chronotype was defined as meeting criteria for "morning types" on the MEQ (MEQ sum score > 58)[11] and having a midpoint of sleep on free days before 02:30 . Late chronotype was defined as meeting criteria for "evening types" on the MEQ (MEQ sum score < 42) and having a midpoint of sleep on free days after 05:30. Subjects not falling into either category are categorized as intermediate chronotypes and were not included in these analyses.

We used Bayesian age- and gender-controlled GLMs to investigate effects of chronotype, time-of-day (morning- vs. evening-testing), and their interaction on subject-wise mean task parameters estimates.

4.5.1 Additional data collection

To improve the precision of estimated circadian effects on task parameters, we increased our sample size by conducting an additional pre-registered data collection (hdps://osf.io/y4ke). We implemented a screening study comprising the MEQ[11] and MCTQ[74]. Taking the chronotyping

approach described above, subjects with an early or late chronotype were identified. Early chronotypes were invited to take part in our study in the evening, late chronotypes in the morning.

We implemented a Bayesian stopping rule to inform our data collection process, taking the following steps. First, participants were screened in batches of 250, and eligible participants were invited to the study session. Next, data resulting from this additional data collection was joined with data resulting from the main data collection and Bayesian GLMs were re-run, as described above. If our precision target of any 95% HDI reaching a maximum width 0.20 was met, we stopped data collection. Was the precision target not met, we returned to step one, and another batch of 250 participants was screened. In any case, data collection would be terminated once 200 eligible participants had completed the main study session.

4.5.2 Differentiating between the effects of neuropsychiatric symptoms and circadian measures on motivational tendency

To investigate how the effects of neuropsychiatric symptoms on motivational tendency (2.3.1) relate to effects of chronotype and time-of-day on motivational tendency we conducted exploratory analyses. In the subsamples of participants with an early or late chronotype (including additionally collected data), we first ran Bayesian GLMs with neuropsychiatric questionnaire scores (SHAPS, DARS, AES respectively) predicting motivational tendency, controlling for age and gender. We next added an interaction term of chronotype and time-of-day into the GLMs, testing how this changes previously observed neuropsychiatric and circadian effects on motivational tendency. Finally, we conducted a model comparison using LOO, comparing between motivational tendency predicted by a neuropsychiatric questionnaire, motivational tendency predicted by chronotype and time-of-day, and motivational tendency predicted by a neuropsychiatric questionnaire and time-of-day (for each neuropsychiatric questionnaire, and controlling for age and gender).

Funding

This study was funded by an AXA Research Fund Fellowship awarded to C.L.N. (G102329) and the Medical Research Council (MC_UU_00030/12). C.L.N. is funded by a Wellcome Career Development Award (226490/Z/22/Z). This research was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215- 20014).

Author contributions

Conceptualization: S.Z.M, C.L.N; Methodology: S.Z.M, C.L.N; Investigation: S.Z.M, C.L.N; Project administration: S.Z.M, C.L.N; Writing (original draft, review & editing): S.Z.M, C.L.N; Formal analysis: S.Z.M; Funding Acquisition: C.L.N; Supervision: C.L.N.

Additional information

Supplementary material

1 Computational modelling

1.1 Mathematical definition of the model space

Model	Cost function	Softmax function
Linear model 1	$SV = (R) - (\boldsymbol{\beta_E} \cdot E)$	$p(accept) = \frac{1}{1 + e^{-(\boldsymbol{\alpha} + SV)}}$
Linear model 2	$SV = (\boldsymbol{\beta_R} \cdot R) - (\boldsymbol{\beta_E} \cdot E)$	$p(accept) = \frac{1}{1 + e^{-SV}}$
Linear model 3	$SV = (\boldsymbol{\beta_R} \cdot R) - (\boldsymbol{\beta_E} \cdot E)$	$p(accept) = \frac{1}{1 + e^{-(\boldsymbol{\alpha} + SV)}}$
Parabolic model 1	$SV = (R) - (\boldsymbol{\beta_E} \cdot E^2)$	$p(accept) = \frac{1}{1 + e^{-(\boldsymbol{\alpha} + SV)}}$
Parabolic model 2	$SV = (\boldsymbol{\beta_R} \cdot R) - (\boldsymbol{\beta_E} \cdot E^2)$	$p(accept) = \frac{1}{1 + e^{-SV}}$
Parabolic model 3	$SV = (\boldsymbol{\beta_R} \cdot R) - (\boldsymbol{\beta_E} \cdot E^2)$	$p(accept) = \frac{1}{1 + e^{-(\boldsymbol{\alpha} + SV)}}$
Exponential model 1	$SV = (R) \cdot e^{(-\boldsymbol{\beta_E} \cdot E)}$	$p(accept) = \frac{1}{1 + e^{-5 \cdot (\boldsymbol{\alpha} + SV)}}$
Exponential model 2	$SV = (\boldsymbol{\beta_R} \cdot R) \cdot e^{(-\boldsymbol{\beta_E} \cdot E)}$	$p(accept) = \frac{1}{1 + e^{-5 \cdot (-0.5 + SV)}}$
Exponential model 3	$SV = (\boldsymbol{\beta_R} \cdot R) \cdot e^{(-\boldsymbol{\beta_E} \cdot E)}$	$p(accept) = \frac{1}{1 + e^{-5 \cdot (\boldsymbol{\alpha} + SV)}}$

Table S1

Mathematical definition of the models included in our model space.

1.2 Model validation

Parameter recoveries were performed to ensure parameter estimates obtained from all models are meaningful. For each model, sets of parameter values were sampled from uniform distributions bound to the respective parameter ranges. Task data was then simulated for *n*=500 agents, using the respective parameters and modelling equations. The resulting simulated data was used for model fitting and resulting "recovered" posterior parameter estimates compared to the underlying parameter values. For all models, underlying parameters correlated highly with the recovered mean parameter estimates (**Table S2**). For the winning model (full parabolic model), relations between underlying and recovered parameters are additionally visualized in **Figure S1A-C**. Importantly, the modelling procedure did not introduce any spurious correlations between free parameters (**Figure S1D**).

Table S2

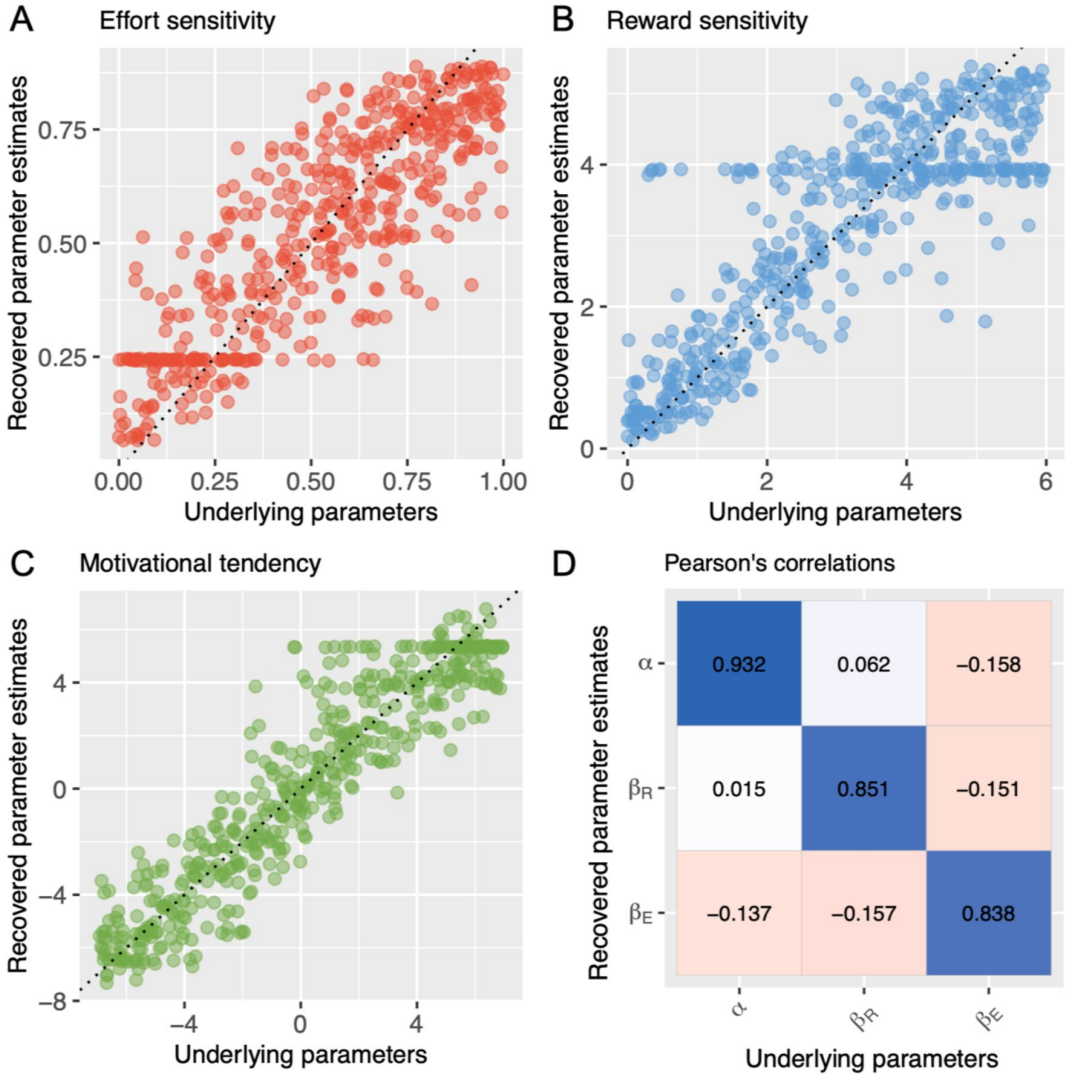
Pearson's correlations between underlying parameters and recovered mean parameter estimates for all models included in the model space.

	Linear models			Parabolic models			Exponential models		
	1	2	3	1	2	3	1	2	3
β_E	.811	.927	.816	.836	.919	.838	.926	.882	.802
β_R	-	.919	.841	-	.928	.851	-	.898	.891
α	.965	-	.932	.978	-	.932	.950	-	.904

Figure S1.

Parameter recovery.

A-C: Comparison between underlying parameters and recovered mean parameter estimates for the three free parameters of the full parabolic model. **D:** Pearson's correlations between all underlying and recovered parameters for the full parabolic model.



1.2.1 Parameter recoveries including inverse temperature

In the process of task and model space development, we also considered models incorportating an inverse temperature paramater. To this end, we conducted parameter recoveries for four models, defined in **Table S3**.

Parameter recoveries indicated that, parameters can be recovered reliably in model 1, which includes only effort sensitivity (β_E) and inverse temperature (τ) as free parameters (on-diagonal correlations: .98 > r > .89, off-diagonal correlations: .04 > |r| > .004). However, as a reward sensitivity parameter is added to the model (model 2), parameter recovery seems to be compromised, as parameters are estimated less accurately (on-diagonal correlations: .80 > r > .68), and spurious correlations between parameters emerge (off-diagonal correlations: .40 > |r| > .17). This issue remains when motivational tendency is added to the model (model 4; on-diagonal correlations: .90 > r > .65; off-diagonal correlations: .28 > |r| > .03), but not when inverse temperature is modelled with effort sensitivity and motivational tendency, but not reward sensitivity (model 3; on-diagonal correlations: .96 > r > .73; off-diagonal correlations: .05 > |r| > .003).

As our pre-registered hypotheses related to the reward sensitivity parameter, we opted to include models with the reward sensitivity parameter rather than the inverse temperature parameter in our model space.

Model	Cost function	Softmax function
1	$SV = (R) - (\boldsymbol{\beta_E} \cdot E^2)$	$p(accept) = \frac{1}{1 + e^{-\tau(SV)}}$
2	$SV = (\boldsymbol{\beta_R} \cdot R) - (\boldsymbol{\beta_E} \cdot E^2)$	$p(accept) = \frac{1}{1 + e^{-\tau(SV)}}$
3	$SV = (R) - (\boldsymbol{\beta_E} \cdot E^2)$	$p(accept) = \frac{1}{1 + e^{-\tau(\alpha + SV)}}$
4	$SV = (\boldsymbol{\beta_R} \cdot R) - (\boldsymbol{\beta_E} \cdot E^2)$	$p(accept) = \frac{1}{1 + e^{-\tau(\alpha + SV)}}$

Table S3

Mathematical definition of models included an inverse temperature parameter.

1.3 Group and subject-wise parameter estimates

As described in the main manuscript, we took a hierarchical Bayesian approach to model fitting, using Markov-Chain Monte Carlo (MCMC) sampling. Hence, we obtained posterior parameter distributions for each parameter, on both the subject and group level. Analyses presented in the main manuscript are based on the mean for each subject-wise parameter estimate distribution. The resulting parameter estimates - both subject-wise and on a group-level - are visualised in **figure S2**.

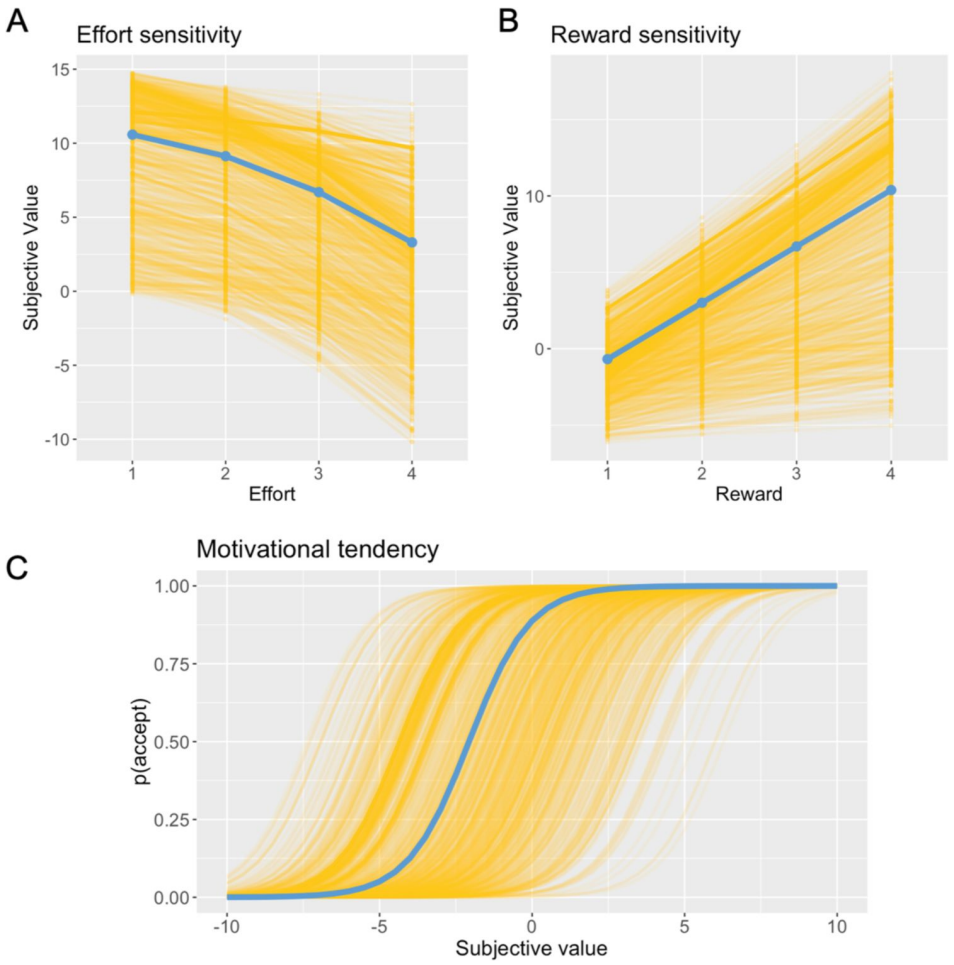


Figure S2.

Parameter estimates.

A-C: Visualisation of individual-level (yellow) and group-level (blue) model parameter estimates for effort sensitivity (A), reward sensitivity (B), and motivational tendency (C).

2 Test-retest reliability

For computational modelling parameters to successfully contribute to research advancing our understanding of mechanisms underlying mental health and disorder, as well as to transform such knowledge into personalized treatment approaches, it is pivotal to ensure reliability of the used measures. The reliability with which a measure captures individual characteristics ultimately sets an upper limit on its usefulness in detecting differences between groups, relationships to other measures, and intervention effects. We assessed the test-retest reliability of the novel effort-expenditure task in a separate in-person sample.

2.1 Methods

2.1.1 Sample

Thirty-three participants were recruited from S.O.N.A. and through advertisements in Cambridge Colleges. Three subjects had to be excluded due to failure during the task calibration. The final sample (*N*=30) consisted of 17 female and 13 male subjects with an average age of *M*=48.1 years (*SD*=21.54). All subjects were native English speakers, none reported neurological conditions and four subjects reported current and/or past psychiatric disorders (depression, anxiety, obsessive—compulsive disorder, and personality disorder).

2.1.2 Study procedure

All participants completed two testing sessions. In the first session, demographic data was collected, followed by the effort-expenditure task and the battery of self-report questionnaires. In the second session, subjects only completed the effort task. The two testing sessions were at least 6 days apart and 9 days at most, with an average difference of 6.93 days (*SD*=0.78). Time-of-day at testing was aimed to be held constant between testing sessions, with the average time difference between starting times of the task being 21.53 minutes (*SD*=30.44, *min*=0, *max*=119).

2.1.3 Analyses

All computational models defined in our model space were fit separately to the task data of each testing session using the same methods and parameters as described for the main sample in the main manuscript. Test-retest analyses were performed on the full parabolic model, given model comparison in our main sample identified this as the winning model.

Intra-class correlation coefficients (ICCs) were used to assess test-retest reliability using the ratio of intra-individual to inter-individual variability. We used two-way mixed effects ICCs considering consistency (reflecting rank order). ICCs below 0.4 indicate that a measure is not reliable, ICCs between 0.4 and 0.75 indicate moderate to good reliability and ICCs above 0.75 indicate excellent reliability[1].

To assess correlations between model parameters from the two sessions, we first calculated Pearson's correlation coefficients between mean posterior parameter estimates resulting from the separate model fitting procedures. Next, we re-fit the model jointly for both testing sessions, while embedding a correlation matrix into the model[2]. Thereby, the full posterior distributions of the model parameters are fed into the calculation of correlation coefficients, rather than solely point estimates. This offers multiple benefits of Bayesian inference: Uncertainty around parameter estimates can be incorporated into the correlation estimation and Bayesian priors can be set over possible values of the correlation matrix. Note that while this model fitting procedure fits the data of both testing sessions simultaneously, separate hyper-parameters are used for the different sessions. Therefore, shrinkage cannot bias the reliability estimates.

Finally, we also made use of the predictive property of the computational modelling approach in the assessment of parameter reliability. If parameters are reliable, the estimates resulting from modelling the data from one session should be able to predict subjects' choices in the other session better than chance[3]. To test this, we calculated the model-predicted choices for each subject and trial and compared this to the observed choices in the respective other session.

Due to our hierarchical Bayesian modelling approach, individual model parameters are subject to shrinkage. To test whether the predictive property of individual parameters is solely due to shrinkage we repeated the procedure using hyper-parameters to make model predictions and then compared the resulting predictive accuracy to that of individual parameters.

2.2 Results

2.2.1 Descriptive task statistics

Data from both sessions reproduced the expected effect of effort discounting. Mixed-effects analyses of variances (ANOVA) of repeated measures confirmed significant main effects of effort (session 1: $F(1,447)$=128.22, p<.001; session 2: $F(1,447)$=129.42, p<.001) and reward (session 1: $F(1,447)$=56.49, p<.001; session 2: reward: $F(1,447)$=46.51, p<.001), as well as an interaction effect (session 1: $F(1,447)$=52.74, p<.001; session 2: $F(1,447)$=49.13, p<.001) for both testing sessions. In post hoc ANOVAs the main effect of effort remained significant at all reward levels (at p<.05) for both sessions and the main effect of reward remained significant at all effort levels (at p<.05), except at the lowest effort level in the first session, and at the two lowest reward levels in the second session.

2.2.2 Computational modelling

All models converged well for both testing sessions. For both session one and two, the full parabolic model was the winning model, based on both the leave-one-out information criterion (LOO) and the expected log predictive density (ELPD) (**Figure S3A**).

2.2.3 Test-retest reliability

The ICC for effort sensitivity showed best test-retest reliability (ICC(C,1)=0.797, p<.001). Reliability for reward sensitivity (ICC(C,1)=0.459, p=.0047) and motivational tendency (ICC(C,1)=0.463, p=.0043), was moderate (**Figure S3B**).

Correlations between point estimates of the modelling parameters across testing sessions were very strong for effort sensitivity (r=.803, p<.01), and moderate for reward sensitivity (r=.467, p<.01) and motivational tendency (r=.517, p<.01). Correlations derived from the embedded correlation matrix and therefore considering the full parameter estimate distribution resulted in slightly higher estimates (effort sensitivity: r=.867; 95%Highest density interval (HDI)=[.667-.999]; reward sensitivity: r=.550; 95%HDI=[.163-.900]; motivational tendency: r=0.585; 95%HDI=[.184-.927]).

Subject-wise parameter estimates could reliably predict individual trial-wise choice data across trials (**Figure S3C**). Session one parameter estimates predicted session two choice data significantly better than chance (t(1919)=46.819, p<.001), as did session two parameters for session one choice data (t(1919)=47.106, p<.001). When considering predictive accuracy of group-level parameter estimates, both session's parameter estimates outperformed chance (session one predicting session two: t(1919)=33.485, p<.001; session two predicting session one: t(1919)=30.291, p<.001). Comparing predictive accuracy of subject-wise parameter estimates to group-level estimates, the subject-level outperformed the group-level for both session one predicting session two (t(3760.7)= 4.951, p<.001), and session two predicting session one (t(3674.2)=7.426, p<.001).

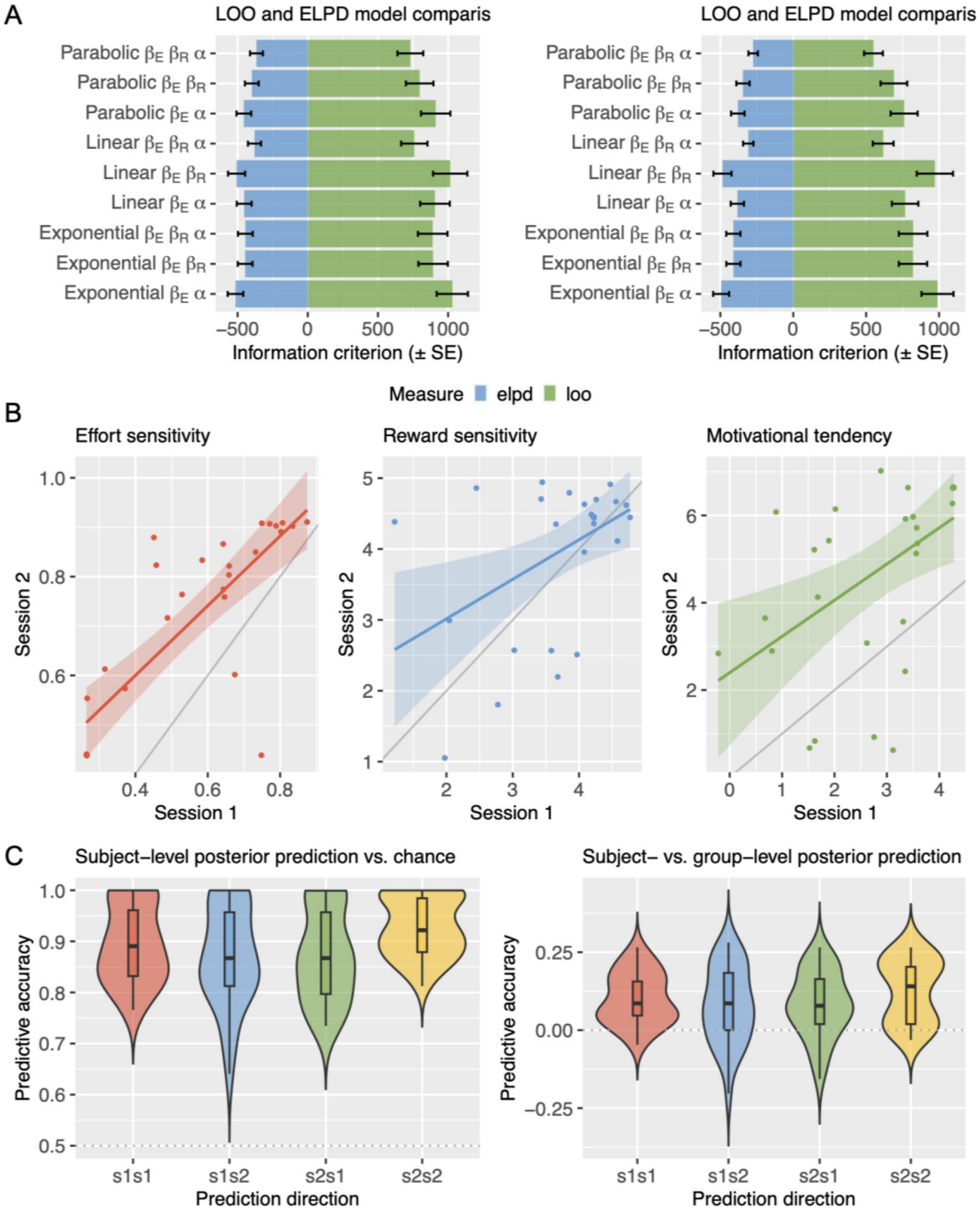


Figure S3.

Computational modelling and test-retest reliability.

A: Model comparison for each testing session based on the leave-one-out information criterion (LOO) and expected log predictive density (ELPD). Error bars indicate standard errors. **B:** Subject-wise parameter estimates compared between testing sessions. **C:** Predictive accuracy against chance (left) and group-level parameters (right; values >0 indicate better performance of subject-level compared to group-level parameters). Labels s1s2 (and s2s1) indicate session 1 (session 2) parameters predicting session 2 (session 1) data, s1s1 (and s2s2) indicate session 1 (session 2) parameters predicting session 1 (session 2) data.

3 Model agnostic task measures relating to questionnaires

3.1 Proportion of accepted trials

To explore the relationship between model agnostic task measures to questionnaire measures of neuropsychiatric symptoms, we conducted Bayesian GLMs, with the proportion of accepted trials predicted by SHAPS scores, controlling for age and gender. The proportion of accepted trials averaged across effort and reward levels was predicted by the Snaith-Hamilton Pleasure Scale (SHAPS) sum scores (M=-0.07; 95%HDI=[-0.12,-0.03]) and the Apathy Evaluation Scale (AES) sum scores (M=-0.05; 95%HDI=[-0.10,-0.002]). Note that this was not driven only by higher effort levels; even confining data to the lowest two effort levels, SHAPS has a predictive value for the proportion of accepted trials: M=-0.05; 95%HDI=[-0.07,-0.02].

A visualisation of model agnostic task measures relating to symptoms is given in **Fig. S4**, comparing subgroups of participants scoring in the highest and lowest quartile on the SHAPS. This shows that participants with a high SHAPS score (i.e., more pronounced anhedonia) are less likely to accept offers than those with a low SHAPS score (**Fig. S4A**). Due to the implemented staircasing procedure, group differences can also be seen in the effort-reward combinations offered per trial. While for both groups, the staircasing procedure seems to devolve towards high effort — low reward offers, this is more pronounced in the subgroup of participants with a lower SHAPS score (**Fig S4B**).

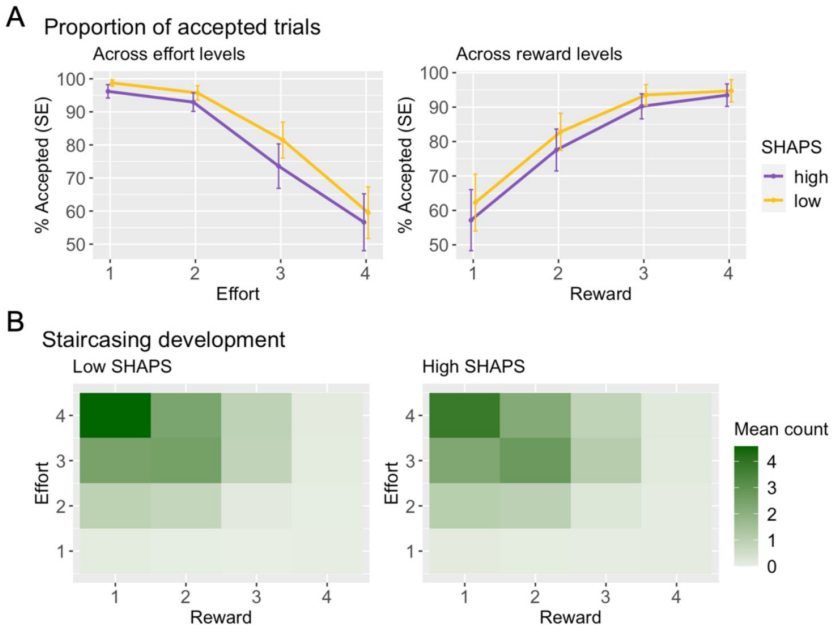


Figure S4.

Model agnostic task measures relation to anhedonia.

A: Comparing the proportion of accepted trials across effort (right) and reward (left) levels in subsamples of participants scoring in the highest and lowest SHAPS quartile. **B:** Distribution of effort-reward combinations, averaged across the final trial of 16 staircases.

3.2 Proportion of accepted but failed trials

For each participant, we computed the proportion of trial in which an offer was accepted, but the required effort then not fulfilled (i.e., failed trials). There was no relationship between average proportion of accepted but failed trials and SHAPS score (controlling for age and gender): M=0.01, 95% HDI=[-0.001,0.02]. However, there are intentionally few accepted but failed trials (*M* = 1.3% trials failed, *SD* = 3.50). This results from several task design characteristics aimed at preventing subjects from failing accepted trials, to avoid confounding of effort discounting with risk discounting.

3.3 Exertion of "extra effort"

We also explored the extent to which participants went "above and beyond" the target in accepted trials. Specifically, considering only accepted and succeeded trials, we computed the factor by which the required number of clicks was exceeded (i.e., if a subject clicked 15 times when 10 clicks were required the factor would be 1.3), averaging across effort and reward level. We then conducted a Bayesian GLM to test whether this subject wise click-exceedance measure can be predicted by apathy or anhedonia, controlling for age and gender. We found neither the SHAPS (M=-0.14, 95% HDI=[-0.43,0.17]) nor the AES (M=0.07, 95% HDI=[-0.26,0.41]) had a predictive value for the amount to which subjects exert "extra effort".

4 Subject-reported decision-making process

After completing the task, subjects were asked whether they used "any strategy to facilitate the game". While this question was initially included to monitor any self-reported cheating strategies, a large subset of subjects (425 subjects, 44.36%) understood this question as a prompt to report their experience of the decision-making process during the effort-expenditure task. Some examples of 'participants' reports are given below:

- *"Only accepting the challenge if the points were equal to or greater than the effort level."*
- *"I decided whether the points were worth the effort."*
- *"I tended to reject if the reward was only 2."*
- *"Measure of how many points to gain against effort and how tired my hand felt."*
- *"As the game progressed I decided what ratio of effort to reward that I would tolerate."*
- *"I rejected the higher difficulties which had low points."*
- *"Was it worth the effort for the points?"*
- *"I didn't do the higher effort challenges."*

References

1. Foster R. G. (2020) **Sleep, circadian rhythms and health** *Interface Focus* **10**

2. McCarthy M. J. (2019) **Missing a beat: assessment of circadian rhythm abnormalities in bipolar disorder in the genomic era** *Psychiatric Genetics* **29**

3. Logan R. W., McClung C. A. (2019) **Rhythms of life: circadian disruption and brain disorders across the lifespan** *Nat Rev Neurosci* **20**:49–65

4. Walker W. H., Walton J. C., DeVries A. C., Nelson R. J. (2020) **Circadian rhythm disruption and mental health** *Transl Psychiatry* **10**:1–13

5. Ashton A., Jagannath A. (2020) **Disrupted Sleep and Circadian Rhythms in Schizophrenia and Their Interaction With Dopamine Signaling** *Frontiers in Neuroscience* **14**

6. Lyall L. M., et al. (2018) **Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank** *The Lancet Psychiatry* **5**:507–514

7. Vadnie C. A., McClung C. A. (2017) **Circadian Rhythm Disturbances in Mood Disorders: Insights into the Role of the Suprachiasmatic Nucleus** *Neural Plasticity* **2017**

8. Emens J., Lewy A., Kinzie J. M., Arntz D., Rough J. (2009) **Circadian misalignment in major depressive disorder** *Psychiatry Research* **168**:259–261

9. Jones S. E., et al. (2019) **Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms** *Nat Commun* **10**

10. Bailey S. L., Heitkemper M. M. (2001) **Circadian rhythmicity of cortisol and body temperature: Morningness-eveningness effects** *Chronobiology International* **18**:249–261

11. Horne J. A., Östberg O. (1976) **A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms** *International Journal of Chronobiology* **4**:97–110

12. Kivelä L., Papadopoulos M. R., Antypa N. (2018) **Chronotype and Psychiatric Disorders** *Curr Sleep Medicine Rep* **4**:94–103

13. Chan J. W. Y., et al. (2014) **Eveningness and Insomnia: Independent Risk Factors of Nonremission in Major Depressive Disorder** *Sleep* **37**:911–917

14. Romo-Nava F., et al. (2020) **Evening chronotype as a discrete clinical subphenotype in bipolar disorder** *Journal of Affective Disorders* **266**:556–562

15. Au J., Reece J. (2017) **The relationship between chronotype and depressive symptoms: A meta-analysis** *Journal of Affective Disorders* **218**:93–104

16. Coleman M. Y., Cain S. W. (2019) **Eveningness is associated with greater subjective cognitive impairment in individuals with self-reported symptoms of unipolar depression** *Journal of Affective Disorders* **256**:404–415

17. McGowan N. M., Saunders K. E. A. (2021) **The Emerging Circadian Phenotype of Borderline Personality Disorder: Mechanisms, Opportunities and Future Directions** *Curr Psychiatry Rep* **23**

18. Carpenter J. S., et al. (2021) **Circadian depression: A mood disorder phenotype** *Neuroscience & Biobehavioral Reviews* **126**:79–101

19. Crouse J. J., et al. (2021) **Circadian rhythm sleep-wake disturbances and depression in young people: implications for prevention and early intervention** *The Lancet Psychiatry* **8**:813–823

20. Treadway M. T., Zald D. H. (2011) **Reconsidering anhedonia in depression: Lessons from translational neuroscience** *Neuroscience & Biobehavioral Reviews* **35**:537–555

21. Horan W. P., Kring A. M., Blanchard J. J. (2006) **Anhedonia in Schizophrenia: A Review of Assessment Strategies** *Schizophrenia Bulletin* **32**:259–273

22. Mazza M., Squillacioti M. R., Pecora R. D., Janiri L., Bria P. (2009) **Effect of aripiprazole on self-reported anhedonia in bipolar depressed patients** *Psychiatry Research* **165**:193–196

23. den Brok M. G. H. E., et al. (2015) **Apathy in Parkinson's disease: A systematic review and meta-analysis** *Movement Disorders* **30**:759–769

24. Husain M., Roiser J. P. (2018) **Neuroscience of apathy and anhedonia: a transdiagnostic approach** *Nat Rev Neurosci* **19**:470–484

25. Ducasse D., et al. (2018) **Anhedonia is associated with suicidal ideation independently of depression: A meta-analysis** *Depression and Anxiety* **35**:382–392

26. Dimick M. K., Hird M. A., Fiksenbaum L. M., Mitchell R. H. B., Goldstein B. I. (2021) **Severe anhedonia among adolescents with bipolar disorder is common and associated with increased psychiatric symptom burden** *Journal of Psychiatric Research* **134**:200–207

27. Calabrese J. R., et al. (2014) **Methodological approaches and magnitude of the clinical unmet need associated with amotivation in mood disorders** *Journal of Affective Disorders* **168**:439–451

28. Gabbay V., et al. (2015) **Anhedonia, but not Irritability, Is Associated with Illness Severity Outcomes in Adolescent Major Depression** *Journal of Child and Adolescent Psychopharmacology* **25**:194–200

29. McMakin D. L., et al. (2012) **Anhedonia Predicts Poorer Recovery Among Youth With Selective Serotonin Reuptake Inhibitor Treatment—Resistant Depression** *Journal of the American Academy of Child & Adolescent Psychiatry* **51**:404–411

30. Treadway M. T., Buckholtz J. W., Schwartzman A. N., Lambert W. E., Zald D. H. (2009) **Worth the 'EEfRT'? The Effort Expenditure for Rewards Task as an Objective Measure of Motivation and Anhedonia** *PLOS ONE* **4**

31. Treadway M. T., Bossaller N., Shelton R. C., Zald D. H. (2012) **Effort-Based Decision-Making in Major Depressive Disorder: A Translational Model of Motivational Anhedonia** *J Abnorm Psychol* **121**:553–558

32. Pessiglione M., Vinckier F., Bouret S., Daunizeau J., Le Bouc R. (2018) **Why not try harder? Computational approach to motivation deficits in neuro-psychiatric diseases** *Brain* **141**:629–650

33. Sugiwaka H., Okouchi H. (2004) **Reformative self-control and discounting of reward value by delay or effort1** *Japanese Psychological Research* **46**:1–9

34. Hartmann M. N., Hager O. M., Tobler P. N., Kaiser S. (2013) **Parabolic discounting of monetary rewards by physical effort** *Behavioural Processes* **100**:192–196

35. Klein-Flügge M. C., Kennerley S. W., Saraiva A. C., Penny W. D., Bestmann S. (2015) **Behavioral Modeling of Human Choices Reveals Dissociable Effects of Physical Effort and Temporal Delay on Reward Devaluation** *PLOS Computational Biology* **11**

36. Białaszek W., Marcowski P., Ostaszewski P. (2017) **Physical and cognitive effort discounting across different reward magnitudes: Tests of discounting models** *PLOS ONE* **12**

37. Ostaszewski P., Bąbel P., Swebodzinski B. (2013) **Physical and cognitive effort discounting of hypothetical monetary rewards** *Japanese Psychological Research* **55**:329–337

38. Tran A. H., et al. (2002) **Altered accumbens neural response to prediction of reward associated with place in dopamine D2 receptor knockout mice** *Proc Natl Acad Sci U S A* **99**:8986–8991

39. Robbins T. W., Roberts D. C., Koob G. F. (1983) **Effects of d-amphetamine and apomorphine upon operant behavior and schedule-induced licking in rats with 6-hydroxydopamine-induced lesions of the nucleus accumbens** *J Pharmacol Exp Ther* **224**:662–673

40. Cagniard B., Balsam P. D., Brunner D., Zhuang X. (2006) **Mice with Chronically Elevated Dopamine Exhibit Enhanced Motivation, but not Learning, for a Food Reward** *Neuropsychopharmacol* **31**:1362–1370

41. Taylor J. R., Robbins T. W. (1984) **Enhanced behavioural control by conditioned reinforcers following microinjections of d-amphetamine into the nucleus accumbens** *Psychopharmacology* **84**:405–412

42. Cawley E., et al. (2013) **Dopamine and light: dissecting effects on mood and motivational states in women with subsyndromal seasonal affective disorder** *J Psychiatry Neurosci* **38**:388–397

43. Venugopalan V. V., et al. (2011) **Acute Phenylalanine/Tyrosine Depletion Reduces Motivation to Smoke Cigaredes Across Stages of Addiction** *Neuropsychopharmacol* **36**:2469–2476

44. Wardle M. C., Treadway M. T., Mayo L. M., Zald D. H., de Wit H. (2011) **Amping Up Effort: Effects of d-Amphetamine on Human Effort-Based Decision-Making** *Journal of Neuroscience* **31**:16597–16602

45. Soder H. E., et al. (2021) **Dose-response effects of d-amphetamine on effort-based decision-making and reinforcement learning** *Neuropsychopharmacol* **46**:1078–1085

46. Treadway M. T., et al. (2012) **Dopaminergic Mechanisms of Individual Differences in Human Effort-Based Decision-Making** *Journal of Neuroscience* **32**:6170–6176

47. Kiehn J.-T., Faltraco F., Palm D., Thome J., Oster H. (2023) **Circadian Clocks in the Regulation of Neurotransmider Systems** *Pharmacopsychiatry* **56**:108–117

48. Castañeda T. R., de Prado B. M., Prieto D., Mora F. (2004) **Circadian rhythms of dopamine, glutamate and GABA in the striatum and nucleus accumbens of the awake rat: modulation by light** *Journal of Pineal Research* **36**:177–185

49. Ferris M. J., et al. (2014) **Dopamine transporters govern diurnal variation in extracellular dopamine tone** *Proc. Natl. Acad. Sci. U.S.A* **111**

50. Weber M., Lauterburg T., Tobler I., Burgunder J.-M. (2004) **Circadian paderns of neurotransmider related gene expression in motor regions of the rat brain** *Neuroscience Letters* **358**:17–20

51. Hampp G., et al. (2008) **Regulation of Monoamine Oxidase A by Circadian-Clock Components Implies Clock Influence on Mood** *Current Biology* **18**:678–683

52. Imbesi M., et al. (2009) **Dopamine receptor-mediated regulation of neuronal "clock" gene expression** *Neuroscience* **158**:537–544

53. Holst S. C., et al. (2014) **Dopaminergic Role in Regulating Neurophysiological Markers of Sleep Homeostasis in Humans** *Journal of Neuroscience* **34**:566–573

54. Valomon A., et al. (2014) **Genetic polymorphisms of DAT1 and COMT differentially associate with actigraphy-derived sleep-wake cycles in young adults** *Chronobiology International* **31**:705–714

55. Zhang R., et al. (2021) **Dopamine D1 and D2 receptors are distinctly associated with rest-activity rhythms and drug reward** *J Clin Invest* **131**

56. Shumay E., et al. (2012) **Repeat variation in the human PER2 gene as a new genetic marker associated with cocaine addiction and brain dopamine D2 receptor availability** *Transl Psychiatry* **2**

57. Boland E. M., et al. (2022) **Poor sleep quality is significantly associated with effort but not temporal discounting of monetary rewards** *Motivation Science* **8**:70–76

58. Boyle C. C., et al. (2020) **Motivation and sensitivity to monetary reward in late-life insomnia: moderating role of sex and the inflammatory marker CRP** *Neuropsychopharmacol* **45**:1664–1671

59. Libedinsky C., et al. (2013) **Sleep Deprivation Alters Effort Discounting but not Delay Discounting of Monetary Rewards** *Sleep* **36**:899–904

60. Bijleveld E., Knufinke M. (2018) **Exposure to bright light biases effort-based decisions** *Behavioral Neuroscience* **132**:183–193

61. Boland E. M., Bertulis K., Leong S. H., Thase M. E., Gehrman P. R. (2019) **Preliminary support for the role of reward relevant effort and chronotype in the depression/insomnia comorbidity** *Journal of Affective Disorders* **242**:220–223

62. Evans S. L., Norbury R. (2021) **Associations between diurnal preference, impulsivity and substance use in a young-adult student sample** *Chronobiology International* **38**:79–89

63. Correa Á., Alguacil S., Ciria L. F., Jiménez A., Ruz M. (2020) **Circadian rhythms and decision-making: a review and new evidence from electroencephalography** *Chronobiology International* **37**:520–541

64. Ingram K. K., et al. (2016) **Molecular insights into chronotype and time-of-day effects on decision-making** *Sci Rep* **6**

65. Hisler G. C., Dickinson D. L., Bruce S. A., Hasler B. P. (2023) **Preliminary evidence that misalignment between sleep and circadian timing alters risk-taking preferences** *Journal of Sleep Research* **32**

66. Bedder R. L., Vaghi M. M., Dolan R. J., Rutledge R. B. (2023) **Risk taking for potential losses but not gains increases with time of day** *Sci Rep* **13**

67. Dercon Q., et al. (2023) **A core component of psychological therapy causes adaptive changes in computational learning mechanisms** *Psychol. Med* :1–11 https://doi.org/10.1017/S0033291723001587

68. McManus S., Bebbington P. E., Jenkins R., Brugha T. (2016) **McManus, S., Bebbington, P. E., Jenkins, R. & Brugha, T. Mental Health and Wellbeing in England: The Adult Psychiatric Morbidity Survey 2014. https://files.digital.nhs.uk/pdf/q/3/mental_health_and_wellbeing_in_england_full_report.pdf (2016).**

69. Office of National Statistics (2016) **CT0570_2011 Census—Sex by age by IMD2004 by ethnic group**

70. Marin R. S., Biedrzycki R. C., Firinciogullari S. (1991) **Reliability and validity of the apathy evaluation scale** *Psychiatry Research* **38**:143–162

71. Snaith R. P., et al. (1995) **A Scale for the Assessment of Hedonic Tone the Snaith—Hamilton Pleasure Scale** *The British Journal of Psychiatry* **167**:99–103

72. Mkrtchian A., Valton V., Roiser J. P. (2023) **Reliability of Decision-Making and Reinforcement Learning Computational Parameters** *Computational Psychiatry* **7**:30–46

73. Rizvi S. J., et al. (2015) **Development and validation of the Dimensional Anhedonia Rating Scale (DARS) in a community sample and individuals with major depression** *Psychiatry Research* **229**:109–119

74. Roenneberg T., Wirz-Justice A., Merrow M. (2003) **Life between Clocks: Daily Temporal Paderns of Human Chronotypes** *J Biol Rhythms* **18**:80–90

75. Lindström J., Tuomilehto J. (2003) **The Diabetes Risk Score** *Diabetes Care* **26**:725–731

76. Lecrubier Y., et al. (1997) **The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI** *European Psychiatry* **12**:224–231

77. Barch D. M., Treadway M., Schoen N. (2014) **Effort, Anhedonia, and Function in Schizophrenia: Reduced Effort Allocation Predicts Amotivation and Functional Impairment** *J Abnorm Psychol* **123**:387–397

78. Berwian I. M., et al. (2020) **Computational Mechanisms of Effort and Reward Decisions in Patients With Depression and Their Association With Relapse After Antidepressant Discontinuation** *JAMA Psychiatry* **77**:513–522

79. Cléry-Melin M.-L., et al. (2011) **Why Don't You Try Harder? An Investigation of Effort Production in Major Depression** *PLOS ONE* **6**

80. Fervaha G., et al. (2013) **Incentive motivation deficits in schizophrenia reflect effort computation impairments during cost-benefit decision-making** *Journal of Psychiatric Research* **47**:1590–1596

81. Gold J. M., et al. (2013) **Negative Symptoms of Schizophrenia Are Associated with Abnormal Effort-Cost Computations** *Biological Psychiatry* **74**:130–136

82. Hershenberg R., et al. (2016) **Diminished effort on a progressive ratio task in both unipolar and bipolar depression** *Journal of Affective Disorders* **196**:97–100

83. Wolf D. H., et al. (2014) **Amotivation in Schizophrenia: Integrated Assessment With Behavioral, Clinical, and Imaging Measures** *Schizophrenia Bulletin* **40**:1328–1337

84. Yang X., et al. (2014) **Motivational deficits in effort-based decision making in individuals with subsyndromal depression, first-episode and remided depression patients** *Psychiatry Research* **220**:874–882

85. Zou Y.-M., et al. (2020) **Effort—cost computation in a transdiagnostic psychiatric sample: Differences among patients with schizophrenia, bipolar disorder, and major depressive disorder** *PsyCh Journal* **9**:210–222

86. Chong T. T.-J., et al. (2015) **Dopamine enhances willingness to exert effort for reward in Parkinson's disease** *Cortex* **69**:40–46

87. Chong T. T.-J., et al. (2018) **Dissociation of reward and effort sensitivity in methcathinone-induced Parkinsonism** *Journal of Neuropsychology* **12**:291–297

88. Le Bouc R., et al. (2016) **Computational Dissection of Dopamine Motor and Motivational Functions in Humans** *J. Neurosci* **36**:6623–6633

89. Le Bouc R., et al. (2023) **Effort avoidance as a core mechanism of apathy in frontotemporal dementia** *Brain* **146**:712–726

90. Le Heron C., et al. (2018) **Distinct effects of apathy and dopamine on effort-based decision-making in Parkinson's disease** *Brain* **141**:1455–1469

91. Tran T., Hagen A. E. F., Hollenstein T., Bowie C. R. (2021) **Physical- and Cognitive-Effort-Based Decision-Making in Depression: Relationships to Symptoms and Functioning** *Clinical Psychological Science* **9**:53–67

92. Bonnelle V., et al. (2015) **Characterization of reward and effort mechanisms in apathy** *Journal of Physiology-Paris* **109**:16–26

93. Jurgelis M., et al. (2021) **Heightened effort discounting is a common feature of both apathy and fatigue** *Sci Rep* **11**

94. Silvia P. J., et al. (2016) **Do depressive symptoms "blunt" effort? An analysis of cardiac engagement and withdrawal for an increasingly difficult task** *Biological Psychology* **118**:52–60

95. Korn C. W., Sharot T., Walter H., Heekeren H. R., Dolan R. J. (2014) **Depression is related to an absence of optimistically biased belief updating about future life events** *Psychological Medicine* **44**:579–592

96. May C. P., Hasher L. (1998) **Synchrony Effects in Inhibitory Control Over Thought and Action** *Journal of Experimental Psychology: Human Perception and Performance*

97. Goldstein D., Hahn C. S., Hasher L., Wiprzycka U. J., Zelazo P. D. (2007) **Time of day, intellectual performance, and behavioral problems in Morning versus Evening type adolescents: Is there a synchrony effect?** *Personality and Individual Differences* **42**:431–440

98. Lehmann C. A., Marks A. D. G., Hanstock T. L. (2013) **Age and synchrony effects in performance on the Rey Auditory Verbal Learning Test** *International Psychogeriatrics* **25**:657–665

99. Barner C., Schmid S. R., Diekelmann S. (2019) **Time-of-day effects on prospective memory** *Behavioural Brain Research* **376**

100. Nowack K., Van Der Meer E. (2018) **The synchrony effect revisited: chronotype, time of day and cognitive performance in a semantic analogy task** *Chronobiology International* **35**:1647–1662

101. Forbes E. E., et al. (2012) **PER2 rs2304672 Polymorphism Moderates Circadian-Relevant Reward Circuitry Activity in Adolescents** *Biological Psychiatry* **71**:451–457

102. Hasler B. P., Sitnick S. L., Shaw D. S., Forbes E. E. (2013) **An altered neural response to reward may contribute to alcohol problems among late adolescents with an evening chronotype** *Psychiatry Research: Neuroimaging* **214**:357–364

103. Holm S. M., et al. (2009) **Reward-Related Brain Function and Sleep in Pre/Early Pubertal and Mid/Late Pubertal Adolescents** *Journal of Adolescent Health* **45**:326–334

104. Chen Z., et al. (2022) **Diurnal mood variation symptoms in major depressive disorder associated with evening chronotype: Evidence from a neuroimaging study** *Journal of Affective Disorders* **298**:151–159

105. Mehrhof S. Z., Fleming H. A., Nord C. (2024) **A cognitive signature of metabolic health in effort-based decision-making** *PsyArXiv* https://doi.org/10.31234/osf.io/4bkm9

106. Gillan C. M., Daw N. D. (2016) **Taking Psychiatry Research Online** *Neuron* **91**:19–23

107. Kantermann T., Sung H., Burgess H. J. (2015) **Comparing the Morningness-Eveningness Questionnaire and Munich ChronoType Questionnaire to the Dim Light Melatonin Onset** *J Biol Rhythms* **30**:449–453

108. Santisteban J. A., Brown T. G., Gruber R. (2018) **Association between the Munich Chronotype Questionnaire and Wrist Actigraphy** *Sleep Disorders* **2018**

109. Nebel L. E., et al. (1996) **The circadian variation of cardiovascular stress levels and reactivity: Relationships to individual differences in morningness/eveningness** *Psychophysiology* **33**:273–281

110. Vrizzi S., Najar A., Lemogne C., Palminteri S., Lebreton M. (2023) **Comparing the test-retest reliability of behavioral, computational and self-reported individual measures of reward and punishment sensitivity in relation to mental health symptoms** *PsyArXiv* https://doi.org/10.31234/osf.io/3u4gp

111. Bhatnagar A., Murray G., Ray S. (2023) **Circadian biology to advance therapeutics for mood disorders** *Trends in Pharmacological Sciences* **44**:689–704

112. Meuret A. E., et al. (2016) **Timing maders: Endogenous cortisol mediates benefits from early-day psychotherapy** *Psychoneuroendocrinology* **74**:197–202

113. de Leeuw J. R. jsPsych (2015) **A JavaScript library for creating behavioral experiments in a Web browser** *Behav Res* **47**:1–12

114. Palan S., Schider C. (2018) **Prolific.ac—A subject pool for online experiments** *Journal of Behavioral and Experimental Finance* **17**:22–27

115. Stansfeld S, et al. (2016) **Chapter 2: Common Mental Disorders** *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014* :37–68

116. Craig C. L., et al. (2003) **International Physical Activity Questionnaire: 12-Country Reliability and Validity** *Medicine & Science in Sports & Exercise* **35**:1381–1395

117. Ahn W.-Y., Krawite A., Kim W., Busmeyer J. R., Brown J. W. (2011) **A ModelBased fMRI Analysis with Hierarchical Bayesian Parameter Estimation** *J Neurosci Psychol Econ* **4**:95–110

118. Stan Development Team (2021) **Stan Modelling Language Users Guide and Reference Manual**

119. Ahn W.-Y., Haines N., Zhang L. (2017) **Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package** *Comput Psychiatr* **1**:24–57

120. Dinga R., et al. (2017) **Evaluating the evidence for biotypes of depression: Methodological replication and extension of Drysdale et al** *NeuroImage: Clinical* **22**

121. Roenneberg Till, Pilz Luisa K., Zerbini Giulia, Winnebeck Eva C. (2019) **Chronotype and Social Jetlag: A (Self-) Critical Review** *Biology* **8**

1. Fleiss J. L. (2011) **Design and Analysis of Clinical Experiments**

2. Haines N., et al. (2020) **Learning from the reliability paradox: How theoretically informed generative models can advance the social, behavioral, and brain sciences** *PsyArXiv* https://doi.org/10.31234/osf.io/xr7y3

3. Mkrtchian A., Valton V., Roiser J. P. (2023) **Reliability of Decision-Making and Reinforcement Learning Computational Parameters** *Computational Psychiatry* **7**:30–46

Editors

Reviewer #1 (Public Review):

Summary:

This study uses an online cognitive task to assess how reward and effort are integrated in a motivated decision-making task. In particular the authors were looking to explore how neuropsychiatric symptoms, in particular, apathy and anhedonia, and circadian rhythms affect behavior in this task. Amongst many results, they found that choice bias (the degree to which integrated reward and effort affect decisions) is reduced in individuals with greater neuropsychiatric symptoms, and late chronotypes (being an 'evening person').

Strengths:

The authors recruited participants to perform the cognitive task both in and out of sync with their chronotypes, allowing for the important insight that individuals with late chronotypes show a more reduced choice bias when tested in the morning.
Overall, this is a well-designed and controlled online experimental study. The modelling approach is robust, with care being taken to both perform and explain to the readers the various tests used to ensure the models allow the authors to sufficiently test their hypotheses.

Weaknesses:

This study was not designed to test the interactions of neuropsychiatric symptoms and chronotypes on decision making, and thus can only make preliminary suggestions regarding how symptoms, chronotypes and time-of-assessment interact.

https://doi.org/10.7554/eLife.96803.2.sa3

Reviewer #2 (Public Review):

Summary:

The study combines computational modeling of choice behavior with an economic, effort-based decision-making task to assess how willingness to exert physical effort for a reward varies as a function of individual differences in apathy and anhedonia, or depression, as well as chronotype. They find an overall reduction in effort selection that scales with apathy, anhedonia and depression. They also find that later chronotypes are less likely to choose effort than earlier chronotypes and, interestingly, an interaction whereby later chronotypes are especially unwilling to exert effort in the morning versus the evening.

Strengths:

This study uses state-of-the-art tools for model fitting and validation and regression methods which rule out multicollinearity among symptom measures and Bayesian methods which estimate effects and uncertainty about those estimates. The replication of results across two

different kinds of samples is another strength. Finally, the study provides new information about the effects not only of chronotype but also chronotype by timepoint interactions which are previously unknown in the subfield of effort-based decision-making.

Weaknesses:

The study has few weaknesses. The biggest drawback is that it does not provide evidence for the idea that a match between chronotype and delay matters is especially relevant for people with depression or continuous measures like anhedonia and apathy. It is unclear whether disorders further interact with chronotype and time of day to determine a bias against effort. On the other hand, the study does provide evidence that future studies should consider such interactions when examining questions about effort expenditure in psychiatric disorders.

https://doi.org/10.7554/eLife.96803.2.sa2

Reviewer #3 (Public Review):

Summary:

In this manuscript, Mehrhof and Nord study a large dataset of participants collected online (n=958 after exclusions) who performed a simple effort-based choice task. They report that the level of effort and reward influence choices in a way that is expected from prior work. They then relate choice preferences to neuropsychiatric syndromes and, in a smaller sample (n<200), to people's circadian preferences, i.e., whether they are a morning-preferring or evening-preferring chronotype. They find relationships between the choice bias (a model parameter capturing the likelihood to accept effort-reward challenges, like an intercept) and anhedonia and apathy, as well as chronotype. People with higher anhedonia and apathy and an evening chronotype are less likely to accept challenges (more negative choice bias). People with an evening chronotype are also more reward sensitive and more likely to accept challenges in the evening, compared to the morning.

Strengths:

This is an interesting and well-written manuscript which replicates some known results and introduces a new consideration related to chronotype relationships which have not been explored before. It uses a large sample size and includes analyses related to transdiagnostic as well as diagnostic criteria.

Weaknesses:

The authors do not explore how chronotype and depression are related (does one mediate the effect of the other etc). Both variables are included in the same model in the revised article now which is a great improvement, but it also means psychopathology and circadian rhythms are treated as distinct phenomena and their relationship in predicting effort-reward preferences is not examined.

https://doi.org/10.7554/eLife.96803.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***eLife assessment***
>
> *This important study provides solid evidence that both psychiatric dimensions (e.g. anhedonia, apathy, or depression) and chronotype (i.e., being a morning or evening person) influence effort-based decision-making. Notably, the current study does not elucidate whether there may be interactive effects of chronotype and psychiatric dimensions on decision-making. This work is of importance to researchers and clinicians alike, who may make inferences about behaviour and cognition without taking into account whether the individual may be tested or observed out-of-sync with their phenotype.*

We thank the three reviewers for their comments, and the Editors at eLife. We have taken the opportunity to revise our manuscript considerably from its original form, not least because we feel a number of the reviewers' suggested analyses strengthen our manuscript considerably (in one instance even clarifying our conclusions, leading us to change our title) —for which we are very appreciative indeed.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public Review):***
>
> *Summary:*
>
> *This study uses an online cognitive task to assess how reward and effort are integrated in a motivated decision-making task. In particular the authors were looking to explore how neuropsychiatric symptoms, in particular apathy and anhedonia, and circadian rhythms affect behavior in this task. Amongst many results, they found that choice bias (the degree to which integrated reward and effort affects decisions) is reduced in individuals with greater neuropsychiatric symptoms, and late chronotypes (being an 'evening person').*
>
> *Strengths:*
>
> *The authors recruited participants to perform the cognitive task both in and out of sync with their chronotypes, allowing for the important insight that individuals with late chronotypes show a more reduced choice bias when tested in the morning.*
> *Overall, this is a well-designed and controlled online experimental study. The modelling approach is robust, with care being taken to both perform and explain to the readers the various tests used to ensure the models allow the authors to sufficiently test their hypotheses.*
>
> *Weaknesses:*
>
> *This study was not designed to test the interactions of neuropsychiatric symptoms and chronotypes on decision making, and thus can only make preliminary suggestions regarding how symptoms, chronotypes and time-of-assessment interact.*

We appreciate the Reviewer's positive view of our research and agree with their assessment of its weaknesses; the study was not designed to assess chronotype-mental health interactions. We hope that our new title and contextualisation makes this clearer. We respond in more detail point-by-point below.

> ***Reviewer #2 (Public Review):***
>
> *Summary:*

> *The study combines computational modeling of choice behavior with an economic, effort-based decision-making task to assess how willingness to exert physical effort for a reward varies as a function of individual differences in apathy and anhedonia, or depression, as well as chronotype. They find an overall reduction in effort selection that scales with apathy and anhedonia and depression. They also find that later chronotypes are less likely to choose effort than earlier chronotypes and, interestingly, an interaction whereby later chronotypes are especially unwilling to exert effort in the morning versus the evening.*
>
> *Strengths:*
>
> *This study uses state-of-the-art tools for model fitting and validation and regression methods which rule out multicollinearity among symptom measures and Bayesian methods which estimate effects and uncertainty about those estimates. The replication of results across two different kinds of samples is another strength. Finally, the study provides new information about the effects not only of chronotype but also chronotype by timepoint interactions which are previously unknown in the subfield of effort-based decision-making.*
>
> *Weaknesses:*
>
> *The study has few weaknesses. One potential concern is that the range of models which were tested was narrow, and other models might have been considered. For example, the Authors might have also tried to fit models with an overall inverse temperature parameter to capture decision noise. One reason for doing so is that some variance in the bias parameter might be attributed to noise, which was not modeled here. Another concern is that the manuscripts discuss effort-based choice as a transdiagnostic feature - and there is evidence in other studies that effort deficits are a transdiagnostic feature of multiple disorders. However, because the present study does not investigate multiple diagnostic categories, it doesn't provide evidence for transdiagnosticity, per se.*

We appreciate Reviewer 2's assessment of our research and agree generally with its weaknesses. We have now addressed the Reviewer's comments regarding transdiagnosticity in the discussion of our revised version and have addressed their detailed recommendations below (see point-by-point responses).

In addition to the below specific changes, in our Discussion section, we now have also added the following (lines 538 – 540):

"Finally, we would like to note that as our study is based on a general population sample, rather than a clinical one. Hence, we cannot speak to transdiagnosticity on the level of multiple diagnostic categories."

> ***Reviewer #3 (Public Review):***
>
> *Summary:*
>
> *In this manuscript, Mehrhof and Nord study a large dataset of participants collected online (n=958 after exclusions) who performed a simple effort-based choice task. They report that the level of effort and reward influence choices in a way that is expected from prior work. They then relate choice preferences to neuropsychiatric syndromes and, in a smaller sample (n<200), to people's circadian preferences, i.e., whether they are a morning-preferring or evening-preferring chronotype. They find relationships between the choice bias (a model parameter capturing the likelihood to accept effort-reward challenges, like an intercept) and anhedonia and apathy, as well as chronotype. People with higher anhedonia and apathy and an evening chronotype are less likely to accept*

challenges (more negative choice bias). People with an evening chronotype are also more reward sensitive and more likely to accept challenges in the evening, compared to the morning.

Strengths:

This is an interesting and well-written manuscript which replicates some known results and introduces a new consideration related to potential chronotype relationships which have not been explored before. It uses a large sample size and includes analyses related to transdiagnostic as well as diagnostic criteria. I have some suggestions for improvements.

Weaknesses:

(1) The novel findings in this manuscript are those pertaining to transdiagnostic and circadian phenotypes. The authors report two separate but "overlapping" effects: individuals high on anhedonia/apathy are less willing to accept offers in the task, and similarly, individuals tested off their chronotype are less willing to accept offers in the task. The authors claim that the latter has implications for studying the former. In other words, because individuals high on anhedonia/apathy predominantly have a late chronotype (but might be tested early in the day), they might accept less offers, which could spuriously look like a link between anhedonia/apathy and choices but might in fact be an effect of the interaction between chronotype and time-of-testing. The authors therefore argue that chronotype needs to be accounted for when studying links between depression and effort tasks.

The authors argue that, if X is associated with Y and Z is associated with Y, X and Z might confound each other. That is possible, but not necessarily true. It would need to be tested explicitly by having X (anhedonia/apathy) and Z (chronotype) in the same regression model. Does the effect of anhedonia/apathy on choices disappear when accounting for chronotype (and time-of-testing)? Similarly, when adding the interaction between anhedonia/apathy, chronotype, and time-of-testing, within the subsample of people tested off their chronotype, is there a residual effect of anhedonia/apathy on choices or not?

If the effect of anhedonia/apathy disappeared (or got weaker) while accounting for chronotype, this result would suggest that chronotype mediates the effect of anhedonia/apathy on effort choices. However, I am not sure it renders the direct effect of anhedonia/apathy on choices entirely spurious. Late chronotype might be a feature (induced by other symptoms) of depression (such as fatigue and insomnia), and the association between anhedonia/apathy and effort choices might be a true and meaningful one. For example, if the effect of anhedonia/apathy on effort choices was mediated by altered connectivity of the dorsal ACC, we would not say that ACC connectivity renders the link between depression and effort choices "spurious", but we would speak of a mechanism that explains this effect. The authors should discuss in a more nuanced way what a significant mediation by the chronotype/time-of-testing congruency means for interpreting effects of depression in computational psychiatry.

We thank the Reviewer for pointing out this crucial weakness in the original version of our manuscript. We have now thought deeply about this and agree with the Reviewer that our original results did not warrant our interpretation that reported effects of anhedonia and apathy on measures of effort-based decision-making could potentially be spurious. At the Reviewer's suggestion, we decided to test this explicitly in our revised version—a decision that has now deepened our understanding of our results, and changed our interpretation thereof.

To investigate how the effects of neuropsychiatric symptoms and the effects of circadian measures relate to each other, we have followed the Reviewer's advice and conducted an additional series of analyses (see below). Surprisingly (to us, but perhaps not the Reviewer) we discovered that all three symptom measures (two of anhedonia, one of apathy) have separable effects from circadian measures on the decision to expend effort (note we have also re-named our key parameter 'motivational tendency' to address this Reviewer's next comment that the term 'choice bias' was unclear). In model comparisons (based on leave-one-out information criterion which penalises for model complexity) the models including both circadian and psychiatric measures always win against the models including either circadian or psychiatric measures. In essence, this strengthens our claims about the importance of measuring circadian rhythm in effort-based tasks generally, as circadian rhythm clearly plays an important role even when considering neuropsychiatric symptoms, but crucially does not support the idea of spurious effects: statistically, circadian measures contributes separably from neuropsychiatric symptoms to the variance in effort-based decision-making. We think this is very interesting indeed, and certainly clarifies (and corrects the inaccuracy in) our original interpretation—and can only express our thanks to the Reviewer for helping us understand our effect more fully.

In response to these new insights, we have made numerous edits to our manuscript. First, we changed the title from "Overlapping effects of neuropsychiatric symptoms and circadian rhythm on effort-based decision-making" to "Both neuropsychiatric symptoms and circadian rhythm alter effort-based decision-making". In the remaining manuscript we now refrain from using the word 'overlapping' (which could be interpreted as overlapping in explained variance), and instead opted to describe the effects as parallel. We hope our new analyses, title, and clarified/improved interpretations together address the Reviewer's valid concern about our manuscript's main weakness.

We detail these new analyses in the Methods section as follows (lines 800 – 814):

"4.5.2. Differentiating between the effects of neuropsychiatric symptoms and circadian measures on motivational tendency

To investigate how the effects of neuropsychiatric symptoms on motivational tendency (2.3.1) relate to effects of chronotype and time-of-day on motivational tendency we conducted exploratory analyses. In the subsamples of participants with an early or late chronotype (including additionally collected data), we first ran Bayesian GLMs with neuropsychiatric questionnaire scores (SHAPS, DARS, AES respectively) predicting motivational tendency, controlling for age and gender. We next added an interaction term of chronotype and time-of-day into the GLMs, testing how this changes previously observed neuropsychiatric and circadian effects on motivational tendency. Finally, we conducted a model comparison using LOO, comparing between motivational tendency predicted by a neuropsychiatric questionnaire, motivational tendency predicted by chronotype and time-of-day, and motivational tendency predicted by a neuropsychiatric questionnaire and time-of-day (for each neuropsychiatric questionnaire, and controlling for age and gender)."

Results of the outlined analyses are reported in the results section as follows (lines 356 – 383):

"2.5.2.1 Neuropsychiatric symptoms and circadian measures have separable effects on motivational tendency

Exploratory analyses testing for the effects of neuropsychiatric questionnaires on motivational tendency in the subsamples of early and late chronotypes confirmed the predictive value of the SHAPS (M=-0.24, 95% HDI=[-0.42,-0.06]), the DARS (M=-0.16, 95% HDI=[-0.31,-0.01]), and the AES (M=-0.18, 95% HDI=[-0.32,-0.02]) on motivational tendency.

For the SHAPS, we find that when adding the measures of chronotype and time-of-day back into the GLMs, the main effect of the SHAPS (M=-0.26, 95% HDI=[-0.43,-0.07]), the main effect of chronotype (M=-0.11, 95% HDI=[-0.22,-0.01]), and the interaction effect of chronotype and time-of-day (M=0.20, 95% HDI=[0.07,0.34]) on motivational tendency remain. Model comparison by LOOIC reveals motivational tendency is best predicted by the model including the SHAPS, chronotype and time-of-day as predictors, followed by the model including only the SHAPS. Note that this approach to model comparison penalizes models for increasing complexity.

Repeating these steps with the DARS, the main effect of the DARS is found numerically, but the 95% HDI just includes 0 (M=-0.15, 95% HDI=[-0.30,0.002]). The main effect of chronotype (M=-0.11, 95% HDI=[-0.21,-0.01]), and the interaction effect of chronotype and time-of-day (M=0.18, 95% HDI=[0.05,0.33]) on motivational tendency remain. Model comparison identifies the model including the DARS and circadian measures as the best model, followed by the model including only the DARS.

For the AES, the main effect of the AES is found (M=-0.19, 95% HDI=[-0.35,-0.04]). For the main effect of chronotype, the 95% narrowly includes 0 (M=-0.10, 95% HDI=[-0.21,0.002]), while the interaction effect of chronotype and time-of-day (M=0.20, 95% HDI=[0.07,0.34]) on motivational tendency remains. Model comparison identifies the model including the AES and circadian measures as the best model, followed by the model including only the AES."

We have now edited parts of our Discussion to discuss and reflect these new insights, including the following.

Lines 399 – 402:

"Various neuropsychiatric disorders are marked by disruptions in circadian rhythm, such as a late chronotype. However, research has rarely investigated how transdiagnostic mechanisms underlying neuropsychiatric conditions may relate to inter-individual differences in circadian rhythm."

Lines 475 – 480:

"It is striking that the effects of neuropsychiatric symptoms on effort-based decision-making largely are paralleled by circadian effects on the same neurocomputational parameter. Exploratory analyses predicting motivational tendency by neuropsychiatric symptoms and circadian measures simultaneously indicate the effects go beyond recapitulating each other, but rather explain separable parts of the variance in motivational tendency."

Lines 528 – 532:

"Our reported analyses investigating neuropsychiatric and circadian effects on effort-based decision-making simultaneously are exploratory, as our study design was not ideally set out to examine this. Further work is needed to disentangle separable effects of neuropsychiatric and circadian measures on effort-based decision-making."

Lines 543 – 550:

"We demonstrate that neuropsychiatric effects on effort-based decision-making are paralleled by effects of circadian rhythm and time-of-day. Exploratory analyses suggest these effects account for separable parts of the variance in effort-based decision-making. It unlikely that effects of neuropsychiatric effects on effort-based decision-making reported here and in previous literature are a spurious result due to multicollinearity with chronotype. Yet, not accounting for chronotype and time of testing, which is the predominant practice in the field, could affect results."

> *(2) It seems that all key results relate to the choice bias in the model (as opposed to reward or effort sensitivity). It would therefore be helpful to understand what fundamental process the choice bias is really capturing in this task. This is not discussed, and the direction of effects is not discussed either, but potentially quite important. It seems that the choice bias captures how many effortful reward challenges are accepted overall which maybe captures general motivation or task engagement. Maybe it is then quite expected that this could be linked with questionnaires measuring general motivation/pleasure/task engagement. Formally, the choice bias is the constant term or intercept in the model for p(accept), but the authors never comment on what its sign means. If I'm not mistaken, people with higher anhedonia but also higher apathy are less likely to accept challenges and thus engage in the task (more negative choice bias). I could not find any discussion or even mention of what these results mean. This similarly pertains to the results on chronotype. In general, "choice bias" may not be the most intuitive term and the authors may want to consider renaming it. Also, given the sign of what the choice bias means could be flipped with a simple sign flip in the model equation (i.e., equating to accepting more vs accepting less offers), it would be helpful to show some basic plots to illustrate the identified differences (e.g., plotting the % accepted for people in the upper and lower tertile for the SHAPS score etc).*

We apologise that this was not made clear previously: the meaning and directionality of "choice bias" is indeed central to our results. We also thank the Reviewer for pointing out the previousely-used term "choice bias" itself might not be intuitive. We have now changed this to 'motivational tendency' (see below) as well as added substantial details on this parameter to the manuscript, including additional explanations and visualisations of the model as suggested by the Reviewer (new Figure 3) and model-agnostic results to aid interpretation (new Figure S3). Note the latter is complex due to our staircasing procedure (see new figure panel D further detailing our staircasing procedure in Figure 2). This shows that participants with more pronounced anhedonia are less likely to accept offers than those with low anhedonia (Fig. S3A), a model-agnostic version of our central result.

Our changes are detailed below:

After careful evaluation we have decided to term the parameter "motivational tendency", hoping that this will present a more intuitive description of the parameter.

To aid with the understanding and interpretation of the model parameters, and motivational tendency in particular, we have added the following explanation to the main text:

Lines 149 – 155:

"The models posit efforts and rewards are joined into a subjective value (SV), weighed by individual effort (and reward sensitivity (parameters. The subjective value is then integrated with an individual motivational tendency (a) parameter to guide decision-making. Specifically, the motivational tendency parameter determines the range at which subjective values are translated to acceptance probabilities: the same subjective value will translate to a higher acceptance probability the higher the motivational tendency."

Further, we have included a new figure, visualizing the model. This demonstrates how the different model parameters contribute to the model (A), and how different values on each parameter affects the model (B-D).

We agree that plotting model agnostic effects in our data may help the reader gain intuition of what our task results mean. We hope to address this with our added section on "Model agnostic task measures relating to questionnaires". We first followed the reviewer's suggestion of extracting subsamples with higher and low anhedonia (as measured with the

SHAPS, highest and lowest quantile) and plotted the acceptance proportion across effort and reward levels (panel A in figure below). However, due to our implemented task design, this only shows part of the picture: the staircasing procedure individualises which effort-reward combination a participant is presented with. Therefore, group differences in choice behaviour will lead to differences in the development of the staircases implemented in our task. Thus, we plotted the count of offered effort-reward combinations for the subsamples of participants with high vs. low SHAPS scores by the end of the task, averaged across staircases and participants.

As the aspect of task development due to the implemented staircasing may not have been explained sufficiently in the main text, we have included panel (D) in figure 2.

Further, we have added the following figure reference to the main text (lines 189 – 193):

"The development of offered effort and reward levels across trials is shown in figure 2D; this shows that as participants generally tend to accept challenges rather than reject them, the implemented staircasing procedure develops toward higher effort and lover reward challenges."

To statistically test effects of model-agnostic task measures on the neuropsychiatric questionnaires, we performed Bayesian GLMs with the proportion of accepted trials predicted by SHAPS and AES. This is reported in the text as follows.

Supplement, lines 172 – 189:

"To explore the relationship between model agnostic task measures to questionnaire measures of neuropsychiatric symptoms, we conducted Bayesian GLMs, with the proportion of accepted trials predicted by SHAPS scores, controlling for age and gender. The proportion of accepted trials averaged across effort and reward levels was predicted by the Snaith-Hamilton Pleasure Scale (SHAPS) sum scores (M=-0.07; 95%HDI=[-0.12,-0.03]) and the Apathy Evaluation Scale (AES) sum scores (M=-0.05; 95%HDI=[-0.10,-0.002]). Note that this was not driven only by higher effort levels; even confining data to the lowest two effort levels, SHAPS has a predictive value for the proportion of accepted trials: M=-0.05; 95%HDI=[-0.07,-0.02]. A visualisation of model agnostic task measures relating to symptoms is given in Fig. S4, comparing subgroups of participants scoring in the highest and lowest quartile on the SHAPS. This shows that participants with a high SHAPS score (i.e., more pronounced anhedonia) are less likely to accept offers than those with a low SHAPS score (Fig. S4A). Due to the implemented staircasing procedure, group differences can also be seen in the effort-reward combinations offered per trial. While for both groups, the staircasing procedure seems to devolve towards high effort – low reward offers, this is more pronounced in the subgroup of participants with a lower SHAPS score (Fig S4B)."

> *(3) None of the key effects relate to effort or reward sensitivity which is somewhat surprising given the previous literature and also means that it is hard to know if choice bias results would be equally found in tasks without any effort component. (The only analysis related to effort sensitivity is exploratory and in a subsample of N=56 per group looking at people meeting criteria for MDD vs matched controls.) Were stimuli constructed such that effort and reward sensitivity could be separated (i.e., are uncorrelated/orthogonal)? Maybe it would be worth looking at the % accepted in the largest or two largest effort value bins in an exploratory analysis. It seems the lowest and 2nd lowest effort level generally lead to accepting the challenge pretty much all the time, so including those effort levels might not be sensitive to individual difference analyses?*

We too were initially surprised by the lack of effect of neuropsychiatric symptoms on reward and effort sensitivity. To address the Reviewer's first comment, the nature of the 'choice bias' parameter (now motivational tendency) is its critical importance in the context of effort-

based decision-making: it is not modelled or measured explicitly in tasks without effort (such as typical reward tasks), so it would be impossible to test this in tasks without an effort component.

For the Reviewer's second comment, the exploratory MDD analysis is not our only one related to effort sensitivity: the effort sensitivity parameter is included in all of our central analyses, and (like reward sensitivity), does not relate to our measured neuropsychiatric symptoms (e.g., see page 15). Note most previous effort tasks do not include a 'choice bias'/motivational tendency parameter, potentially explaining this discrepancy. However, our model was quantitatively superior to models without this parameter, for example with only effort- and reward-sensitivity (page 11, Fig. 3).

Our three model parameters (reward sensitivity, effort sensitivity, and choice bias/motivational tendency) were indeed uncorrelated/orthogonal to one another (see parameter orthogonality analyses below), making it unlikely that the variance and effect captured by our motivational tendency parameter (previously termed "choice bias") should really be attributed to reward sensitivity. As per the Reviewer's suggestion, we also examined whether the lowest two effort levels might not be sensitive to individual differences; in fact, we found out proportion of accepted trials on the lowest effort levels alone was nevertheless predicted by anhedonia (see ceiling effect analyses below).

Specifically, in terms of parameter orthogonality:

When developing our task design and computational modelling approach we were careful to ensure that meaningful neurocomputational parameters could be estimated and that no spurious correlations between parameters would be introduced by modelling. By conducting parameter recoveries for all models, we showed that our modelling approach could reliably estimate parameters, and that estimated parameters are orthogonal to the other underlying parameters (as can be seen in Figure S1 in the supplement). It is thus unlikely that the variance and effect captured by our motivational tendency parameter (previously termed "choice bias") should really be attributed to reward sensitivity.

And finally, regarding the possibility of a ceiling effect for low effort levels:

We agree that visual inspection of the proportion of accepted results across effort and reward values can lead to the belief that a ceiling effect prevents the two lowest effort levels from capturing any inter-individual differences. To test whether this is the case, we ran a Bayesian GLM with the SHAPS sum score predicting the proportion of accepted trials (controlling for age and gender), in a subset of the data including only trials with an effort level of 1 or 2. We found the SHAPS has a predictive value for the proportion of accepted trials in the lowest two effort levels: M=-0.05; 95%HDI=[-0.07,-0.02]). This is noted in the text as follows.

Supplement, lines 175 – 180:

"The proportion of accepted trials averaged across effort and reward levels was predicted by the Snaith-Hamilton Pleasure Scale (SHAPS) sum scores (M=-0.07; 95%HDI=[-0.12,-0.03]) and the Apathy Evaluation Scale (AES) sum scores (M=-0.05; 95%HDI=[-0.10,-0.002]). Note that this was not driven only by higher effort levels; even confining data to the lowest two effort levels, SHAPS has a predictive value for the proportion of accepted trials: M=-0.05; 95%HDI=[-0.07,-0.02]."

> *(4) The abstract and discussion seem overstated (implications for the school system and statements on circadian rhythms which were not measured here). They should be toned down to reflect conclusions supported by the data.*

We thank the Reviewer for pointing this out, and have now removed these claims from the abstract and Discussion; we hope they now better reflect conclusions supported by these data

directly.

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations For The Authors):***
>
> *(1) Suggestions for improved or additional experiments, data or analyses.*
>
> *- For a non-computational audience, it would be useful to unpack the influence of the choice bias on behavior, as it is less clear how this would affect decision-making than sensitivity to effort or reward. Perhaps a figure showing accept/reject decisions when sensitivities are held and choice bias is high would be beneficial.*

We thank the Reviewer for suggesting additional explanations of the choice bias parameter to aid interpretation for non-computational readers; as per the Reviewer's suggestion, we have now included additional explanations and visualisations (Figure 3) to make this as clear as possible. Please note also that, in response to one of the other Reviewers and after careful considerations, we have decided to rename the "choice bias" parameter to "motivational tendency", hoping this will prove more intuitive.

To aid with the understanding and interpretation of this and the other model parameters, we have added the following explanation to the main text.

Lines 149 – 155:

"The models posit efforts and rewards are joined into a subjective value (SV), weighed by individual effort (and reward sensitivity (parameters. The subjective value is then integrated with an individual motivational tendency (a) parameter to guide decision-making. Specifically, the motivational tendency parameter determines the range at which subjective values are translated to acceptance probabilities: the same subjective value will translate to a higher acceptance probability the higher the motivational tendency."

Additionally, we add the following explanation to the Methods section.

Lines 698 – 709:

First, a cost function transforms costs and rewards associated with an action into a subjective value (SV):

$$SV = (\beta_R \cdot R) - (\beta_E \cdot E) \quad (1)$$

with β_R and β_E for reward and effort sensitivity, and $\mathcal{R}$ and E for reward and effort. Higher effort and reward sensitivity mean the SV is more strongly influenced by changes in effort and reward, respectively (Fig. 3B-C). Hence, low effort and reward sensitivity mean the SV, and with that decision-making, is less guided by effort and reward offers, as would be in random decision-making.

This SV is then transformed to an acceptance probability by a softmax function:

$$p(accept) = \frac{1}{1 + e^{-(\alpha + SV)}} \quad (2)$$

with $_{p(accept)}$ for the predicted acceptance probability and α for the intercept representing

motivational tendency. A high motivational tendency means a subjects has a tendency, or bias, to accept rather than reject offers (Fig. 3D).

Our new figure (panels A-D in figure 3) visualizes the model. This demonstrates how the different model parameters come at play in the model (A), and how different values on each parameter affects the model (B-D).

> *- The early and late chronotype groups have significant differences in ages and gender. Additional supplementary analysis here may mitigate any concerns from readers.*

The Reviewer is right to notice that our subsamples of early and late chronotypes differ significantly in age and gender, but it important to note that all our analyses comparing these two groups take this into account, statistically controlling for age and gender. We regret that this was previously only mentioned in the Methods section, so this information was not accessible where most relevant. To remedy this, we have amended the Results section as follows.

Lines 317 – 323:

“Bayesian GLMs, controlling for age and gender, predicting task parameters by time-of-day and chronotype showed effects of chronotype on reward sensitivity (i.e. those with a late chronotype had a higher reward sensitivity; M= 0.325, 95% HDI=[0.19,0.46]) and motivational tendency (higher in early chronotypes; M=-0.248, 95% HDI=[-0.37,-0.11]), as well as an interaction between chronotype and time-of-day on motivational tendency (M=0.309, 95% HDI=[0.15,0.48]).”

> *(2) Recommendations for improving the writing and presentation.*
>
> *- I found the term 'overlapping' a little jarring. I think the authors use it to mean both neuropsychiatric symptoms and chronotypes affect task parameters, but they are are not tested to be 'separable', nor is an interaction tested. Perhaps being upfront about how interactions are not being tested here (in the introduction, and not waiting until the discussion) would give an opportunity to operationalize this term.*

We agree with the Reviewer that our previously-used term “overlapping” was not ideal: it may have been misleading, and was not necessarily reflective of the nature of our findings. We now state explicitly that we are not testing an interaction between neuropsychiatric symptoms and chronotypes in our primary analyses. Additionally, following suggestions made by Reviewer 3, we ran new exploratory analyses to investigate how the effects of neuropsychiatric symptoms and circadian measures on motivational tendency relate to one another. These results in fact show that all three symptom measures have separable effects from circadian measures on motivational tendency. This supports the Reviewer’s view that ‘overlapping’ was entirely the wrong word—although it nevertheless shows the important contribution of circadian rhythm as well as neuropsychiatric symptoms in effort-based decision-making. We have changed the manuscript throughout to better describe this important, more accurate interpretation of our findings, including replacing the term “overlapping”. We changed the title from “Overlapping effects of neuropsychiatric symptoms and circadian rhythm on effort-based decision-making” to “Both neuropsychiatric symptoms and circadian rhythm alter effort-based decision-making”.

To clarify the intention of our primary analyses, we have added the following to the last paragraph of the introduction.

Lines 107 – 112:

“Next, we pre-registered a follow-up experiment to directly investigate how circadian preference interacts with time-of-day on motivational decision-making, using the same task and computational modelling approach. While this allows us to test how circadian effects on motivational decision-making compare to neuropsychiatric effects, we do not test for possible interactions between neuropsychiatric symptoms and chronobiology.”

We detail our new analyses in the Methods section as follows.

Lines 800 – 814:

“4.5.2 Differentiating between the effects of neuropsychiatric symptoms and circadian measures on motivational tendency

To investigate how the effects of neuropsychiatric symptoms on motivational tendency (2.3.1) relate to effects of chronotype and time-of-day on motivational tendency we conducted exploratory analyses. In the subsamples of participants with an early or late chronotype (including additionally collected data), we first ran Bayesian GLMs with neuropsychiatric questionnaire scores (SHAPS, DARS, AES respectively) predicting motivational tendency, controlling for age and gender. We next added an interaction term of chronotype and time-of-day into the GLMs, testing how this changes previously observed neuropsychiatric and circadian effects on motivational tendency. Finally, we conducted a model comparison using LOO, comparing between motivational tendency predicted by a neuropsychiatric questionnaire, motivational tendency predicted by chronotype and time-of-day, and motivational tendency predicted by a neuropsychiatric questionnaire and time-of-day (for each neuropsychiatric questionnaire, and controlling for age and gender).”

Results of the outlined analyses are reported in the Results section as follows.

Lines 356 – 383:

“2.5.2.1 Neuropsychiatric symptoms and circadian measures have separable effects on motivational tendency

Exploratory analyses testing for the effects of neuropsychiatric questionnaires on motivational tendency in the subsamples of early and late chronotypes confirmed the predictive value of the SHAPS (M=-0.24, 95% HDI=[-0.42,-0.06]), the DARS (M=-0.16, 95% HDI=[-0.31,-0.01]), and the AES (M=-0.18, 95% HDI=[-0.32,-0.02]) on motivational tendency.

For the SHAPS, we find that when adding the measures of chronotype and time-of-day back into the GLMs, the main effect of the SHAPS (M=-0.26, 95% HDI=[-0.43,-0.07]), the main effect of chronotype (M=-0.11, 95% HDI=[-0.22,-0.01]), and the interaction effect of chronotype and time-of-day (M=0.20, 95% HDI=[0.07,0.34]) on motivational tendency remain. Model comparison by LOOIC reveals motivational tendency is best predicted by the model including the SHAPS, chronotype and time-of-day as predictors, followed by the model including only the SHAPS. Note that this approach to model comparison penalizes models for increasing complexity.

Repeating these steps with the DARS, the main effect of the DARS is found numerically, but the 95% HDI just includes 0 (M=-0.15, 95% HDI=[-0.30,0.002]). The main effect of chronotype (M=-0.11, 95% HDI=[-0.21,-0.01]), and the interaction effect of chronotype and time-of-day (M=0.18, 95% HDI=[0.05,0.33]) on motivational tendency remain. Model comparison identifies the model including the DARS and circadian measures as the best model, followed by the model including only the DARS.

For the AES, the main effect of the AES is found (M=-0.19, 95% HDI=[-0.35,-0.04]). For the main effect of chronotype, the 95% narrowly includes 0 (M=-0.10, 95% HDI=[-0.21,0.002]), while the interaction effect of chronotype and time-of-day (M=0.20, 95% HDI=[0.07,0.34]) on motivational tendency remains. Model comparison identifies the model including the AES and circadian measures as the best model, followed by the model including only the AES."

In addition to the title change, we edited our Discussion to discuss and reflect these new insights, including the following.

Lines 399 – 402:

"Various neuropsychiatric disorders are marked by disruptions in circadian rhythm, such as a late chronotype. However, research has rarely investigated how transdiagnostic mechanisms underlying neuropsychiatric conditions may relate to inter-individual differences in circadian rhythm."

Lines 475 – 480:

"It is striking that the effects of neuropsychiatric symptoms on effort-based decision-making largely are paralleled by circadian effects on the same neurocomputational parameter. Exploratory analyses predicting motivational tendency by neuropsychiatric symptoms and circadian measures simultaneously indicate the effects go beyond recapitulating each other, but rather explain separable parts of the variance in motivational tendency."

Lines 528 – 532:

"Our reported analyses investigating neuropsychiatric and circadian effects on effort-based decision-making simultaneously are exploratory, as our study design was not ideally set out to examine this. Further work is needed to disentangle separable effects of neuropsychiatric and circadian measures on effort-based decision-making."

Lines 543 – 550:

"We demonstrate that neuropsychiatric effects on effort-based decision-making are paralleled by effects of circadian rhythm and time-of-day. Exploratory analyses suggest these effects account for separable parts of the variance in effort-based decision-making. It unlikely that effects of neuropsychiatric effects on effort-based decision-making reported here and in previous literature are a spurious result due to multicollinearity with chronotype. Yet, not accounting for chronotype and time of testing, which is the predominant practice in the field, could affect results."

> *- A minor point, but it could be made clearer that many neurotransmitters have circadian rhythms (and not just dopamine).*

We agree this should have been made clearer, and have added the following to the Introduction.

Lines 83 – 84:

"Bi-directional links between chronobiology and several neurotransmitter systems have been reported, including dopamine47.

(47) Kiehn, J.-T., Faltraco, F., Palm, D., Thome, J. & Oster, H. Circadian Clocks in the Regulation of Neurotransmitter Systems. Pharmacopsychiatry 56, 108–117 (2023)."

> *- Making reference to other studies which have explored circadian rhythms in cognitive tasks would allow interested readers to explore the broader field. One such paper is:*

> *Bedder, R. L., Vaghi, M. M., Dolan, R. J., & Rutledge, R. B. (2023). Risk taking for potential losses but not gains increases with time of day. Scientific reports, 13(1), 5534, which also includes references to other similar studies in the discussion.*

We thank the Reviewer for pointing out that we failed to cite this relevant work. We have now included it in the Introduction as follows.

Lines 97 – 98:

"A circadian effect on decision-making under risk is reported, with the sensitivity to losses decreasing with time-of-day66.

(66) Bedder, R. L., Vaghi, M. M., Dolan, R. J. & Rutledge, R. B. Risk taking for potential losses but not gains increases with time of day. Sci Rep 13, 5534 (2023)."

> *(3) Minor corrections to the text and figures.*
>
> *None, clearly written and structured. Figures are high quality and significantly aid understanding.*
>
> ***Reviewer #2 (Recommendations For The Authors):***
>
> *I did have a few more minor comments:*
>
> *- The manuscript doesn't clarify whether trials had time limits - so that participants might fail to earn points - or instead they did not and participants had to continue exerting effort until they were done. This is important to know since it impacts on decision-strategies and behavioral outcomes that might be analyzed. For example, if there is no time limit, it might be useful to examine the amount of time it took participants to complete their effort - and whether that had any relationship to choice patterns or symptomatology. Or, if they did, it might be interesting to test whether the relationship between choices and exerted effort depended on symptoms. For example, someone with depression might be less willing to choose effort, but just as, if not more likely to successfully complete a trial once it is selected.*

We thank the Reviewer for pointing out this important detail in the task design, which we should have made clearer. The trials did indeed have a time limit which was dependent on the effort level. To clarify this in the manuscript, we have made changes to Figure 2 and the Methods section. We agree it would be interesting to explore whether the exerted effort in the task related to symptoms. We explored this in our data by predicting the participant average proportion of accepted but failed trials by SHAPS score (controlling for age and gender). We found no relationship: M=0.01, 95% HDI=[-0.001,0.02]. However, it should be noted that the measure of proportion of failed trials may not be suitable here, as there are only few accepted but failed trials (M = 1.3% trials failed, SD = 3.50). This results from several task design characteristics aimed at preventing subjects from failing accepted trials, to avoid confounding of effort discounting with risk discounting. As an alternative measure, we explored the extent to which participants went "above and beyond" the target in accepted trials. Specifically, considering only accepted and succeeded trials, we computed the factor by which the required number of clicks was exceeded (i.e., if a subject clicked 15 times when 10 clicks were required the factor would be 1.3), averaging across effort and reward level. We then conducted a Bayesian GLM to test whether this subject wise click-exceedance measure can be predicted by apathy or anhedonia, controlling for age and gender. We found neither the SHAPS (M=-0.14, 95% HDI=[-0.43,0.17]) nor the AES (M=0.07, 95% HDI=[-0.26,0.41]) had a predictive value for the amount to which subjects exert "extra effort". We have now added this to the manuscript.

In Figure 2, which explains the task design in the results section, we have added the following to the figure description.

Lines 161 – 165:

“Each trial consists of an offer with a reward (2,3,4, or 5 points) and an effort level (1,2,3, or 4, scaled to the required clicking speed and time the clicking must be sustained for) that subjects accept or reject. If accepted, a challenge at the respective effort level must be fulfilled for the required time to win the points.”

In the Methods section, we have added the following.

Lines 617 – 622:

“We used four effort-levels, corresponding to a clicking speed at 30% of a participant’s maximal capacity for 8 seconds (level 1), 50% for 11 seconds (level 2), 70% for 14 seconds (level 3), and 90% for 17 seconds (level 4). Therefore, in each trial, participants had to fulfil a certain number of mouse clicks (dependent on their capacity and the effort level) in a specific time (dependent on the effort level).”

In the Supplement, we have added the additional analyses suggested by the Reviewer.

Lines 195 – 213:

“3.2 Proportion of accepted but failed trials

For each participant, we computed the proportion of trial in which an offer was accepted, but the required effort then not fulfilled (i.e., failed trials). There was no relationship between average proportion of accepted but failed trials and SHAPS score (controlling for age and gender): M=0.01, 95% HDI=[-0.001,0.02]. However, there are intentionally few accepted but failed trials (M = 1.3% trials failed, SD = 3.50). This results from several task design characteristics aimed at preventing subjects from failing accepted trials, to avoid confounding of effort discounting with risk discounting.”

“3.3 Exertion of “extra effort”

We also explored the extent to which participants went “above and beyond” the target in accepted trials. Specifically, considering only accepted and succeeded trials, we computed the factor by which the required number of clicks was exceeded (i.e., if a subject clicked 15 times when 10 clicks were required the factor would be 1.3), averaging across effort and reward level. We then conducted a Bayesian GLM to test whether this subject wise click-exceedance measure can be predicted by apathy or anhedonia, controlling for age and gender. We found neither the SHAPS (M=-0.14, 95% HDI=[-0.43,0.17]) nor the AES (M=0.07, 95% HDI=[-0.26,0.41]) had a predictive value for the amount to which subjects exert “extra effort”.”

> *- Perhaps relatedly, there is evidence that people with depression show less of an optimism bias in their predictions about future outcomes. As such, they show more "rational" choices in probabilistic decision tasks. I'm curious whether the Authors think that a weaker choice bias among those with stronger depression/anhedonia/apathy might be related. Also, are choices better matched with actual effort production among those with depression?*

We think this is a very interesting comment, but unfortunately feel our manuscript cannot properly speak to it: as in our response to the previous comment, our exploratory analysis linking the proportion of accepted but failed trials to anhedonia symptoms (i.e. less anhedonic people making more optimistic judgments of their likelihood of success) did not show a relationship between the two. However, this null finding may be the result of our task

design which is not laid out to capture such an effect (in fact to minimize trials of this nature). We have added to the Discussion section.

Lines 442 – 445:

“It is possible that a higher motivational tendency reflects a more optimistic assessment of future task success, in line with work on the optimism bias95; however our task intentionally minimized unsuccessful trials by titrating effort and reward; future studies should explore this more directly.

(95) Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R. & Dolan, R. J. Depression is related to an absence of optimistically biased belief updating about future life events. Psychological Medicine 44, 579–592 (2014).”

> *- The manuscript does not clarify: How did the Authors ensure that each subject received each effort-reward combination at least once if a given subject always accepted or always rejected offers?*

We have made the following edit to the Methods section to better explain this aspect of our task design.

Lines 642 – 655:

“For each subject, trial-by-trial presentation of effort-reward combinations were made semi-adaptively by 16 randomly interleaved staircases. Each of the 16 possible offers (4 effort-levels x 4 reward-levels) served as the starting point of one of the 16 staircase. Within each staircase, after a subject accepted a challenge, the next trial’s offer on that staircase was adjusted (by increasing effort or decreasing reward). After a subject rejected a challenge, the next offer on that staircase was adjusted by decreasing effort or increasing reward. This ensured subjects received each effort-reward combination at least once (as each participant completed all 16 staircases), while individualizing trial presentation to maximize the trials’ informative value. Therefore, in practice, even in the case of a subject rejecing all offers (and hence the staircasing procedures always adapting by decreasing effort or increasing reward), the full range of effort-reward combinations will be represented in the task across the startingpoints of all staircases (and therefore before adaption takeplace).”

> *- The word "metabolic" is misspelled in Table 1*
>
> *- Figure 2 is missing panel label "C"*
>
> *- The word "effort" is repeated on line 448.*

We thank the Reviewer for their attentive reading of our manuscript and have corrected the mistakes mentioned.

> ***Reviewer #3 (Recommendations For The Authors):***
>
> *It is a bit difficult to get a sense of people's discounting from the plots provided. Could the authors show a few example individuals and their fits (i.e., how steep was effort discounting on average and how much variance was there across individuals; maybe they could show the mean discount function or some examples etc)*

We appreciate very much the Reviewer's suggestion to visualise our parameter estimates within and across individuals. We have implemented this in Figure .S2

> *It would be helpful if correlations between the various markers used as dependent variables (SHAPS, DARS, AES, chronotype etc) could plotted as part of each related figure*

> *(e.g., next to the relevant effects shown).*

We agree with the Reviewer that a visual representation of the various correlations between dependent variables would be a better and more assessable communication than our current paragraph listing the correlations. We have implemented this by adding a new figure plotting all correlations in a heat map, with asterisks indicating significance.

> *The authors use the term "meaningful relationship" - how is this defined? If undefined, maybe consider changing (do they mean significant?)*

We understand how our use of the term “(no) meaningful relationship” was confusing here. As we conducted most analyses in a Bayesian fashion, this is a formal definition of ‘meaningful’: the 95% highest density interval does not span across 0. However, we do not want this to be misunderstood as frequentist “significance” and agree clarity can be improved here, To avoid confusion, we have amended the manuscript where relevant (i.e., we now state “we found a (/no) relationship / effect” rather than “we found a meaningful relationship”.

> *The authors do not include an inverse temperature parameter in their discounting models-can they motivate why? If a participant chose nearly randomly, which set of parameter values would they get assigned?*

Our decision to not include an inverse temperature parameter was made after an extensive simulation-based investigation of different models and task designs. A series of parameter recovery studies including models with an inverse temperature parameter revealed the inverse temperature parameter could not be distinguished from the reward sensitivity parameter. Specifically, inverse temperature seemed to capture the variance of the true underlying reward sensitivity parameter, leading to confounding between the two. Hence, including both reward sensitivity and inverse temperature would not have allowed us to reliably estimate either parameter. As our pre-registered hypotheses related to the reward sensitivity parameter, we opted to include models with the reward sensitivity parameter rather than the inverse temperature parameter in our model space. We have now added these simulations to our supplement.

Nevertheless, we believe our models can capture random decision-making. The parameters of effort and reward sensitivity capture how sensitive one is to changes in effort/reward level. Hence, random decision-making can be interpreted as low effort and reward sensitivity, such that one’s decision-making is not guided by changes in effort and reward magnitude. With low effort/reward sensitivity, the motivational tendency parameter (previously “choice bias”) would capture to what extend this random decision-making is biased toward accepting or rejecting offers.

The simulation results are now detailed in the Supplement.

Lines 25 – 46:

“1.2.1 Parameter recoveries including inverse temperature

In the process of task and model space development, we also considered models incorportating an inverse temperature paramater. To this end, we conducted parameter recoveries for four models, defined in Table S3.

Parameter recoveries indicated that, parameters can be recovered reliably in model 1, which

includes only effort sensitivity (β_E) and inverse temperature (τ) as free parameters (on-

diagonal correlations: .98 > r > .89, off-diagonal correlations: .04 > |r| > .004). However, as a reward sensitivity parameter is added to the model (model 2), parameter recovery seems to be compromised, as parameters are estimated less accurately (on-diagonal correlations: .80 > r > .68), and spurious correlations between parameters emerge (off-diagonal correlations: .40 > |r| > .17). This issue remains when motivational tendency is added to the model (model 4; on-diagonal correlations: .90 > r > .65; off-diagonal correlations: .28 > |r| > .03), but not when inverse temperature is modelled with effort sensitivity and motivational tendency, but not reward sensitivity (model 3; on-diagonal correlations: .96 > r > .73; off-diagonal correlations: .05 > |r| > .003).

As our pre-registered hypotheses related to the reward sensitivity parameter, we opted to include models with the reward sensitivity parameter rather than the inverse temperature parameter in our model space."

And we now discuss random decision-making specifically in the Methods section.

Lines 698 – 709:

"First, a cost function transforms costs and rewards associated with an action into a subjective value (SV):

$$SV = (\beta_R \cdot R) - (\beta_E \cdot E) \quad (1)$$

with β_R and β_E for reward and effort sensitivity, and and for reward and effort. Higher effort and reward sensitivity mean the SV is more strongly influenced by changes in effort and reward, respectively (Fig. 3B-C). Hence, low effort and reward sensitivity mean the SV, and with that decision-making, is less guided by effort and reward offers, as would be in random decision-making.

This SV is then transformed to an acceptance probability by a softmax function:

$$p(accept) = \frac{1}{1 + e^{-(\alpha + SV)}} \quad (2)$$

with $p(accept)$ for the predicted acceptance probability and for the intercept representing motivational tendency. A high motivational tendency means a subjects has a tendency, or bias, to accept rather than reject offers (Fig. 3D)."

> *The pre-registration mentions effects of BMI and risk of metabolic disease-those are briefly reported the in factor loadings, but not discussed afterwards-although the authors stated hypotheses regarding these measures in their preregistration. Were those hypotheses supported?*

We reported these results (albeit only briefly) in the factor loadings resulting from our PLS regression and results from follow-up GLMs (see below). We have now amended the Discussion to enable further elaboration on whether they confirmed our hypotheses (this evidence was unclear, but we have subsequently followed up in a sample with type-2 diabetes, who also show reduced motivational tendency).

Lines 258 – 261:

"For the MEQ (95%HDI=[-0.09,0.06]), MCTQ (95%HDI=[-0.17,0.05]), BMI (95%HDI=[-0.19,0.01]), and FINDRISC (95%HDI=[-0.09,0.03]) no relationship with motivational tendency was found, consistent with the smaller magnitude of reported component loadings from the PLS regression."

We have added the following paragraph to our discussion.

Lines 491 – 502:

"To our surprise, we did not find statistical evidence for a relationship between effort-based decision-making and measures of metabolic health (BMI and risk for type-2 diabetes). Our analyses linking BMI to motivational tendency reveal a numeric effect in line with our hypothesis: a higher BMI relating to a lower motivational tendency. However, the 95% HDI for this effect narrowly included zero (95%HDI=[-0.19,0.01]). Possibly, our sample did not have sufficient variance in metabolic health to detect dimensional metabolic effects in a current general population sample. A recent study by our group investigates the same neurocomputational parameters of effort-based decision-making in participants with type-2 diabetes and non-diabetic controls matched by age, gender, and physical activity105. We report a group effect on the motivational tendency parameter, with type-2 diabetic patients showing a lower tendency to exert effort for reward."

"(105) Mehrhof, S. Z., Fleming, H. A. & Nord, C. A cognitive signature of metabolic health in effort-based decision-making. Preprint at https://doi.org/10.31234/osf.io/4bkm9 (2024)."

> *R-values are indicated as a range (e.g., from 0.07-0.72 for the last one in 2.1 which is a large range). As mentioned above, the full correlation matrix should be reported in figures as heatmaps.*

We agree with the Reviewer that a heatmap is a better way of conveying this information – see Figure 1 in response to their previous comment.

> *The answer on whether data was already collected is missing on the second preregistration link. Maybe this is worth commenting on somewhere in the manuscript.*

This question appears missing because, as detailed in the manuscript, we felt that technically some data *was* already collected by the time our second pre-registration was posted. This is because the second pre-registration detailed an additional data collection, with the goal of extending data from the original dataset to include extreme chronotypes and increase precision of analyses. To avoid any confusion regarding the lack of reply to this question in the pre-registration, we have added the following disclaimer to the description of the second pre-registration:

"Please note the lack of response to the question regarding already collected data. This is because the data collection in the current pre-registration extends data from the original dataset to increase the precision of analyses. While this original data is already collected, none of the data collection described here has taken place."

> *Some referencing is not reflective of the current state of the field (e.g., for effort discounting: Sugiwaka et al., 2004 is cited). There are multiple labs that have published on this since then including Philippe Tobler's and Sven Bestmann's groups (e.g., Hartmann et al., 2013; Klein-Flügge et al., Plos CB, 2015).*

We agree absolutely, and have added additional, more recent references on effort discounting.

Lines 67 – 68:

“Higher costs devalue associated rewards, an effect referred to as effort-discounting33–37.”

(33) Sugiwaka, H. & Okouchi, H. Reformative self-control and discounting of reward value by delay or effort1. Japanese Psychological Research 46, 1–9 (2004).

(34) Hartmann, M. N., Hager, O. M., Tobler, P. N. & Kaiser, S. Parabolic discounting of monetary rewards by physical effort. Behavioural Processes 100, 192–196 (2013).

(35) Klein-Flügge, M. C., Kennerley, S. W., Saraiva, A. C., Penny, W. D. & Bestmann, S. Behavioral Modeling of Human Choices Reveals Dissociable Effects of Physical Effort and Temporal Delay on Reward Devaluation. PLOS Computational Biology 11, e1004116 (2015).

(36) Białaszek, W., Marcowski, P. & Ostaszewski, P. Physical and cognitive effort discounting across different reward magnitudes: Tests of discounting models. PLOS ONE 12, e0182353 (2017).

(37) Ostaszewski, P., Bąbel, P. & Swebodziński, B. Physical and cognitive effort discounting of hypothetical monetary rewards. Japanese Psychological Research 55, 329–337 (2013).

> *There are lots of typos throughout (e.g., Supplementary martial, Mornignness etc)*

We thank the Reviewer for their attentive reading of our manuscript and have corrected our mistakes.

> *In Table 1, it is not clear what the numbers given in parentheses are. The figure note mentions SD, IQR, and those are explicitly specified for some rows, but not all.*

After reviewing Table 1 we understand the comment regarding the clarity of the number in parentheses. In our original manuscript, for some variables, numbers were given per category (e.g. for gender and ethnicity), rather than per row, in which case the parenthetical statistic was indicated in the header row only. However, we now see that the clarity of the table would have been improved by adding the reported statistic for each row—we have corrected this.

> *In Figure 1C, it would be much more helpful if the different panels were combined into one single panel (using differently coloured dots/lines instead of bars).*

We agree visualizing the proportion of accepted trials across effort and reward levels in one single panel aids interpretability. We have implemented it in the following plot (now Figure 2C).

> *In Sections 2.2.1 and 4.2.1, the authors mention "mixed-effects analysis of variance (ANOVA) of repeated measures" (same in the preregistration). It is not clear if this is a standard RM-ANOVA (aggregating data per participant per condition) or a mixed-effects model (analysing data on a trial-by-trial level). This model seems to only include within-subjects variable, so it isn't a "mixed ANOVA" mixing within and between subjects effects.*

We apologise that our use of the term "mixed-effects analysis of variance (ANOVA) of repeated measures" is indeed incorrectly applied here. We aggregate data per participant and effort-by-reward combination, meaning there are no between-subject effects tested. We have corrected this to “repeated measures ANOVA”.

> *In Section 2.2.2, the authors write "R-hats>1.002" but probably mean "R-hats < 1.002". ESS is hard to evaluate unless the total number of samples is given.*

We thank the Reviewer for noticing this mistake and have corrected it in the manuscript.

> *In Section 2.3, the inference criterion is unclear. The authors first report "factor loadings" and then perform a permutation test that is not further explained. Which of these factors are actually needed for predicting choice bias out of chance? The permutation test suggests that the null hypothesis is just "none of these measures contributes anything to predicting choice bias", which is already falsified if only one of them shows an association with choice bias. It would be relevant to know for which measures this is the case. Specifically, it would be relevant to know whether adding circadian measures into a model that already contains apathy/anhedonia improves predictive performance.*

We understand the Reviewer's concerns regarding the detail of explanation we have provided for this part of our analysis, but we believe there may have been a misunderstanding regarding the partial least squares (PLS) regression. Rather than identifying a number of factors to predict the outcome variable, a PLS regression identifies a model with one or multiple components, with various factor loadings of differing magnitude. In our case, the PLS regression identified a model with one component to best predict our outcome variable (motivational tendency, which in our previous various we called choice bias). This one component had factor loadings of our questionnaire-based measures, with measures of apathy and anhedonia having highest weights, followed by lesser weighted factor loadings by measures of circadian rhythm and metabolic health. The permutation test tests whether this component (consisting of the combination of factor loadings) can predict the outcome variable out of sample.

We hope we have improved clarity on this in the manuscript by making the following edits to the Results section.

Lines 248 – 251:

"Permutation testing indicated the predictive value of the resulting component (with factor loadings described above) was significant out-of-sample (root-mean-squared error [RMSE]=0.203, p=.001)."

Further, we hope to provide a more in-depth explanation of these results in the Methods section.

Lines 755 – 759:

"Statistical significance of obtained effects (i.e., the predictive accuracy of the identified component and factor loadings) was assessed by permutation tests, probing the proportion of root-mean-squared errors (RMSEs) indicating stronger or equally strong predictive accuracy under the null hypothesis."

> *In Section 2.5, the authors simply report "that chronotype showed effects of chronotype on reward sensitivity", but the direction of the effect (higher reward sensitivity in early vs. late chronotype) remains unclear.*

We thank the Reviewer for pointing this out. While we did report the direction of effect, this was only presented in the subsequent parentheticals and could have been made much clearer. To assist with this, we have made the following addition to the text.

Lines 317 – 320:

“Bayesian GLMs, controlling for age and gender, predicting task parameters by time-of-day and chronotype showed effects of chronotype on reward sensitivity (i.e. those with a late chronotype had a higher reward sensitivity; M= 0.325, 95% HDI=[0.19,0.46])”

> *In Section 4.2, the authors write that they "implemented a previously-described procedure using Prolific pre-screeners", but no reference to this previous description is given.*

We thank the Reviewer for bringing our attention to this missing reference, which has now been added to the manuscript.

> *In Supplementary Table S2, only the "on-diagonal correlations" are given, but off-diagonal correlations (indicative of trade-offs between parameters) would also be informative.*

We agree with the Reviewer that off-diagonal correlations between underlying and recovered parameters are crucial to assess confounding between parameters during model estimation. We reported this in figure S1D, where we present the full correlation matric between underlying and recovered parameters in a heatmap. We have now noticed that this plot was missing axis labels, which have been added now.

> *I found it somewhat difficult to follow the results section without having read the methods section beforehand. At the beginning of the Results section, could the authors briefly sketch the outline of their study? Also, given they have a pre-registration, could the authors introduce each section with a statement of what they expected to find, and close with whether the data confirmed their expectations? In the current version of the manuscript, many results are presented without much context of what they mean.*

We agree a brief outline of the study procedure before reporting the results would be beneficial to following the subsequently text and have added the following to the end of our Introduction.

Lines 101 – 106:

“Here, we tested the relationship between motivational decision-making and three key neuropsychiatric syndromes: anhedonia, apathy, and depression, taking both a transdiagnostic and categorical (diagnostic) approach. To do this, we validate a newly developed effort-expenditure task, designed for online testing, and gamified to increase engagement. Participants completed the effort-expenditure task online, followed by a series of self-report questionnaires.”

We have added references to our pre-registered hypotheses at multiple points in our manuscript.

Lines 185 – 187:

“In line with our pre-registered hypotheses, we found significant main effects for effort ($F(1,14367)=4961.07$, $p<.0001$) and reward ($F(1,14367)=3037.91$, $p<.001$), and a significant interaction between the two ($F(1,14367)=1703.24$, $p<.001$).”

Lines 215 – 221:

“Model comparison by out-of-sample predictive accuracy identified the model implementing three parameters (motivational tendency a, reward sensitivity β_R, and effort sensitivity β_E),

with a parabolic cost function (subsequently referred to as the full parabolic model) as the winning model (leave-one-out information criterion [LOOIC; lower is better] = 29734.8; expected log posterior density [ELPD; higher is better] = -14867.4; Fig. 31ED). This was in line with our pre-registered hypotheses."

Lines 252 – 258:

"Bayesian GLMs confirmed evidence for psychiatric questionnaire measures predicting motivational tendency (SHAPS: M=-0.109; 95% highest density interval (HDI)=[-0.17,-0.04]; AES: M=-0.096; 95%HDI=[-0.15,-0.03]; DARS: M=-0.061; 95%HDI=[-0.13,-0.01]; Fig. 4A). Post-hoc GLMs on DARS sub-scales showed an effect for the sensory subscale (M=-0.050; 95%HDI=[-0.10,-0.01]). This result of neuropsychiatric symptoms predicting a lower motivational tendency is in line with our pre-registered hypothesis."

Lines 258 – 263:

"For the MEQ (95%HDI=[-0.09,0.06]), MCTQ (95%HDI=[-0.17,0.05]), BMI (95%HDI=[-0.19,0.01]), and FINDRISC (95%HDI=[-0.09,0.03]) no meaningful relationship with choice biasmotivational tendency was found, consistent with the smaller magnitude of reported component loadings from the PLS regression. This null finding for dimensional measures of circadian rhythm and metabolic health was not in line with our pre-registered hypotheses."

Lines 268 – 270:

"For reward sensitivity, the intercept-only model outperformed models incorporating questionnaire predictors based on RMSE. This result was not in line with our pre-registered expectations."

Lines 295 – 298:

"As in our transdiagnostic analyses of continuous neuropsychiatric measures (Results 2.3), we found evidence for a lower motivational tendency parameter in the MDD group compared to HCs (M=-0.111, 95% HDI=[-0.20,-0.03]) (Fig. 4B). This result confirmed our pre-registered hypothesis."

Lines 344 – 355:

"Late chronotypes showed a lower motivational tendency than early chronotypes (M=-0.11, 95% HDI=[-0.22,-0.02])—comparable to effects of transdiagnostic measures of apathy and anhedonia, as well as diagnostic criteria for depression. Crucially, we found motivational tendency was modulated by an interaction between chronotype and time-of-day (M=0.19, 95% HDI=[0.05,0.33]): post-hoc GLMs in each chronotype group showed this was driven by a time-of-day effect within late, rather than early, chronotype participants (M=0.12, 95% HDI=[0.02,0.22], such that late chronotype participants showed a lower motivational tendency in the morning testing sessions, and a higher motivational tendency in the evening testing sessions; early chronotype: 95% HDI=[-0.16,0.04]) (Fig. 5A). These results of a main effect and an interaction effect of chronotype on motivational tendency confirmed our pre-registered hypothesis."

Lines 390 – 393:

"Participants with an early chronotype had a lower reward sensitivity parameter than those with a late chronotype (M=0.27, 95% HDI=[0.16,0.38]). We found no effect of time-of-day on reward sensitivity (95%HDI=[-0.09,0.11]) (Fig. 5B). These results were in line with our pre-registered hypotheses."

https://doi.org/10.7554/eLife.96803.2.sa0