Cognitive phenotypes of risk and resilience for the relationship between social media and mental health

Lukas J. Gunschera[1], Michelle Achterberg[2], Lydia Gabriela Speyer[3], Toby Wise[4], Amy Orben[1]

[1] MRC Cognition and Brain Sciences Unit, University of Cambridge, UK
[2] Department of Psychology, Education and Child Studies, Erasmus School of Social and Behavioral Sciences, Erasmus University Rotterdam, the Netherlands
[3] Early Life Care Institute, Paracelsus Medical University Salzburg, Austria
[4] Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London

Author Note

Correspondence concerning this article should be addressed to Lukas J. Gunschera, MRC Cognition and Brain Sciences Unit, University of Cambridge. Email: lukas.gunschera@mrc-cbu.cam.ac.uk.

Acknowledgements: L.J.G. was funded by the Medical Research Council PhD scholarship (G116768) and Wolfson College Cambridge (ID1061400). L.J.G and A.O. were funded by the Medical Research Council (MC_UU_00030/13), the Jacobs Foundation, and a UKRI Future Leaders Fellowship (MR/X034925/1). M.A. was funded by the Netherlands Organization for Scientific Research (NWO-VENI). T.W. was supported by a Career Development Award from the Wellcome Trust (225945/Z/22/Z).

Conflicts of interest: The authors have no conflicts of interest to disclose.

Abstract

Research on social media's impact on adolescent mental health has yielded inconsistent findings, and many demographic moderators have been investigated to little effect. We move beyond such comparisons to examine how cognitive processes might shape this relationship. In a pre-registered longitudinal study, we applied computational modelling to repeated decision-making tasks to test whether preference for immediate rewards moderates the longitudinal relationship between social media use and mental health. Adolescents with a greater preference for immediate rewards exhibited larger declines in behavioural activation as social media use increased. This effect was consistent across timepoints, indicating heightened vulnerability among those favouring immediate gratification. No moderation was observed for other mental health outcomes, and delay discounting was unrelated to time spent on social media or compulsive internet use. Cognitive processes could therefore provide insights into individual differences in social media's mental health effects, offering future mechanistic explanations and potential targets for intervention.

Introduction

Social media shapes the lives of many adolescents, and the growing concerns about the psychological implications of this development outpace the provision of conclusive scientific evidence (Orben & Matias, 2025). Across the population, studies tend to find small negative correlations between social media use and mental health and wellbeing (Fassi et al., 2024; Ivie et al., 2020; Keles et al., 2020). However, the evidence suggests that these associations differ markedly across individuals (Valkenburg et al., 2021, 2024). Subsequent efforts have therefore attempted to determine which individual differences (enduring characteristics that distinguish people from one another) predict risk or resilience to the effects of social media. This includes, for example, age (Crone & Konijn, 2018; Orben et al., 2022; Orben & Blakemore, 2023), gender (Booker et al., 2018; Coyne et al., 2023), and socioeconomic status (Kurten et al., 2025).

Despite theoretical reasons why high-level demographic differences, such as age, might moderate the relationship between social media use and mental health, empirical support for these moderations has been mixed (Booker et al., 2018; Cunningham et al., 2021; Orben et al., 2022; Vahedi & Zannella, 2021). This inconsistency may stem from the predominant focus on broad demographic categories as isolated predictors, overlooking how these interact with the underlying psychological and biological processes that drive cognition and behaviour.

Individual differences likely operate hierarchically across three interconnected levels: broad macro-level factors such as demographics (e.g. age), nuanced meso-level cognitive mechanisms (reward learning processes that determine how individuals learn from rewards), and fine-grained micro-level biological processes that implement these cognitive mechanisms in the brain or body (e.g. neural networks)(see Fig. 1). Macro-level factors tend to be structural or demographic in nature and are often measured via self-report or administrative data, which provide a high-level lens for examining population differences. Meso-level factors, by contrast, capture the cognitive, behavioural and biological mechanisms that vary both between individuals and within individuals over time, determining how individuals process information, make decisions, and respond to their environment. Micro-level factors represent the biological implementations of these processes. Crucially, these levels are interconnected, and meso-level mechanisms may covary with higher-order macro-level differences. For example, age-related differences in the association between social media use and mental health may be explained by underlying differences in reward sensitivity – a meso-level cognitive mechanism that changes across development and determines how individuals pursue rewards in their environment.

This hierarchical framework may help explain why previous research, which has predominantly examined macro-level individual differences, has produced heterogeneous findings. Macro-level measures may obscure meaningful patterns by aggregating distinct meso-level processes that differentially relate to outcomes of interest. This pattern is evidence in clinical research, where findings have suggested that while age and gender (macro-level factors) fail to predict antidepressant treatment outcomes, hypothalamic-pituitary-adrenal axis functioning – a meso-level neurocognitive process – helps predict patient treatment response (Ising et al., 2007). Similarly, in social media research, age alone may mask important differences in underlying cognitive processes and their influence on the effects of social media use on mental health. This illustrates the importance of targeting lower-level (i.e. micro- and meso-level) for understanding, predicting, and treating the complex link between social media use and mental health outcomes.

Research across the mental health field has largely recognised the importance of meso-level processes for effective mechanistic research. Within the field of computational psychiatry, researchers use mathematical models of cognition and behaviour to formalise and test the meso-level mechanisms that underpin mental health, enabling more precise quantification of these nuanced processes. This mathematical approach has enabled researchers to examine detailed neurocognitive processes that are implicated in the onset, maintenance, and treatment of mental distress (Browning et al., 2023; Montague et al., 2012; Stern et al., 2018; Wise et al., 2023).

Building on these advances from mental health research, a cognitive-computational approach can potentially provide a valuable lens for examining the nuanced meso-level factors of risk and resilience for the relationship between social media use and mental health. Cognitive processes – including impulsivity, reward learning, and attention – shape a) how individuals seek and process information (Kelly & Sharot, 2021; Sharot & Sunstein, 2020), b) how individuals interpret and learn from their experiences (Poulton & Hester, 2020), and c) how these experiences relate to downstream emotional and behavioural consequences (Amlung et al., 2019;

Pike & Robinson, 2022). It is eminently plausible that these influences are not limited to offline environments and that cognitive processes systematically shape how social media affects individuals. Therefore, a meso-level approach investigating cognitive individual differences in risk and resilience can potentially provide an effective framework for studying the impacts of social media on mental health.

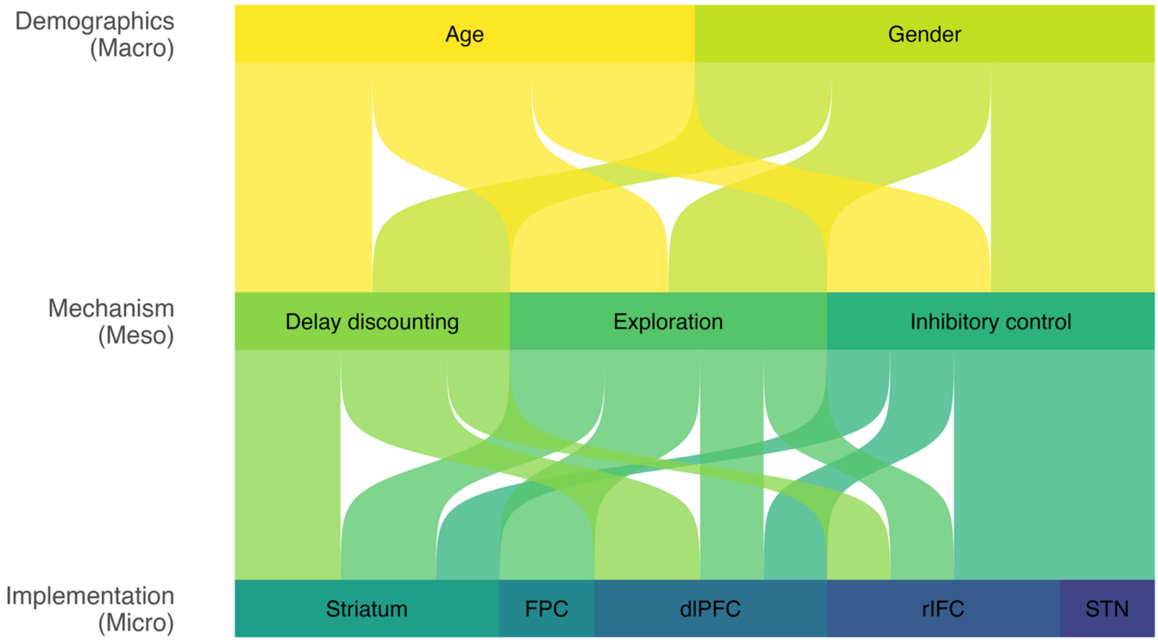

Fig. 1: Schematic Illustration of Hierarchical Framework of Individual Differences Operating Across Micro-, Meso-, and Macro-levels. *This schematic diagram maps out the hierarchical and interconnected nature of the macro-, meso-, and micro-levels. We present a simplified subset of relevant variables to illustrate the framework. In reality, the hierarchical structure encompasses many more macro-level factors, each underpinned by multiple meso-level processes, which in turn are implemented through numerous micro-level neural mechanisms. The strength of the connections is not based on data and is solely illustrative of how different factors may differentially relate to factors on other levels. For instance, the meso-level process delay discounting is implemented across numerous micro-level brain regions and relates differentially to macro-level factors such as age and gender.*

Among the numerous potential cognitive processes of interest, one process that has been implicated across a range of mental health conditions is delay discounting (Amlung et al., 2019; Lempert et al., 2019), a measure of impulsive preference for smaller, immediate rewards over larger, delayed rewards (da Matta et al., 2012). Steeper discounting of delayed rewards – meaning a stronger preference for immediate rewards – has been associated with addictive disorders (Bickel et al., 2012, 2014), attention-deficit/hyperactivity disorder (Jackson & MacKillop, 2016), and eating-related disorders (Amlung et al., 2016). Therefore, scholars have proposed delay discounting as a transdiagnostic process involved in psychiatric conditions that provides insights into the underlying features of a range of disorders and marks a meaningful treatment target (Amlung et al., 2019; Levin et al., 2018). Moreover, delay discounting varies across age (Achterberg et al., 2016; Lu et al., 2023) and gender (Lv et al., 2025). Individuals who exhibit steeper discounting may gravitate towards experiences that promise more proximal rewards, such as short-form entertainment, rather than pursuing long-term rewarding outcomes, such as social connection (Endert & Mohr, 2020). Therefore, delay discounting may shape how users interact with and are affected by their online experiences. For instance, users who seek long-term rewards may foster meaningful interactions that contribute to their overall wellbeing, whereas users in pursuit of immediate rewards may be more vulnerable to potential negative effects of social media.

The present pre-registered study examines this premise using a unique combination of a large longitudinal dataset of adolescents and rich cognitive task data concerning delay discounting, which develops across adolescence and has been linked to a range of mental health outcomes. Specifically, we investigate whether individual differences in delay discounting moderate the relationship between social media use and mental health outcomes over time, as well as whether it relates to time spent social media and compulsive internet use. In doing so, we illustrate how researching nuanced cognitive processes can further our understanding of the mechanisms linking social media use and mental health and generate testable pathways of risk and resilience to the mental health effects of social media use.

Results

Sample characteristics

We analysed data from 512 children part of the seven-year Leiden Consortium on Individual Development longitudinal twin study (Crone et al., 2020). Participants were aged 7-9 at baseline and completed annual measurements for seven years. The data available at the seven measurement occasions are presented in Fig. 2. The included mental health, social media use, and behavioural task measures are presented in Table 1, and a detailed overview of all measurements and their timings is available online (Crone et al., 2020). The hypotheses, analyses, and exclusion criteria for the present research were preregistered at https://osf.io/jat54.

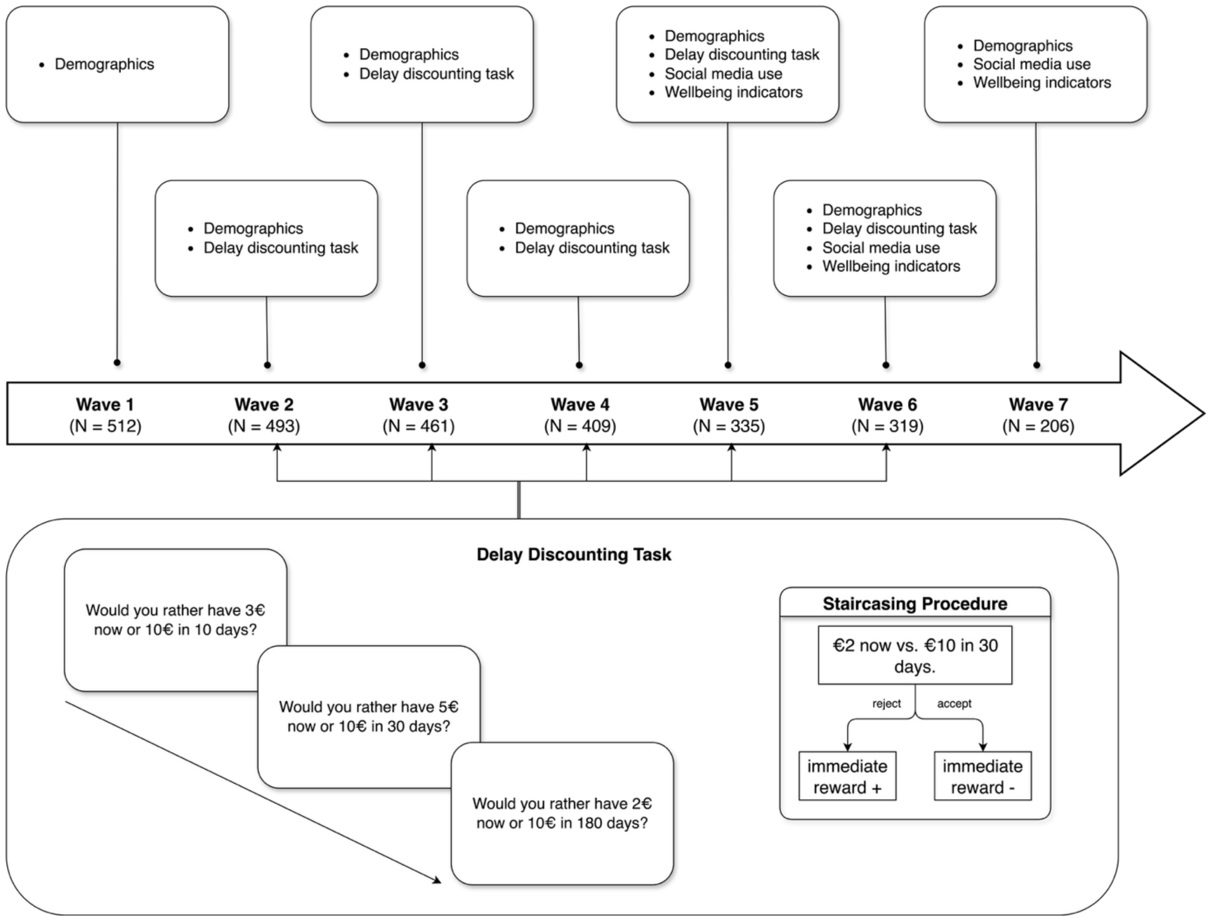

Fig. 2: Overview of the Data Available for each of the Seven Measurement Occasions in the Leiden Consortium on Individual Development. *This figure summarises the data available for analysis. Demographic data for all participants were available across all seven measurement occasions. The monetary delay discounting task was administered at measurement occasions 2-6. The task involved repeated presentations of hypothetical sooner and later offers that were adjusted based on a staircasing procedure. Measures of social media use and compulsive internet use were included at measurement occasions 5-7. Measures of mental health were included at measurement occasions 5-7. The subsequent longitudinal analyses are limited to data from measurement occasions 5-7 due to the availability of data on social media use and mental health.*

Variable	Time 1		Time 2		Time 3		Time 4		Time 5		Time 6		Time 7	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
N	250	262	241	252	221	240	199	210	159	176	149	170	103	103
Age (years)	7.891	7.997	8.900	8.982	9.955	10.000	11.196	11.257	12.401	12.417	13.294	13.380	15.299	15.520
Delay discounting	/	/	2.943	2.541	3.345	3.359	4.082	3.468	4.363	3.322	5.207	4.126	/	/
Social media use	/	/	/	/	/	/	/	/	1.091	1.625	1.170	1.750	1.700	2.183
CIUS	/	/	/	/	/	/	/	/	20.156	21.348	20.627	22.830	23.490	25.796
BIS	/	/	/	/	/	/	/	/	17.817	19.762	17.010	19.687	17.733	20.776
BAS	/	/	/	/	/	/	/	/	38.679	38.295	37.592	37.461	38.953	37.908
SDQ	/	/	/	/	/	/	/	/	21.280	22.315	21.168	23.176	21.484	24.063
HSCS	/	/	/	/	/	/	/	/	53.024	57.372	52.862	57.135	52.094	58.538
EATQ-EC	/	/	/	/	/	/	/	/	55.098	54.617	54.706	54.268	53.312	52.989

Table 1: Descriptive Statistics. *The mean values for all variables included in subsequent analyses are displayed and broken down by measurement occasion and participant gender. Social media use was quantified as the self-reported time spent posting and scrolling on social media (0 = never, 1 = less than one hour, 2 = one to two hours, 3 = two to three hours, 4 = three to four hours, 5 = more than four hours). The following acronyms are the mental health measures: CIUS = compulsive internet use scale; BAS = behavioural activation subscale; BIS = behavioural inhibition subscale; SDQ = strength and difficulties questionnaire; HSCS = highly sensitive child scale; EATQ-EC = early adolescent temperament scale effortful control subscale.*

Delay discounting

We employed a modelling approach to disentangle the distinct psychological mechanisms underlying delay discounting behaviour (Story et al., 2016). We modelled participants' behavioural choices on the delay discounting task using a hyperbolic function. This model assumes that the subjective value of an objective reward changes as a function of increasing delay. The model includes two free parameters: discounting rate (k) and inverse temperature (β). The discounting rate quantifies the rate at which rewards are devalued with delay, with higher values indicating heightened sensitivity to delays. The inverse temperature parameter captures behavioural consistency, with higher values indicating more deterministic choice patterns.

The negative log-transformed delay discounting parameter estimate (-log(k)) – hereafter referred to as delay discounting – showed strong correlations in line with task behaviour (see Fig. S4). Across measurements, the delay discounting parameter estimate showed a strong negative correlation with the observed probability of choosing the sooner response option ($t(1260) = -93.426$, 95% Confidence Interval (CI) [-0.941, -0.927], $p < 0.001$). Moreover, delay discounting decreased with age, as shown in a mixed effects model ($t(1260.620) = 6.413$, $\beta = 0.161$, 95% CI [0.112, 0.210], $p < 0.001$, R^2 marginal = 0.046, R^2 conditional = 0.447).

Social media use and compulsive internet use

First, we examined how delay discounting related to both time spent on social media and the compulsive internet use scale (CIUS; Meerkerk et al., 2009).

On average, participants reported they spent between one and two hours a day viewing messages, photos, or watching videos on social media (never = 22.296%, <1 hour = 34.565%, 1-2 hours 21.768%, 2-3 hours 12.533%, 3-4 hours 4.354%, >4 hours 4.485%). This time increased with age (mixed effects model; $t(679.010) = 5.393$, $\beta = 0.184$, 95% CI [0.117, 0.251], $p < 0.001$, R^2 marginal = 0.034, R^2 conditional = 0.360); with 11-year-olds reporting spending closer to <1 hour a day on social media ($X_{time,11} = 1.098$, 95% CI [0.910, 1.286]) in contrast to 1-2 hours in 17-year-olds ($X_{time,17} = 2.131$, 95% CI [1.881, 2.381]). Similarly, there was a positive linear association between age and social media use ($t(679.010) = 5.393$, $\beta = 0.184$, 95% CI [0.117, 0.251], $p < 0.001$, R^2 marginal = 0.034, R^2 conditional = 0.360), as well as compulsive internet use ($t(614.295) = 10.372$, $\beta = 0.321$, 95% CI [0.260, 0.382], $p < 0.001$, R^2 marginal = 0.104, R^2 conditional = 0.539).

Delay discounting estimates did not relate to reported time spent on social media ($t(372.279) = 0.374$, $\beta = 0.020$, 95% CI [-0.083, 0.122], $p = 0.708$, R^2 marginal < 0.001, R^2 conditional = 0.438) or compulsive internet use ($t(498.044) = 0.218$, $\beta = 0.012$, 95% CI [-0.093, 0.116], $p = 0.827$, R^2 marginal < 0.001, R^2 conditional = 0.529).

Longitudinal models of social media use and mental health outcomes

We proceeded to investigate the bidirectional within-person developmental associations between social media use and various mental health outcomes. The models decomposed stable between-person differences from within-person fluctuations, allowing us to test how deviations from one's average levels in social media use relate to subsequent changes in mental health and vice versa. Specifically, we were interested in the extent to which individual differences in delay discounting moderated these within-subject longitudinal associations.

We pre-registered the longitudinal models for the following mental health outcomes: Behavioural Inhibition and Activation System scale (BIS/BAS; Carver & White, 1994), the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997, 2001), the Highly Sensitive Child Scale (HSC; Pluess et al., 2018), and the effortful control subscale of the Early Adolescent Temperament questionnaire (EATQ-EC; Capaldi & Rothbart, 1992). The BIS/BAS scale is a 20-item self-report measure of individual differences in responsiveness to non-reward punishment (BIS; Cronbach's alpha = 0.81) and sensitivity to reward and approach-related behaviour (BAS; Cronbach's alpha = .84). People with high BIS are more likely to feel anxious or worried, potentially avoiding situations that could lead to punishment or unpleasant experiences (Carver & White, 1994; Svihra & Katzman, 2004); in contrast people with high BAS are more likely to be energetic and driven to seek out new experiences or rewards (Carver & White, 1994; Karimpour-Vazifehkhorani et al., 2020; McFarland et al., 2006).

The scale has been validated across studies, and evidence has supported both BIS and BAS as separate unidimensional measures (Carver & White, 1994; Jorm et al., 1998; Maack & Ebesutani, 2018). Prior to conducting the analyses, we deviated from the preregistration and decided to analyse the BIS and BAS subscales separately. This deviation was required to correct an error in the preregistration, where we proposed to analyse it as one concept, which would have failed to draw psychometrically and theoretically valid conclusions. Given that these systems are not opposite extremes on a continuum but orthogonal constructs, analyses on the total score would have obscured meaningful patterns and produced uninterpretable results. This matches standard practice in the literature, which consistently treats the subscales as separate outcomes (Jorm et al., 1998; Maack & Ebesutani, 2018; Quilty & Oakman, 2004).

The SDQ scale is a widely used 25-item measure of child mental health and assesses psychological difficulties across five domains. The total score indicates overall psychological difficulties and is internally consistent (Cronbach's alpha: 0.73) and relatively stable across time (Goodman, 1997, 2001; Goodman et al., 2000). The HSC is a 27-item measure of environmental sensitivity with acceptable internal consistency (Cronbach's alpha = 0.69; Pluess et al., 2018; Weyn et al., 2021). The L-CID used a subset of 13 items. The EATQ-EC subscale comprises 11 items that assess effortful control, the ability to flexibly regulate attention and behaviour, and has good psychometric properties (Capaldi & Rothbart, 1992; Clinciu, 2012; Latham et al., 2020).

Therefore, we fit a total of five baseline and five moderation RI-CLPMs examining the bidirectional relationship between social media use and 1) Behavioural Inhibition Subscale, 2) Behavioural Activation Subscale, 3) Strengths and Difficulties Questionnaire, 4) Highly Sensitive Child Scale, and 5) Early Adolescent Temperament Questionnaire Effortful Control Subscale.

Baseline Random Intercept Cross-lagged Panel Models

To examine the relationship between social media use and mental health, we fit five baseline models. An overview and visualisations of the model estimates are presented in Supplement 4 (Table S1, Fig. S5, S6a – S9a). The significant effects for each model are presented below.

All five baseline models had good fit (CFIs = 1.000, 90% CIs [>0.934, 1.000], TLIs = 1.000, 90% CIs [>0.522, 1.000], RMSEAs = 0.000, 90% CIs [0.000, <0.111]), with stable convergence (Potential Scale Reduction (PSR) values < 1.100). Within-person reciprocal associations between social media use and mental health were non-significant across all baseline models (all *p*s > 0.062). We found positive autoregressive effects for social media use between measurement points five and six across all baseline models (all *p*s < 0.025, all *β*s > 0.349). Moreover, we observed positive autoregressive effects for mental health measures at timepoints five and six for BIS (β = 0.295, p = 0.032), BAS (β = 0.323, p = 0.007), and HSCS (β = 0.373, p = 0.001). The only cross-sectional association between social media use and a mental health outcome observed was the positive association between social media use and EATQ-EC scores at measurement occasion six (β = 0.253, p = 0.013).

Moderation models of delay discounting

To examine whether individual differences in delay discounting moderated the relationship between social media use and mental health, we added delay discounting estimates as moderators to the cross-lagged panel models. We applied the delay discounting estimates from each measurement occasion to moderate the cross-lagged paths originating from that same timepoint. For example, delay discounting measured at timepoint five moderated the path from social media use at timepoint five to mental health outcome at timepoint six. This approach allowed us to test whether within-subject lagged associations from social media use to mental health varied depending on the level of delay discounting. Below, we examine the results for each mental health outcome separately. An overview and visualisations of the model estimates are presented in Supplement 4 and Fig. 3 below (Table S2, Fig. S6b – S9b)..

Our findings indicated no relationship between between-person differences in delay discounting and cross-lagged associations between social media use and BIS (p = 0.402), SDQ (p = 0.428), HSCS (p = 0.211), and EATQ-EC (p = 0.241). We therefore did not find evidence that delay discounting is a potential risk factor for negative social media effects on these four mental health outcomes.

For the Behavioural Activation Subscale, we observed a negative cross-lagged relationship between social media use and behavioural activation ($\beta_{5\text{-}6}$ = -0.198, $p_{5\text{-}6}$ = 0.038; $\beta_{6\text{-}7}$ = -0.065, $p_{6\text{-}7}$ = 0.353) when adding the moderator delay discounting (Fig. 3). Therefore, individuals who scored below their model-predicted amount of social media use showed lower behavioural activation in the subsequent year, i.e. they are less driven to pursue goals or pleasurable experiences and may thus experience less positive affect (Kasch et al., 2002; McFarland et al., 2006). This negative relationship was moderated by individual differences in delay discounting across measurement occasions (β_{logk5} = 0.265, p = 0.011; β_{logk6} = 0.252, p = 0.011); the negative association between social media and behavioural activation was more pronounced for individuals who showed a stronger preference for immediate rewards.

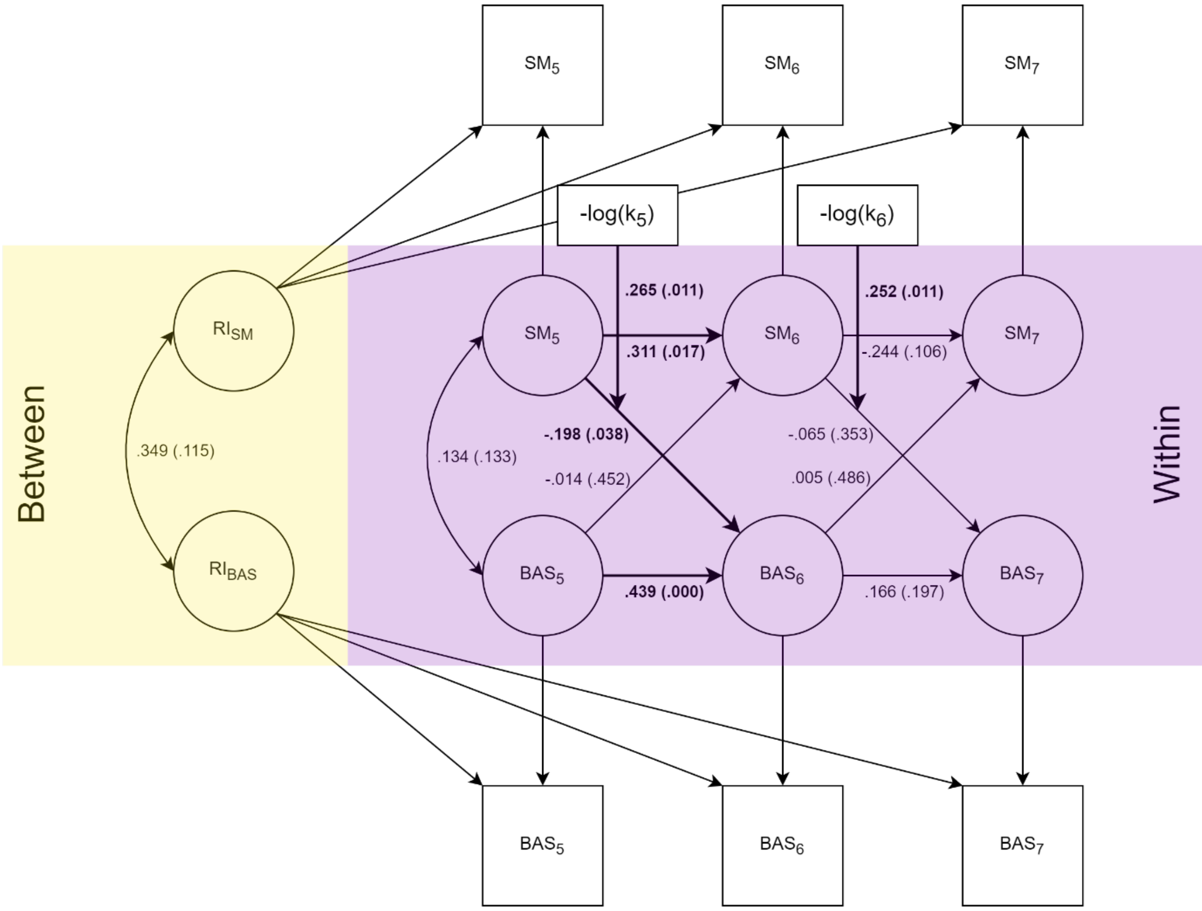

Fig. 3: Random-Intercept Cross-Lagged Panel Model for BAS. *The displayed random-intercept cross-lagged panel model shows the cross-lagged paths from the Behavioural Activation Subscale (BAS) to subsequent measures of time spent on social media (SM) and vice versa. The presented model includes the moderation paths from delay discounting estimates on the cross-lagged associations between time spent on social media and subsequent BAS measures. The significant moderation on the negative cross-lagged relationship suggests that individuals with a stronger preference for immediate rewards showed stronger negative cross-lagged relationships between social media use and behavioural activation. RI = random intercept.*

Discussion

The mechanisms that underlie individual susceptibility to social media's mental health effects remain poorly understood. Cognitive processes have the potential to underlie individuals' risk or resilience to experiencing negative mental health outcomes on social media, but have remained largely unexplored. In the first study of its kind, we therefore tested whether delay discounting – a cognitive process reflecting preference for immediate versus delayed rewards – moderates longitudinal associations between social media use and mental health outcomes in over 500 adolescents across three measurement occasions. We also conducted exploratory analyses examining direct relationships between delay discounting and both social media engagement and compulsive internet use.

Across five mental health outcomes examined, we found one significant moderation effect of delay discounting. Specifically, higher social media use predicted decreases in behavioural activation (i.e. approach-related motivation and behaviour), and this negative association was stronger among individuals who showed a stronger preference for immediate rewards. This pattern suggests that individuals who prefer immediate rewards may be more susceptible to the negative mental health effects of social media use, as deficits in approach-oriented behaviour (i.e. not seeking out rewarding experiences) have been associated with the onset and maintenance of mental health conditions, including depression (Hasler et al., 2010; Kasch et al., 2002; Markarian et al., 2013; McFarland et al., 2006).

However, we did not find delay discounting to moderate the relationship between social media use and the remaining four mental health outcomes. A possible explanation for these mixed findings is that delay discounting may be of particular relevance to the relationship between social media use and approach motivation for reward-related outcomes. Previous studies have illustrated the importance of reward learning for explaining social media engagement (Lindström et al., 2021; Turner et al., 2024), which may subsequently affect user wellbeing. However, the mixed nature of our findings calls for caution, and more research will be required to establish the robustness of the observed findings.

We further examined the direct relationship between delay discounting and social media use and compulsive internet use. Our findings indicate no differences in time spent on social media or compulsiveness of internet use as a function of differences in delay discounting. Whereas some studies have reported a direct relationship between delay discounting and time spent on social media (Endert & Mohr, 2020), previous research has only been completed on adult samples. More generally, it is plausible that cognitive processes operate through more complex mechanisms that determine the qualitatively different usage patterns that go beyond the mere amount of consumption (Parry et al., 2021; Turel et al., 2018).

The present research takes an innovative approach to studying the relationship between social media use and mental health that capitalises on advances in computational and longitudinal modelling of human behaviour. It represents the first attempt to examine individual differences in the relationship between social media use and mental health through the lens of meso-level cognitive processes. We used some of the best longitudinal data available, which spanned critical developmental periods of adolescence and enabled us to examine the influence of cognitive processes on the bidirectional relationship between social media use and mental health. Our computational approach was integrated into longitudinal models to estimate individual differences in cognitive processes that could provide nuanced insights. Contrasting previous research attempting to characterise risk and resilience to the effects of social media use on mental health, the present findings constitute the first evidence suggesting that individual differences in fundamental cognitive processes have the potential to generate compelling insights into the mechanisms underlying the relationship between social media use and mental health outcomes. This research opens new avenues for understanding and potentially intervening in the complex relationship between digital technologies and adolescent mental health.

Our findings and their implications for future research should however be considered in light of two main limitations. First, the measurement of individual differences in cognitive processes is a contentious issue in the field of computational psychiatry. Although a computational approach has allowed for the identification of latent cognitive processes, heterogeneous task design and implementation have raised concerns about the specificity and sensitivity of parameter estimates to real-world outcomes (Bailey et al., 2021). However, previous studies have reported acceptable or strong temporal stability of delay discounting estimates in both child (Burns et al., 2020; Klein et al., 2022) and adolescent samples (Gelino et al., 2024). Measurement-related properties are important to consider when attempting to extend the presented framework to other cognitive processes. Moreover, the present findings were based on the discounting of monetary rewards, whereas previous research has indicated that discounting of social rewards may be more relevant in moderating relationships between social processes and affective outcomes (Hill et al., 2023; Pegg et al., 2021; Politte-Corn et al., 2024). Future studies may expand on the present findings by examining cognitive processes within social decision-making contexts.

Another limitation of the findings follows from delays between measurement occasions. Whereas the present results are based on annual measurements, it is plausible that the time scales at which cognitive processes exert their influences on user behaviour and mental health are shorter. Studies that involve shorter intervals between measurements may provide important further insights into the dynamics of social media use and mental health, and the extent to which these dynamics are shaped by individual differences in cognitive processes.

Conclusion

Previous attempts to understand individual differences in mental health outcomes of social media use have focused on macro-level differences such as age or gender. While demographic variables establish important population-level patterns, they must be complemented to understand the specific cognitive mechanisms through which these effects manifest. This cognitive lens enables researchers to identify not only who might be at risk but precisely how and why certain individuals experience adverse outcomes, creating pathways for intervention and prevention. Previous research has demonstrated the malleability of cognitive processes through therapeutic intervention (Bickel et al., 2011) or non-invasive brain stimulation (Figner et al., 2010) to manipulate discounting rates, illustrating how the study of cognitive processes may generate actionable insights and can be used to advance targeted intervention (Koffarnus et al., 2013).

The cognitive approach will also become increasingly relevant as social media platforms are engineered to align with and exploit users’ cognitive processes, such as attention allocation or reward sensitivity, to maximise user engagement metrics and corporate revenue (Strümke et al., 2023). Understanding the relationship between cognition, technologies and well-being in this context will be especially critical for safeguarding adolescents, who face unique susceptibilities due to converging neurodevelopmental factors. The rapid design of evidence-based strategies to improve digital well-being might well rest on a cognitive understanding of how – and for whom – specific aspects of social media facilitate negative outcomes.

Methods

This data collection was approved by the Dutch Central Committee for Human Research and conducted in accordance with the standards set by the Declaration of Helsinki. The hypotheses, analyses, and exclusion criteria were preregistered on the Open Science Framework (osf.io/jat54), and the corresponding code is available on GitHub (https://doi.org/10.5281/zenodo.17347224). We discuss deviations from the preregistration in the relevant sections and provide a detailed discussion of any deviations in the preregistration deviation table available in Supplement 9.

Dataset

The L-CID dataset is an accelerated cohort-sequential longitudinal twin sample, with longitudinal data of children between the ages of 3 and 17 from 2013 to 2023. The study was designed in two overlapping cohorts starting at the ages of 3 to 4 and 7 to 8 years, to capture a 10-year developmental period. Participation invitations were sent to families with same-sex twins born between 2006 and 2009, within a two-hour radius of the city of Leiden, the Netherlands. Eligibility criteria included being fluent in Dutch and having parents and grandparents born in Europe (for genetic analyses). Further exclusion criteria included congenital disability, chronic illness, hereditary disease, or a visual or hearing impairment if the disorder disabled the child from performing the behavioural tasks or from participating in imaging procedures. The data collection was approved by the Dutch Central Committee on Research Involving Human Subjects (CCMO, number NL50277.058.14), and written informed consent was obtained from both parents and children. For a detailed description of the dataset see Crone et al. (2020).

The present research is based on a subset of the data from seven measurement waves of the middle childhood cohort. This cohort contains 512 participants at the first measurement occasion, from 256 families. Participants (female = 262, male = 250) age ranged from seven to nine years at inclusion (M = 7.946, SD = 0.671). Participants were of low (9%), medium (46%), and high (45%) socio-economic status, predominantly Caucasian (90%) and scoring in a normal range on IQ (*M* = 103.580, *SD* = 11.760). Moreover, 87% of participants were right-handed, 55% monozygotic twins, and 2.1% diagnosed with an AXIS-I disorder. Subsequent waves included data from 493, 461, 409, 335, 319, and 206 participants at waves two to seven, respectively.

Measures

The subset of task and questionnaire measures available for the present research is outlined in Table 1, and the procedure in Fig. 2. Both are described in detail in the sections below. We set out to analyse participants' trial-level data from the temporal delay discounting task and the stop signal reaction time task, however we report analyses and results for the delay discounting task data only. This deviation from the pre-registration followed concerns surrounding the data quality of the stop signal reaction time trial level data, which was inconsistent and featured errors in coding and data values in the sixth measurement wave. Since it was not possible to recover the source of errors, we had to exclude the sixth wave of the stop signal reaction time task and subsequently exclude the task altogether, given that the planned longitudinal analyses were not possible with two measurements spaced two years apart.

Delay discounting task

A hypothetical delay discounting task (Achterberg et al., 2016; Peper et al., 2013; Richards et al., 1999) was administered at waves two, three, four, five, and six. The task involved a series of choices between small (€0-10), immediately available monetary rewards or a larger (€10) reward available after a variable delay (i.e. 2, 30, 180, or 365 days). Trials of different delays were presented in random order and took the form of "What would you rather have: €2 right away or €10 in 30 days?". Rewards were adjusted based on a staircasing procedure aimed at maximising information gained on a given trial, the exact implementation of which is available at https://zenodo.org/records/17338997 and adapted from Du et al. (2002). An exemplary trial of the delay discounting task is displayed in Fig. 2.

Self-report questionnaires

All participants completed a questionnaire battery at each annual measurement wave. The questionnaire batteries varied across the measurement occasions and contained questionnaires pertaining to a range of constructs such as social, emotional, and physical well-being. The self-report measures relevant to the present project include the Behavioural Inhibition and Activation System scale (BIS/BAS; Carver & White, 1994), the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997, 2001), the Highly Sensitive Child Scale (HSC; Pluess et al., 2018), and the effortful control subscale of the Early Adolescent Temperament questionnaire (EATQ-EC; Capaldi & Rothbart, 1992).

Social media use was assessed using items adapted from the Media Multitasking Index (Pea et al., 2012), assessing time spent engaging in social media-related activities, rated on a six-point Likert Scale (never, less than one hour, one to two hours, two to three hours, three to four hours, more than four hours). Social media use was quantified based on participants' responses to the question "On an average weekday, how much time do you spend on your phone or iPad/ tablet viewing public messages, photos, or videos (e.g. Facebook, Instagram, TikTok, Snapchat, Twitter)". Moreover, the Compulsive Internet Use Scale (CIUS; Meerkerk et al., 2009) was administered at measurement waves five, six, and seven. This self-report measure is often used to assess problematic patterns of internet use (Guertler et al., 2014).

Exclusion criteria

All exclusion criteria were preregistered on the Open Science Framework at https://osf.io/jat54. Exclusion criteria were split into participant- and trial-level criteria. Participants were removed if they showed no variation in their response pattern on the delay discounting task (i.e. accepting or rejecting all offers), or had fewer than twenty trials of data available. These criteria led to the exclusion of no participants across all waves.

Data analysis

Our primary analyses examined the moderating effect of the cognitive process (i.e. delay discounting) on the bidirectional, longitudinal relationships between time spent on social media and mental health outcomes. We fit random intercept cross-lagged panel models for each mental health outcome (five models). Delay discounting was quantified based on the trial-level behavioural choices on the monetary delay discounting task. The mental health outcomes included the BAS, the BIS, the HSCS, the SDQ, and the EATQ-EC. Whereas we preregistered analyses for the BIS/BAS total score, this mistake was corrected prior to the analyses, and we performed analyses separately for the BIS and BAS. This decision was based on the fact that the behavioural inhibition and activation systems are considered independent systems, the combination of which would have compromised the validity of our interpretations (Jorm et al., 1998; Maack & Ebesutani, 2018). We provide a detailed description of the deviation in the preregistration deviation table S7.

Delay discounting task

We modelled participants' temporal preferences on the delay discounting task using a hyperbolic function. The hyperbolic model is an established framework for examining intertemporal choice behaviour and posits that the subjective value *V* of a delayed reward decreases non-linearly as a function of time (Rubinstein, 2003; Story et al., 2016; van der Pol & Cairns, 2002). The hyperbolic function formulates how the subjective value (*V*) of an objective reward (*A*) changes as a function of increases in reward delay (*D*).

$$V = \frac{A}{1 + k * D} \tag{1}$$

The free parameter *k* determines the rate at which increases in delay result in a decreased subjective reward value. Larger values of *k* translate to steeper decreases in subjective reward value with increasing delays. On a given trial (*t*), the subjective reward value for the sooner (V_{SS}) and later (V_{LL}) options were estimated and translated into choice probabilities using a softmax function (Luce's choice rule).

$$P(LLoverSS) = \frac{1}{1 + e^{-\beta(V_{SS(t)} - V_{LL(t)}}} \tag{2}$$

The softmax function includes an inverse temperature parameter, the second free parameter of the model, and controls an individual's sensitivity to differences in subjective value between the sooner and later option. A higher *β* indicates more consistent (deterministic) choice patterns, whereas a lower *β* indicates more stochastic choices on the delay discounting task. The hyperbolic and softmax functions for different parameter values are visualised in Fig. 4.

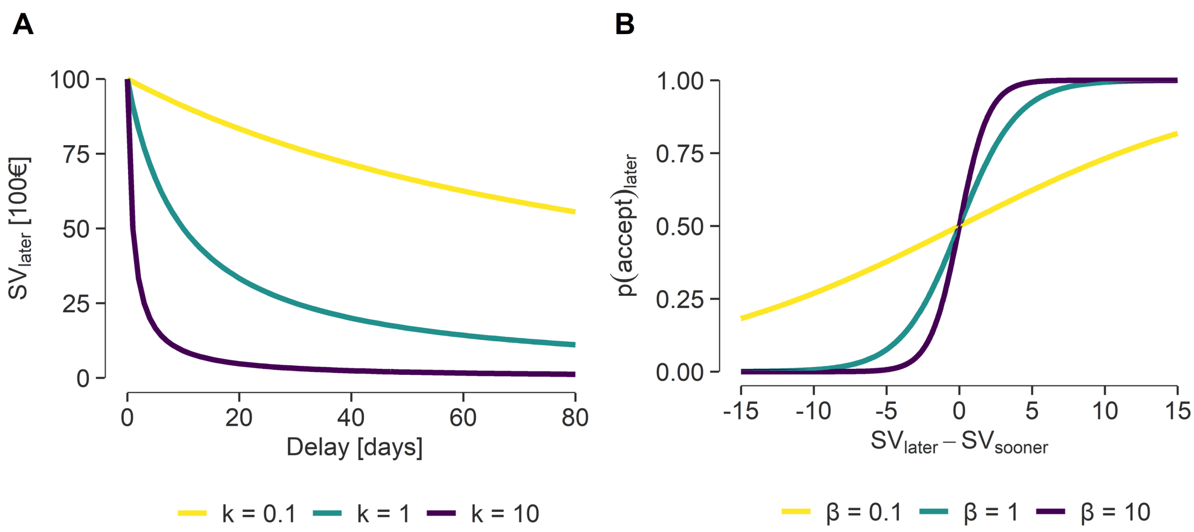

Fig. 4: Hyperbolic and Softmax Function for Different Parameter Configurations. A: *The estimated subjective value of the later response option for different delay discounting parameter (k) values. Larger delay discounting parameters capture steeper discounting of rewards with delay.* **B:** *The estimated acceptance probability of the later response option for different inverse temperature parameter (β) values. Larger inverse temperature values indicate more deterministic responding.*

We took a hierarchical Bayesian approach to model fitting (Ahn et al., 2013; Piray et al., 2019), implemented in *CmdStan R* (Gabry et al., 2023), with code adapted from *hBayesDM* (Ahn et al., 2017). All models were fit using Markov-Chain Monte Carlo (MCMC), with 2000 warm-up iterations and 10,000 sampling iterations, across four chains. Model convergence was established by visual inspection of trace plots, and numeric diagnostics, effective sample size (ESS) and R-hats. The R-hat statistic is a quantitative measure of model convergence and is based on the variance estimates within and between MCMC chains, with values greater than 1.1 indicating that the chains have not converged. We further validated our modelling procedure by examining behavioural correlates of computational processes, parameter recovery, and posterior predictive checks. We negative log-transformed the delay discounting parameter (-log*k*; Weinsztok et al., 2021). This means that lower values of *-logk* are interpreted as steeper discounting of rewards, and vice versa. The posterior predictive checks and parameter recovery are presented in the Supplement (Fig. S2, S3).

Cross-sectional analyses

To examine the cross-sectional relationships of social media use and compulsive internet use with age and delay discounting, we fit linear mixed effects models with fixed effects for the predictor and random intercepts for participants. We fit the models using the *lme4* R package (Bates et al., 2015) and examined the marginal and conditional explained variance using the *MuMIn* R package (Bartoń, 2025).

Random intercept cross-lagged panel models

We fit separate random intercept cross-lagged panel models (RI-CLPMs) for each mental health outcome (n = 5) and cognitive parameter 'delay discounting' (-logk) to estimate the bi-directional relationship between social media use and well-being and the moderating effect of 'delay discounting' on this relationship. All models were fit using a Bayesian estimator in Mplus, which treats missing data as additional unknown quantities that are sampled from their conditional posterior distribution (McNeish & Hamaker, 2020). We examined model fit using the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI). Model fit was considered good if RMSEA < 0.06, CFI > 0.95, and TLI > 0.95. Model convergence was assessed using the criterion of potential scale reduction (PSR), indicating whether between-chain variation (relative to the total of between- and within-chain variation) of two or more Markov chains is close to zero (Asparouhov & Muthen, 2010; Asparouhov & Muthén, 2021). Upon convergence (i.e. PSR < 1.1), the number of iterations was doubled to ensure stable convergence.

Baseline random intercept cross-lagged panel models

We fit five baseline RI-CLPMs for each well-being outcome to examine whether social media use and well-being are reciprocally related at the within-person level. We estimated random intercepts for social media use and mental health outcome, and cross-lagged and autoregressive effects were specified on the within-person centred residuals.

Moderation models on the between x within interaction

We fit five RI-CLPMs, including interaction terms between the cognitive process 'delay discounting' and social media use to predict within-person deviations in mental health outcomes at each measurement wave. The approach assumes stability of the cognitive process at the within-person level and estimates represent a combination of within- and between-person moderation effects if the assumption is not met. Measurement error variances were fixed to 0.2 to improve model convergence. This approach should not introduce substantial bias as long as measurement error variances are an approximation of zero (Asparouhov & Muthén, 2021; Speyer et al., 2023).

References

Achterberg, M., Peper, J. S., Duijvenvoorde, A. C. K. van, Mandl, R. C. W., & Crone, E. A. (2016). Frontostriatal White Matter Integrity Predicts Development of Delay of Gratification: A Longitudinal Study. *Journal of Neuroscience*, *36*(6), 1954–1961. https://doi.org/10.1523/JNEUROSCI.3459-15.2016

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. *Computational Psychiatry (Cambridge, Mass.)*, *1*, 24–57. https://doi.org/10.1162/CPSY_a_00002

Ahn, W.-Y., Krawitz, A., Kim, W., Busemeyer, J. R., & Brown, J. W. (2013). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *Decision*, *1*(S), 8–23. https://doi.org/10.1037/2325-9965.1.S.8

Amlung, M., Marsden, E., Holshausen, K., Morris, V., Patel, H., Vedelago, L., Naish, K. R., Reed, D. D., & McCabe, R. E. (2019). Delay Discounting as a Transdiagnostic Process in Psychiatric Disorders: A Meta-analysis. *JAMA Psychiatry*, *76*(11), 1176. https://doi.org/10.1001/jamapsychiatry.2019.2102

Amlung, M., Petker, T., Jackson, J., Balodis, I., & MacKillop, J. (2016). Steep discounting of delayed monetary and food rewards in obesity: A meta-analysis. *Psychological Medicine*, *46*(11), 2423–2434. https://doi.org/10.1017/S0033291716000866

Asparouhov, T., & Muthen, B. (2010). *Bayesian Analysis Using Mplus: Technical Implementation* (Version 3) [Computer software].

Asparouhov, T., & Muthén, B. (2021). Bayesian estimation of single and multilevel models with latent variable interactions. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(2), 314–328. https://doi.org/10.1080/10705511.2020.1761808

Bailey, A. J., Romeu, R. J., & Finn, P. R. (2021). The problems with delay discounting: A critical review of current practices and clinical applications. *Psychological Medicine*, *51*(11), 1799–1806. https://doi.org/10.1017/S0033291721002282

Bartoń, K. (2025). *MuMIn: Multi-Model Inference* (Version 1.48.11) [Computer software]. https://cran.r-project.org/web/packages/MuMIn/index.html

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bickel, W. K., Jarmolowicz, D. P., Mueller, E. T., Koffarnus, M. N., & Gatchalian, K. M. (2012). Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: Emerging evidence. *Pharmacology & Therapeutics*, *134*(3), 287–297. https://doi.org/10.1016/j.pharmthera.2012.02.004

Bickel, W. K., Koffarnus, M. N., Moody, L., & Wilson, A. G. (2014). The Behavioral- and Neuro-Economic Process of Temporal Discounting: A Candidate Behavioral Marker of Addiction. *Neuropharmacology*, *76*(0 0), 10.1016/j.neuropharm.2013.06.013. https://doi.org/10.1016/j.neuropharm.2013.06.013

Bickel, W. K., Yi, R., Landes, R. D., Hill, P. F., & Baxter, C. (2011). Remember the Future: Working Memory Training Decreases Delay Discounting Among Stimulant Addicts. *Biological Psychiatry*, *69*(3), 260–265. https://doi.org/10.1016/j.biopsych.2010.08.017

Booker, C. L., Kelly, Y. J., & Sacker, A. (2018). Gender differences in the associations between age trends of social media interaction and well-being among 10-15 year olds in the UK. *BMC Public Health*, *18*(1), 321. https://doi.org/10.1186/s12889-018-5220-4

Browning, M., Paulus, M., & Huys, Q. J. M. (2023). What is Computational Psychiatry Good For? *Biological Psychiatry*, *93*(8), 658–660. https://doi.org/10.1016/j.biopsych.2022.08.030

Burns, P., Fay, O., McCafferty, M.-F., McKeever, V., Atance, C., & McCormack, T. (2020). Examining children's ability to delay reward: Is the delay discounting task a suitable measure? *Journal of Behavioral Decision Making*, *33*(2), 208–219. https://doi.org/10.1002/bdm.2154

Capaldi, D. M., & Rothbart, M. K. (1992). Development and validation of an early adolescent temperament measure. *The Journal of Early Adolescence*, *12*(2), 153–173. https://doi.org/10.1177/0272431692012002002

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS

Scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Clinciu, A. I. (2012). Convergent validation of EATQ-R questionnaire against Eysenck's PEN model of personality. *Procedia - Social and Behavioral Sciences*, *33*, 408–412. https://doi.org/10.1016/j.sbspro.2012.01.153

Coyne, S. M., Weinstein, E., Sheppard, J. A., James, S., Gale, M., Van Alfen, M., Ririe, N., Monson, C., Ashby, S., Weston, A., & Banks, K. (2023). Analysis of Social Media Use, Mental Health, and Gender Identity Among US Youths. *JAMA Network Open*, *6*(7), e2324389. https://doi.org/10.1001/jamanetworkopen.2023.24389

Crone, E. A., Achterberg, M., Dobbelaar, S., Euser, S., van den Bulk, B., der Meulen, M. van, van Drunen, L., Wierenga, L. M., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2020). Neural and behavioral signatures of social evaluation and adaptation in childhood and adolescence: The Leiden consortium on individual development (L-CID). *Developmental Cognitive Neuroscience*, *45*, 100805. https://doi.org/10.1016/j.dcn.2020.100805

Crone, E. A., & Konijn, E. A. (2018). Media use and brain development during adolescence. *Nature Communications*, *9*(1), 588. https://doi.org/10.1038/s41467-018-03126-x

Cunningham, S., Hudson, C. C., & Harkness, K. (2021). Social Media and Depression Symptoms: A Meta-Analysis. *Research on Child and Adolescent Psychopathology*, *49*(2), 241–253. https://doi.org/chu

da Matta, A., Gonçalves, F. L., & Bizarro, L. (2012). Delay discounting: Concepts and measures. *Psychology & Neuroscience*, *5*(2), 135–146. https://doi.org/10.3922/j.psns.2012.2.03

Du, W., Green, L., & Myerson, J. (2002). Cross-Cultural Comparisons of Discounting Delayed and Probabilistic Rewards. *The Psychological Record*, *52*(4), 479–492. https://doi.org/10.1007/BF03395199

Endert, T. S. van, & Mohr, P. N. C. (2020). Likes and impulsivity: Investigating the relationship between actual smartphone use and delay discounting. *PLOS ONE*, *15*(11), e0241383. https://doi.org/10.1371/journal.pone.0241383

Fassi, L., Thomas, K., Parry, D. A., Leyland-Craggs, A., Ford, T. J., & Orben, A. (2024). Social Media Use and Internalizing Symptoms in Clinical and Community Adolescent Samples: A Systematic Review and Meta-Analysis.

JAMA Pediatrics, *178*(8), 814–822. https://doi.org/10.1001/jamapediatrics.2024.2078

Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., & Weber, E. U. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience*, *13*(5), 538–539. https://doi.org/10.1038/nn.2516

Gabry, J., Češnovar, R., & Johnson, A. (2023). *cmdstanr: R interface to CmdStan* [Computer software]. https://mc-stan.org/cmdstanr/, https://discourse.mc-stan.org

Gelino, B. W., Schlitzer, R. D., Reed, D. D., & Strickland, J. C. (2024). A systematic review and meta-analysis of test–retest reliability and stability of delay and probability discounting. *Journal of the Experimental Analysis of Behavior*, *121*(3), 358–372. https://doi.org/10.1002/jeab.910

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(11), 1337–1345. https://doi.org/10.1097/00004583-200111000-00015

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry*, *177*(6), 534–539. https://doi.org/10.1192/bjp.177.6.534

Guertler, D., Rumpf, H.-J., Bischof, A., Kastirke, N., Petersen, K. U., John, U., & Meyer, C. (2014). Assessment of Problematic Internet Use by the Compulsive Internet Use Scale and the Internet Addiction Test: A Sample of Problematic and Pathological Gamblers. *European Addiction Research*, *20*(2), 75–81. https://doi.org/10.1159/000355076

Hasler, B. P., Allen, J. J. B., Sbarra, D. A., Bootzin, R. R., & Bernert, R. A. (2010). Morningness–eveningness and depression: Preliminary evidence for the role of the behavioral activation system and positive affect. *Psychiatry Research*, *176*(2), 166–173. https://doi.org/10.1016/j.psychres.2009.06.006

Hill, K. E., Dickey, L., Pegg, S., Dao, A., Arfer, K. B., & Kujawa, A. (2023). Associations between parental conflict and social and monetary reward responsiveness in adolescents with clinical depression. *Research on Child and Adolescent Psychopathology*, *51*(1), 119–131. https://doi.org/10.1007/s10802-022-00949-7

Ising, M., Horstmann, S., Kloiber, S., Lucae, S., Binder, E. B., Kern, N., Künzel, H. E., Pfennig, A., Uhr, M., & Holsboer, F. (2007). Combined Dexamethasone/Corticotropin Releasing Hormone Test Predicts Treatment Response in Major Depression–A Potential Biomarker? *Biological Psychiatry*, *62*(1), 47–54. https://doi.org/10.1016/j.biopsych.2006.07.039

Ivie, E. J., Pettitt, A., Moses, L. J., & Allen, N. B. (2020). A meta-analysis of the association between adolescent social media use and depressive symptoms. *Journal of Affective Disorders*, *275*, 165–174. https://doi.org/10.1016/j.jad.2020.06.014

Jackson, J. N. S., & MacKillop, J. (2016). Attention-Deficit/Hyperactivity Disorder and Monetary Delay Discounting: A Meta-Analysis of Case-Control Studies. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(4), 316–325. https://doi.org/10.1016/j.bpsc.2016.01.007

Jorm, A. F., Christensen, H., Henderson, A. S., Jacomb, P. A., Korten, A. E., & Rodgers, B. (1998). Using the BIS/BAS scales to measure behavioural inhibition and behavioural activation: Factor structure, validity and norms in a large community sample. *Personality and Individual Differences*, *26*(1), 49–58. https://doi.org/10.1016/S0191-8869(98)00143-3

Karimpour-Vazifehkhorani, A., Bakhshipour Rudsari, A., Rezvanizadeh, A., Kehtary-Harzang, L., & Hasanzadeh, K. (2020). Behavioral Activation Therapy on Reward Seeking Behaviors in Depressed People: An Experimental study. *Journal of Caring Sciences*, *9*(4), 195–202. https://doi.org/10.34172/jcs.2020.030

Kasch, K. L., Rottenberg, J., Arnow, B. A., & Gotlib, I. H. (2002). Behavioral activation and inhibition systems and the severity and course of depression. *Journal of Abnormal Psychology*, *111*(4), 589–597. https://doi.org/10.1037//0021-843x.111.4.589

Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*. https://www.tandfonline.com/doi/abs/10.1080/02673843.2019.1590851

Kelly, C. A., & Sharot, T. (2021). Individual differences in information-seeking. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-27046-5

Klein, S. D., Collins, P. F., & Luciana, M. (2022). Developmental trajectories of delay discounting from childhood to young adulthood: Longitudinal associations and test-retest reliability. *Cognitive Psychology*, *139*, 101518. https://doi.org/10.1016/j.cogpsych.2022.101518

Koffarnus, M. N., Jarmolowicz, D. P., Mueller, E. T., & Bickel, W. K. (2013). Changing delay discounting in the light of the competing neurobehavioral decision systems theory: A review. *Journal of the Experimental Analysis of Behavior*, *99*(1), 32–57. https://doi.org/10.1002/jeab.2

Kurten, S., Ghai, S., Odgers, C., Kievit, R. A., & Orben, A. (2025). Deprivation’s role in adolescent social media use and its links to life satisfaction. *Computers in Human Behavior*, *165*, 108541. https://doi.org/10.1016/j.chb.2024.108541

Latham, M. D., Dudgeon, P., Yap, M. B. H., Simmons, J. G., Byrne, M. L., Schwartz, O. S., Ivie, E., Whittle, S., & Allen, N. B. (2020). Factor Structure of the Early Adolescent Temperament Questionnaire–Revised. *Assessment*, *27*(7), 1547–1561. https://doi.org/10.1177/1073191119831789

Lempert, K. M., Steinglass, J. E., Pinto, A., Kable, J. W., & Simpson, H. B. (2019). Can delay discounting deliver on the promise of RDoC? *Psychological Medicine*, *49*(2), 190–199. https://doi.org/10.1017/S0033291718001770

Levin, M. E., Haeger, J., Ong, C. W., & Twohig, M. P. (2018). An Examination of the Transdiagnostic Role of Delay Discounting in Psychological Inflexibility and Mental Health Problems. *The Psychological Record*, *68*(2), 201–210. https://doi.org/10.1007/s40732-018-0281-4

Lindström, B., Bellander, M., Schultner, D. T., Chang, A., Tobler, P. N., & Amodio, D. M. (2021). A computational reward learning account of social media engagement. *Nature Communications*, *12*(1), 1311. https://doi.org/10.1038/s41467-020-19607-x

Lu, J., Yao, J., Zhou, Z., & Wang, X. T. (2023). Age effects on delay discounting across the lifespan: A meta-analytical approach to theory comparison and model development. *Psychological Bulletin*, *149*(7–8), 447–486. https://doi.org/10.1037/bul0000396

Lv, C., Liu, Z., Turel, O., & He, Q. (2025). Understanding gender differences in delay discounting: A systematic review and meta-analysis. *Humanities and Social*

Sciences Communications, *12*(1), 1–15. https://doi.org/10.1057/s41599-025-04843-7

Maack, D. J., & Ebesutani, C. (2018). A re-examination of the BIS/BAS scales: Evidence for BIS and BAS as unidimensional scales. *International Journal of Methods in Psychiatric Research*, *27*(2), e1612. https://doi.org/10.1002/mpr.1612

Markarian, S. A., Pickett, S. M., Deveson, D. F., & Kanona, B. B. (2013). A model of BIS/BAS sensitivity, emotion regulation difficulties, and depression, anxiety, and stress symptoms in relation to sleep quality. *Psychiatry Research*, *210*(1), 281–286. https://doi.org/10.1016/j.psychres.2013.06.004

McFarland, B. R., Shankman, S. A., Tenke, C. E., Bruder, G. E., & Klein, D. N. (2006). Behavioral activation system deficits predict the six-month course of depression. *Journal of Affective Disorders*, *91*(2–3), 229–234. https://doi.org/10.1016/j.jad.2006.01.012

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods*, *25*(5), 610–635. https://doi.org/10.1037/met0000250

Meerkerk, G.-J., Van Den Eijnden, R. J. J. M., Vermulst, A. A., & Garretsen, H. F. L. (2009). The Compulsive Internet Use Scale (CIUS): Some Psychometric Properties. *CyberPsychology & Behavior*, *12*(1), 1–6. https://doi.org/10.1089/cpb.2008.0181

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, *16*(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018

Orben, A., & Blakemore, S.-J. (2023). How social media affects teen mental health: A missing link. *Nature*, *614*(7948), 410–412. https://doi.org/10.1038/d41586-023-00402-9

Orben, A., & Matias, J. N. (2025). Fixing the science of digital technology harms. *Science*, *388*(6743), 152–155. https://doi.org/10.1126/science.adt6807

Orben, A., Przybylski, A. K., Blakemore, S.-J., & Kievit, R. A. (2022). Windows of developmental sensitivity to social media. *Nature Communications*, *13*(1), 1649. https://doi.org/10.1038/s41467-022-29296-3

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Pea, R., Nass, C., Meheula, L., Rance, M., Kumar, A., Bamford, H., Nass, M., Simha, A., Stillerman, B., Yang, S., & Zhou, M. (2012). Media use, face-to-face communication, media multitasking, and social well-being among 8- to 12-year-old girls. *Developmental Psychology*, *48*(2), 327–336. https://doi.org/10.1037/a0027030

Pegg, S., Arfer, K. B., & Kujawa, A. (2021). Altered Reward Responsiveness and Depressive Symptoms: An Examination of Social and Monetary Reward Domains and Interactions with Rejection Sensitivity. *Journal of Affective Disorders*, *282*, 717–725. https://doi.org/10.1016/j.jad.2020.12.093

Peper, J. S., Mandl, R. C. W., Braams, B. R., de Water, E., Heijboer, A. C., Koolschijn, P. C. M. P., & Crone, E. A. (2013). Delay Discounting and Frontostriatal Fiber Tracts: A Combined DTI and MTR Study on Impulsive Choices in Healthy Young Adults. *Cerebral Cortex (New York, NY)*, *23*(7), 1695–1702. https://doi.org/10.1093/cercor/bhs163

Pike, A. C., & Robinson, O. J. (2022). Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control Individuals: A Systematic Review and Meta-analysis. *JAMA Psychiatry*, *79*(4), 313–322. https://doi.org/10.1001/jamapsychiatry.2022.0051

Piray, P., Dezfouli, A., Heskes, T., Frank, M. J., & Daw, N. D. (2019). Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. *PLOS Computational Biology*, *15*(6), e1007043. https://doi.org/10.1371/journal.pcbi.1007043

Pluess, M., Assary, E., Lionetti, F., Lester, K. J., Krapohl, E., Aron, E. N., & Aron, A. (2018). Environmental sensitivity in children: Development of the Highly Sensitive Child Scale and identification of sensitivity groups. *Developmental Psychology*, *54*(1), 51–70. https://doi.org/10.1037/dev0000406

Politte-Corn, M., Pegg, S., Dickey, L., & Kujawa, A. (2024). Neural Reactivity to Social Reward Moderates the Association Between Social Media Use and Momentary Positive Affect in Adolescents. *Affective Science*, *5*(4), 281–294. https://doi.org/10.1007/s42761-024-00237-1

Poulton, A., & Hester, R. (2020). Transition to substance use disorders: Impulsivity for reward and learning from reward. *Social Cognitive and Affective Neuroscience*, *15*(10), 1182–1191. https://doi.org/10.1093/scan/nsz077

Quilty, L. C., & Oakman, J. M. (2004). The assessment of behavioural activation—the relationship between impulsivity and behavioural activation. *Personality and Individual Differences*, *37*(2), 429–442. https://doi.org/10.1016/j.paid.2003.09.014

Richards, J. B., Zhang, L., Mitchell, S. H., & de Wit, H. (1999). Delay or Probability Discounting in a Model of Impulsive Behavior: Effect of Alcohol. *Journal of the Experimental Analysis of Behavior*, *71*(2), 121–143. https://doi.org/10.1901/jeab.1999.71-121

Rubinstein, A. (2003). "Economics and Psychology"? The Case of Hyperbolic Discounting. *International Economic Review*, *44*(4), 1207–1216. https://doi.org/10.1111/1468-2354.t01-1-00106

Sharot, T., & Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour*, *4*(1), 14–19. https://doi.org/10.1038/s41562-019-0793-1

Speyer, L. G., Ushakova, A., Blakemore, S.-J., Murray, A. L., & Kievit, R. (2023). Testing for *Within* × *Within* and *Between* × *Within* Moderation Using Random Intercept Cross-Lagged Panel Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(2), 315–327. https://doi.org/10.1080/10705511.2022.2096613

Stern, S., Linker, S., Vadodaria, K. C., Marchetto, M. C., & Gage, F. H. (2018). Prediction of response to drug therapy in psychiatric disorders. *Open Biology*, *8*(5), 180031. https://doi.org/10.1098/rsob.180031

Story, G. W., Moutoussis, M., & Dolan, R. J. (2016). A Computational Analysis of Aberrant Delay Discounting in Psychiatric Disorders. *Frontiers in Psychology*, *6*, 1948. https://doi.org/10.3389/fpsyg.2015.01948

Strümke, I., Slavkovik, M., & Stachl, C. (2023). *Against Algorithmic Exploitation of Human Vulnerabilities* (No. arXiv:2301.04993). arXiv. https://doi.org/10.48550/arXiv.2301.04993

Svihra, M., & Katzman, M. A. (2004). Behavioural inhibition: A predictor of anxiety. *Paediatrics & Child Health*, *9*(8), 547–550. https://doi.org/10.1093/pch/9.8.547

Turel, O., He, Q., Brevers, D., & Bechara, A. (2018). Delay discounting mediates the association between posterior insular cortex volume and social media addiction symptoms. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(4), 694–704. https://doi.org/10.3758/s13415-018-0597-1

Turner, G., Gunschera, L. J., Subrahmanya, S., Salecha, A., Eichstaedt, J. C., Palminteri, S., & Orben, A. (2024). *A computational model of reward learning and habits on social media*. https://doi.org/10.31234/osf.io/xe25k

Vahedi, Z., & Zannella, L. (2021). The association between self-reported depressive symptoms and the use of social networking sites (SNS): A meta-analysis. *Current Psychology*, *40*(5), 2174–2189. https://doi.org/10.1007/s12144-019-0150-6

Valkenburg, P. M., Beyens, I., de Vaate, N. B., Janssen, L., & Wal, A. V. D. (2024). Person-Specific Media Effects. In *Communication Research into the Digital Society* (pp. 233–245). Amsterdam University Press. https://doi.org/10.1515/9789048560608-015

Valkenburg, P. M., Beyens, I., Pouwels, J. L., van Driel, I. I., & Keijsers, L. (2021). Social Media Use and Adolescents' Self-Esteem: Heading for a Person-Specific Media Effects Paradigm. *Journal of Communication*, *71*(1), 56–78. https://doi.org/10.1093/joc/jqaa039

van der Pol, M., & Cairns, J. (2002). A comparison of the discounted utility model and hyperbolic discounting models in the case of social and private intertemporal preferences for health. *Journal of Economic Behavior & Organization*, *49*(1), 79–96. https://doi.org/10.1016/S0167-2681(02)00059-8

Weinsztok, S., Brassard, S., Balodis, I., Martin, L. E., & Amlung, M. (2021). Delay Discounting in Established and Proposed Behavioral Addictions: A Systematic Review and Meta-Analysis. *Frontiers in Behavioral Neuroscience*, *15*. https://doi.org/10.3389/fnbeh.2021.786358

Weyn, S., Van Leeuwen, K., Pluess, M., Lionetti, F., Greven, C. U., Goossens, L., Colpin, H., Van Den Noortgate, W., Verschueren, K., Bastin, M., Van Hoof, E., De Fruyt, F., & Bijttebier, P. (2021). Psychometric properties of the Highly Sensitive Child scale across developmental stage, gender, and country. *Current Psychology*, *40*(7), 3309–3325. https://doi.org/10.1007/s12144-019-00254-5

Wise, T., Robinson, O. J., & Gillan, C. M. (2023). Identifying Transdiagnostic Mechanisms in Mental Health Using Computational Factor Modeling. *Biological Psychiatry*, *93*(8), 690–703. https://doi.org/10.1016/j.biopsych.2022.09.034

Supplementary Material

Supplementary Method 1: Hyperbolic Delay Discounting Model

We modelled participants' temporal preferences on the delay discounting task using a hyperbolic function. The hyperbolic model is an established framework of examining intertemporal choice behaviour and posits that the subjective value V of a delayed reward decreases non-linearly as a function of time (Rubinstein, 2003; Story et al., 2016; van der Pol & Cairns, 2002). The hyperbolic function formulates how the subjective value (V) of an objective reward (A) changes as a function of increases in reward delay (D).

$$V = \frac{A}{1 + k * D} \tag{1}$$

The free parameter k determines the rate at which increases in delay result in a decreased subjective reward value. Larger values of k translate to steeper decreases in subjective reward value with increasing delays. On a given trial (t), the subjective reward value for the sooner (V_{SS}) and later (V_{LL}) options were estimated and translated into choice probabilities using softmax (Luce's choice rule).

$$P(\mathrm{LLoverSS}) = \frac{1}{1 + e^{-\beta(V_{SS(t)} - V_{LL(t)}}} \tag{2}$$

The softmax includes an inverse temperature parameter, the second free parameter of the model, and controls an individual's sensitivity to differences in subjective value between the sooner and later option. A higher β indicates more consistent (deterministic) choice patterns, whereas a lower β indicates more stochastic choices on the delay discounting task.

We took a hierarchical Bayesian approach to model fitting (Ahn et al., 2013; Piray et al., 2019), implemented in CmdStan R (Gabry et al., 2023), with code adapted from hBayesDM (Ahn et al., 2017). All models were fit using Markov-Chain Monte Carlo (MCMC), with 2000 warm-up iterations and 10,000 sampling iterations, by four chains. Model convergence was established by visual inspection of trace plots (Supplement Figure S1), and effective sample size (ESS) and R-hats. We validated our modelling procedure by examining behavioural correlates of computational processes, parameter recovery, and posterior predictive checks. We log-transformed the delay discounting parameter (-log(*k*); Weinsztok et al., 2021).

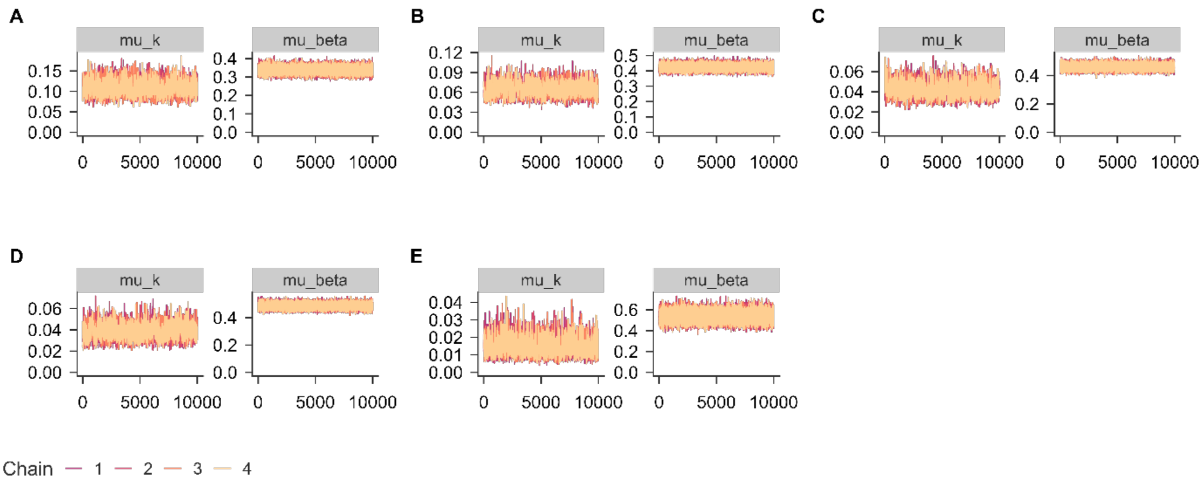

Figure S1: Trace Plots for the Hyperbolic Delay Discounting Model. A: *Trace plot for parameters k and β for L-CID measurement wave two.* **B:** *Trace plot for parameters k and β for L-CID measurement wave three.* **C:** *Trace plot for parameters k and β for L-CID measurement wave four.* **D:** *Trace plot for parameters k and β for L-CID measurement wave five.* **E:** *Trace plot for parameters k and β for L-CID measurement wave six.*

Supplementary Method 2: Hyperbolic Delay Discounting Parameter Recovery

To validate the model-based analyses, we performed a parameter recovery. We generated synthetic data from the hyperbolic model using known parameter values. We sampled delay discounting parameters (k) from a beta distribution (α = .75, β = 3) and inverse temperature parameters (β) from a uniform distribution (0-5). A beta distribution for the delay discounting parameters was chosen to reflect the skew in k parameters, reflected in the fact that most studies report the natural logarithms of the delay discounting parameter estimates. We simulated task data of 10,000 agents based on the above modelling equations and parameter configurations. We compared the parameter estimates (recovered parameters) returned when fitting the hyperbolic model to the synthetic data to the generating parameter values. The generating parameters correlated strongly with the recovered parameter values (Pearson's $r(998) = 0.795$, 95% CI [0.771, 0.817], $p < 0.001$). The parameter recovery is displayed in Figure S2.

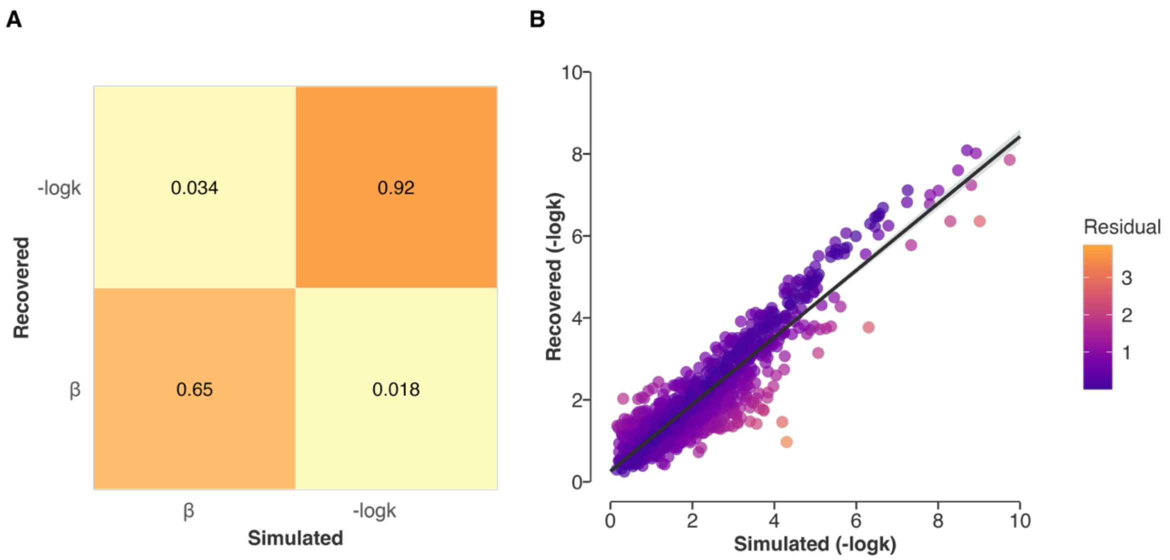

Figure S2: Parameter Recovery for the Hyperbolic Model. A: *Correlation matrix of the simulated and recovered free model parameters, and their cross-correlations.* **B:** *Delay discounting parameters were simulated from a beta (α = .75, β = 3) distribution and negative log-transformed. Colour is used to indicate the residual of the predicted from the simulated parameter values, with brighter colours indicating less accurate predictions.*

Supplementary Method 3: Hyperbolic Delay Discounting Posterior Predictive Checks

To examine whether the hyperbolic delay discounting model is sensitive to capturing relevant features of the data, we performed posterior predictive checks. We simulated data based on the fitted model parameters and compared the observed and model-predicted choice data. Our results indicated adequate agreement between observed and model-predicted choice data across measurement occasions and across effort levels. Accordance between the observed and predicted choice data increased as children became older, from Pearson's r = 0.683 (95% CI [0.658, 0.707], p < 0.001) in measurement wave two to Pearson's r = 0.754 (95% CI [0.704, 0.796], p < 0.001) in wave six.

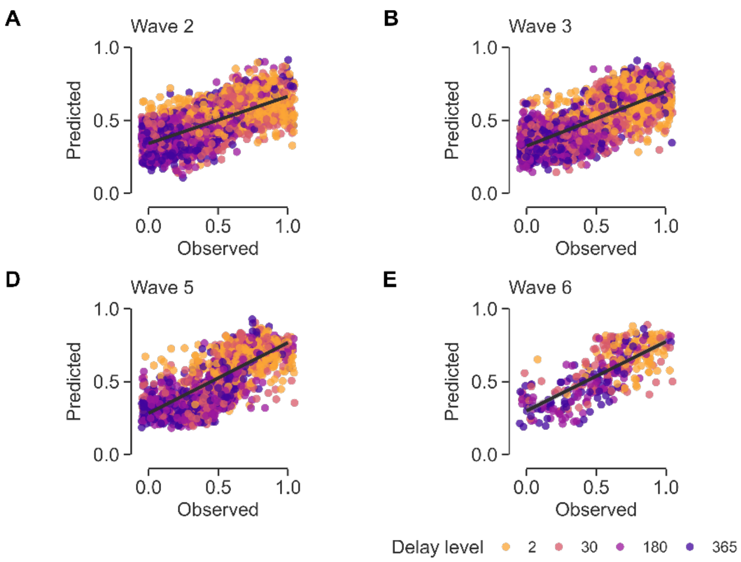

Figure S3: Posterior Predictive Model Performance Across Measurement Occasions. A-E: *Visualisation of the observed and predicted proportions of later offers accepted, with one equivalent to accepting all and zero equivalent to accepting no later offers. The posterior predictive checks are presented for each measurement wave of the Leiden Consortium on Individual Development dataset. Choice data were simulated using the parameters estimated by the hyperbolic model of the observed choice data.*

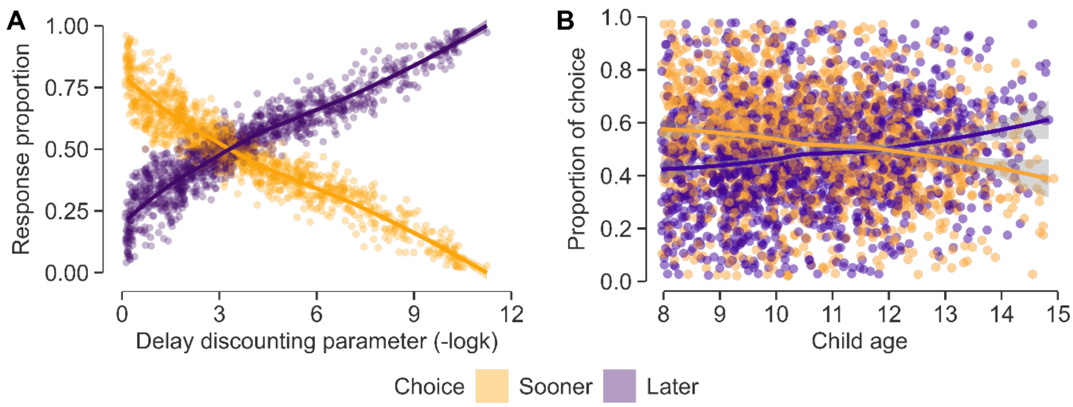

Figure S4: Proportion of Sooner and Later Offers Accepted and Estimated Delay Discounting Parameter. A: *The proportion of responses that participants accept the sooner and later options relates to the estimated delay discounting parameter. Specifically, participants who accept a larger proportion of sooner offers should have a smaller negative log-transformed delay discounting parameter.* **B:** *The proportion of sooner and later response options which participants selected as a function of age. Across the entire sample, we observe an increase in the proportion of accepted later response options, which is reflected in the model delay discounting estimates.*

Supplementary Method 4: Summarised Model Estimates

We fit separate random intercept cross-lagged panel models for each mental health outcome. The baseline models include the autoregressive (AR) and cross-lagged (CL) paths between social media use and the respective mental health outcome. All models were fit using a Bayesian estimator in Mplus, treating missing data as additional unknown quantities that were sampled from the conditional posterior distribution. The estimates for the baseline model, and the moderation model – which includes moderation terms for the cross-lagged paths between social media use and mental health outcome – are presented below.

Variables	**BIS**	**BAS**	**SDQ**	**HSCS**	**EATQ-EC**
Within-Effects	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)
AR $SM_5 \rightarrow SM_6$	**0.364** (0.102 /0.557)	**0.349** (0.068/0.557)	**0.416** (0.103/0.592)	**0.362** (0.075/0.570)	**0.364** (0.083/0.562)
AR $SM_6 \rightarrow SM_7$	-0.095 (-0.509/0.232)	-0.097 (-0.508/0.230)	0.026 (-0.373/0.306)	-0.058 (-0.460/0.254)	-0.007 (-0.445/0.314)
AR $MH_5 \rightarrow MH_6$	0.295 (-0.019/0.560)	**0.323** (0.068/0.524)	0.142 (-0.213/0.386)	**0.373** (0.141/0.546)	0.093 (-0.112/0.266)
AR $MH_6 \rightarrow MH_7$	-0.005 (-0.417/0.341)	-0.038 (-0.335/0.329)	0.117 (-0.131/0.347)	0.215 (-0.066/0.442)	0.074 (-0.239/0.323)
CL $SM_5 \rightarrow MH_6$	-0.182 (-0.397/0.055)	-0.076 (-0.297 /0.167)	0.137 (-0.130/0.354)	0.027 (-0.193/0.223)	0.066 (-0.158/0.305)
CL $SM_6 \rightarrow MH_7$	-0.171 (-0.554/0.154)	-0.006 (-0.319 /0.338)	0.192 (-0.103/0.458)	-0.042 (-0.345/0.250)	-0.129 (-0.401/0.196)
CL $MH_5 \rightarrow SM_6$	-0.090 (-0.349/0.205)	-0.045 (-0.264/0.197)	-0.052 (-0.326/0.184)	0.009 (-0.229/0.233)	0.112 (-0.061/0.297)
CL $MH_6 \rightarrow SM_7$	-0.193 (-0.515/0.099)	-0.044 (-0.353/0.262)	0.141 (-0.098/0.371)	-0.062 (-0.323/0.196)	-0.068 (-0.315/0.258)

Table S1. Summarised Model Estimates for all Baseline Models. *All estimates are standardised. CI_L = 2.5% Credible Interval, CI_U = 97.5% Credible Interval, AR = autoregressive effect, CL = cross-lagged effect, SM = social media use, MH = mental health indicator, ^{B} = between component, numbers in subscript indicate measurement wave, significant standardised effects are indicated by bold font and based on credible intervals not containing zero.*

Variables	BIS	BAS	SDQ	HSCS	EATQ-EC
Within-Effects	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)
AR $SM_5 \rightarrow SM_6$	0.323 (-0.001/0.556)	**0.292** (0.021/0.538)	**0.383** (0.082/0.581)	**0.350** (0.057/0.563)	**0.361** (0.076/0.567)
AR $SM_6 \rightarrow SM_7$	-0.191 (-0.615/0.181)	-0.334 (-1.015/0.177)	-0.007 (-0.404/0.289)	-0.087 (-0.477/0.242)	-0.022 (-0.450/0.296)
AR $MH_5 \rightarrow MH_6$	0.102 (-0.432/0.454)	**0.397** (0.162/0.584)	0.129 (-0.244/0.394)	**0.355** (0.069/0.535)	0.089 (-0.124/0.261)
AR $MH_6 \rightarrow MH_7$	0.027 (-0.330/0.360)	0.178 (-0.278/0.457)	0.113 (-0.145/0.357)	0.215 (-0.133/0.434)	0.075 (-0.247/0.322)
CL $SM_5 \rightarrow MH_6$	-0.257 (-0.535/0.057)	-0.834 (-1.811/0.090)	0.125 (-0.128/0.347)	0.041 (-0.196/0.243)	0.051 (-0.172/0.291)
CL $SM_6 \rightarrow MH_7$	-0.215 (-0.557/0.114)	-0.302 (-2.015/1.336)	0.176 (-0.147/0.447)	-0.049 (-0.329/0.231)	-0.146 (-0.418/0.185)
CL $MH_5 \rightarrow SM_6$	0.041 (-0.253/0.373)	-0.003 (-0.049/0.043)	-0.059 (-0.337/0.170)	0.027 (-0.210/0.243)	0.106 (-0.069/0.296)
CL $MH_6 \rightarrow SM_7$	-0.139 (-0.453/0.174)	0.002 (-0.101/0.103)	0.129 (-0.115/0.357)	-0.032 (-0.295/0.243)	-0.086 (-0.340/0.237)
Between*Within	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)	Estimate (CI_L/CI_U)
DD^B : CL $SM_5 \rightarrow MH_6$	-0.032 (-0.281/0.237)	**0.470** (0.071/0.872)	0.017 (-0.170/0.186)	-0.069 (-0.233/0.100)	0.065 (-0.113/0.241)
DD^B : CL $SM_6 \rightarrow MH_7$	-0.025 (-0.230/0.168)	**0.470** (0.071/0.872)	0.014 (-0.149/0.165)	-0.067 (-0.232/0.097)	0.057 (-0.105/0.216)

Table S2. Summarised Model Estimates for all Moderation Models. *All estimates are standardised. CI_L = 2.5% Credible Interval, CI_U = 97.5% Credible Interval, AR = autoregressive effect, CL = cross-lagged effect, SM = social media use, MH = mental health indicator, ^{B} = between component, numbers in subscript indicate measurement wave, significant standardised effects are indicated by bold font and based on credible intervals not containing zero.*

Supplementary Method 5: Random Intercept Cross-lagged Panel Models

We fit five random-intercept cross-lagged panel models, one for each well-being outcome included: BIS subscale, BAS subscale, SDQ, HSCS, EATQ-EC. Below, we display the results of each of the models, including all cross-lagged, autoregressive, and moderation effects. We first present the baseline model for the BAS subscale, as the corresponding moderation model is presented in the main manuscript.

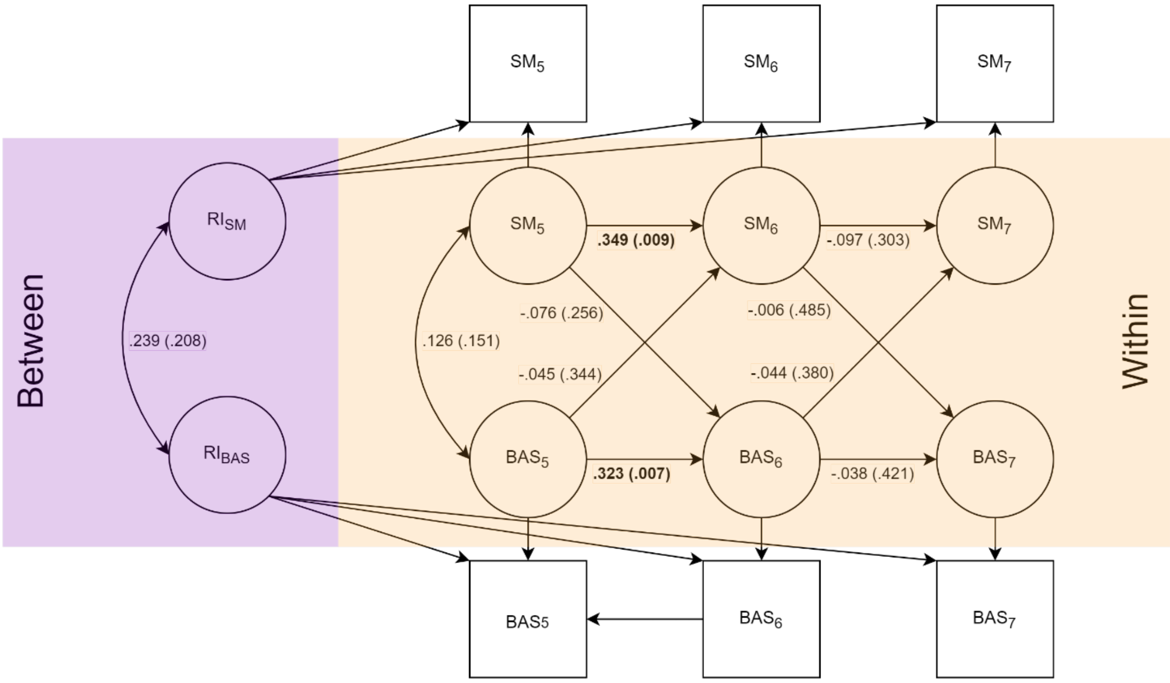

Figure S5: Baseline RI-CLPM for BAS Subscale. *Displayed are the standardised regression coefficients and the corresponding p-values. The corresponding RI-CLPM for the BAS Subscale including the moderation terms is presented in the manuscript, Figure 3.*

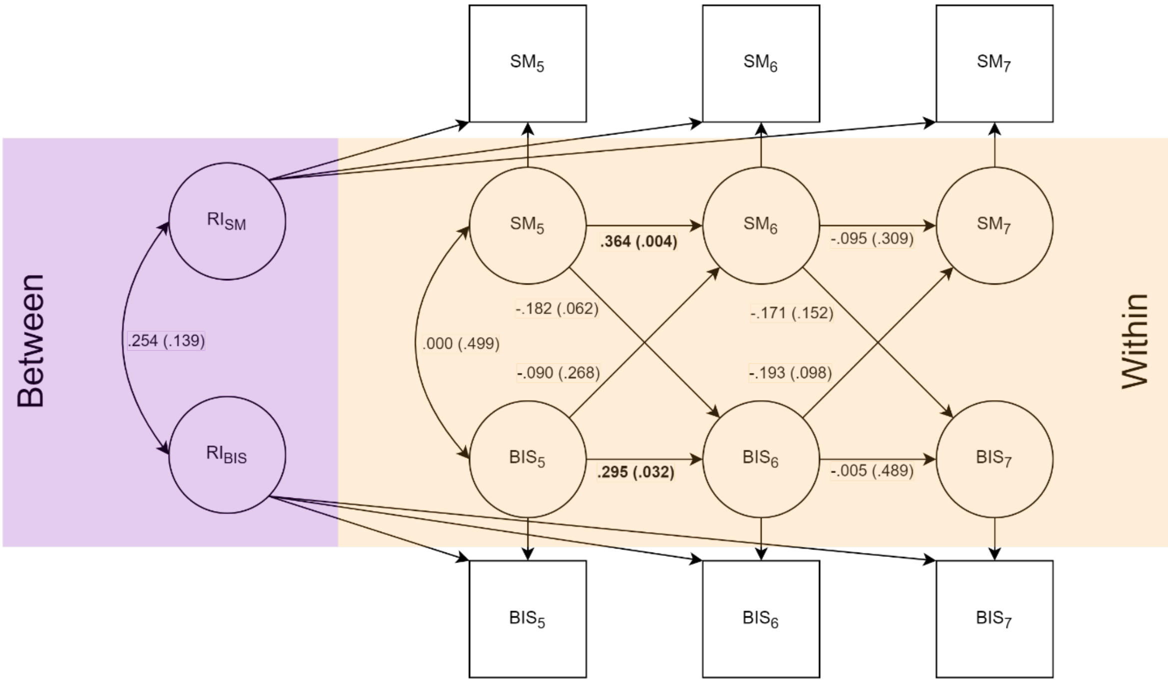

Figure S6a: Baseline RI-CLPM for the BIS Subscale. *Displayed are the standardised regression coefficients and the corresponding p-values.*

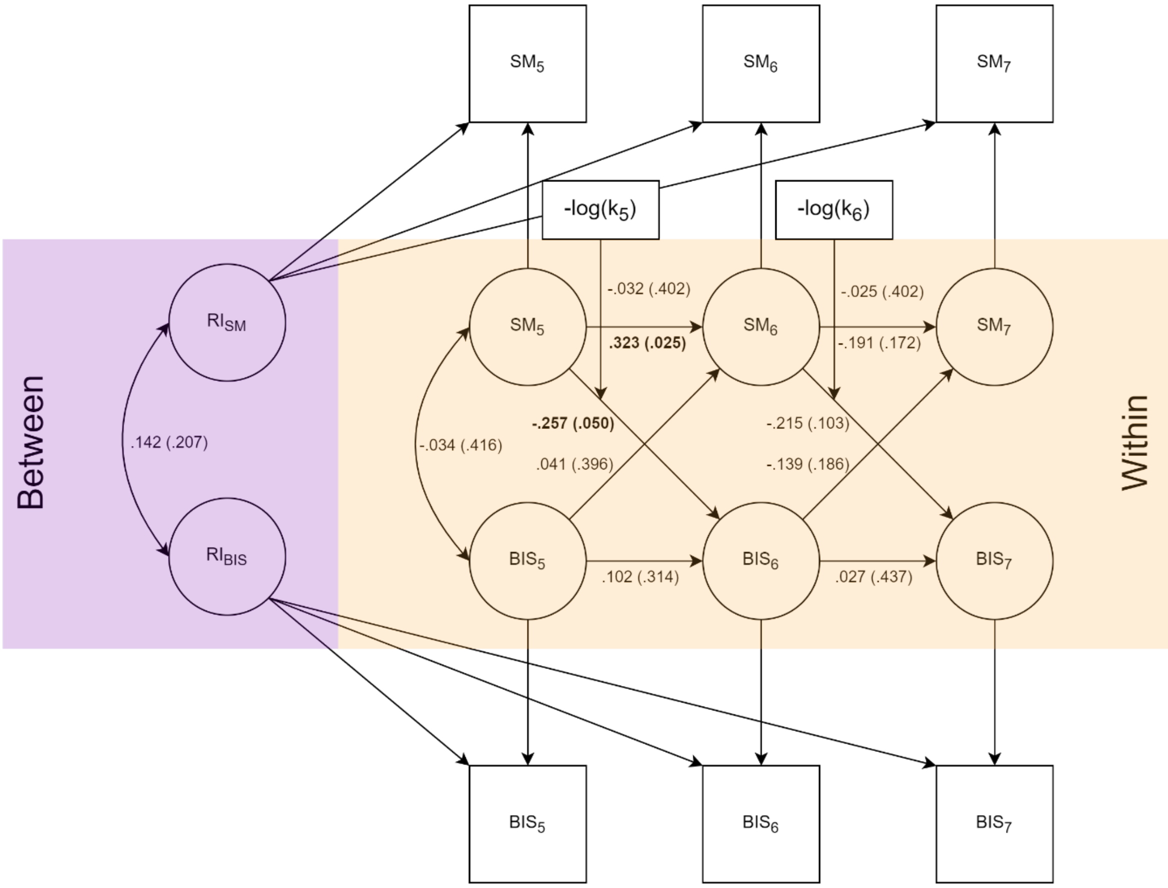

Figure S6b: Within x Between Interaction RI-CLPM for BIS Subscale. *Displayed are the standardised regression coefficients and the corresponding p-values.*

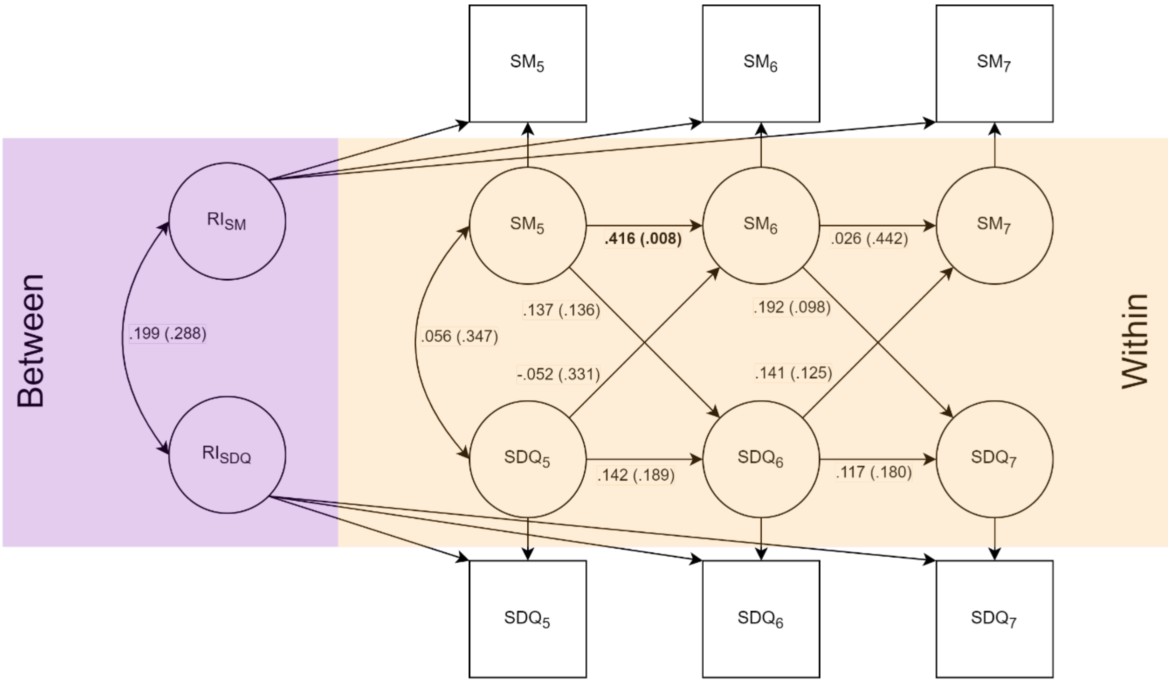

Figure S7a: Baseline RI-CLPM for SDQ. *Displayed are the standardised regression coefficients and the corresponding p-values.*

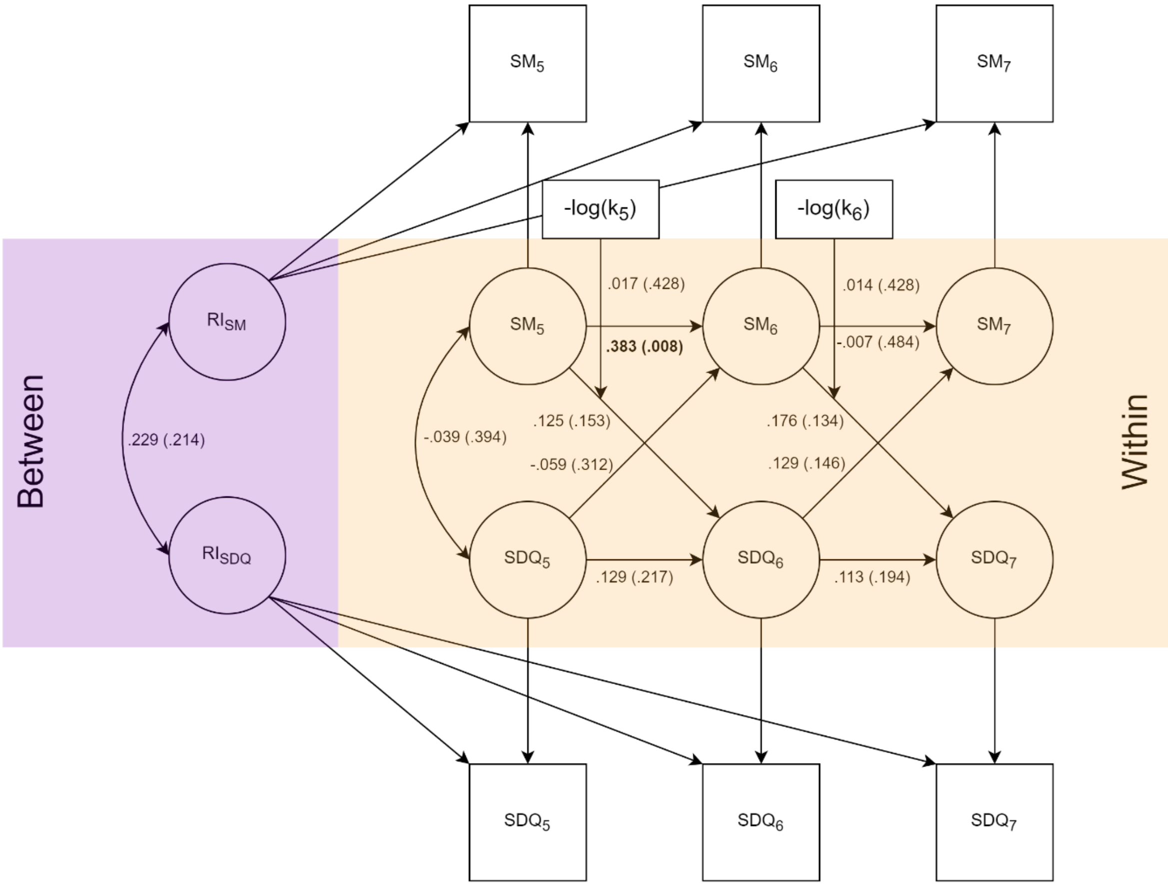

Figure S7b: Within x Between Interaction RI-CLPM for SDQ. *Displayed are the standardised regression coefficients and the corresponding p-values.*

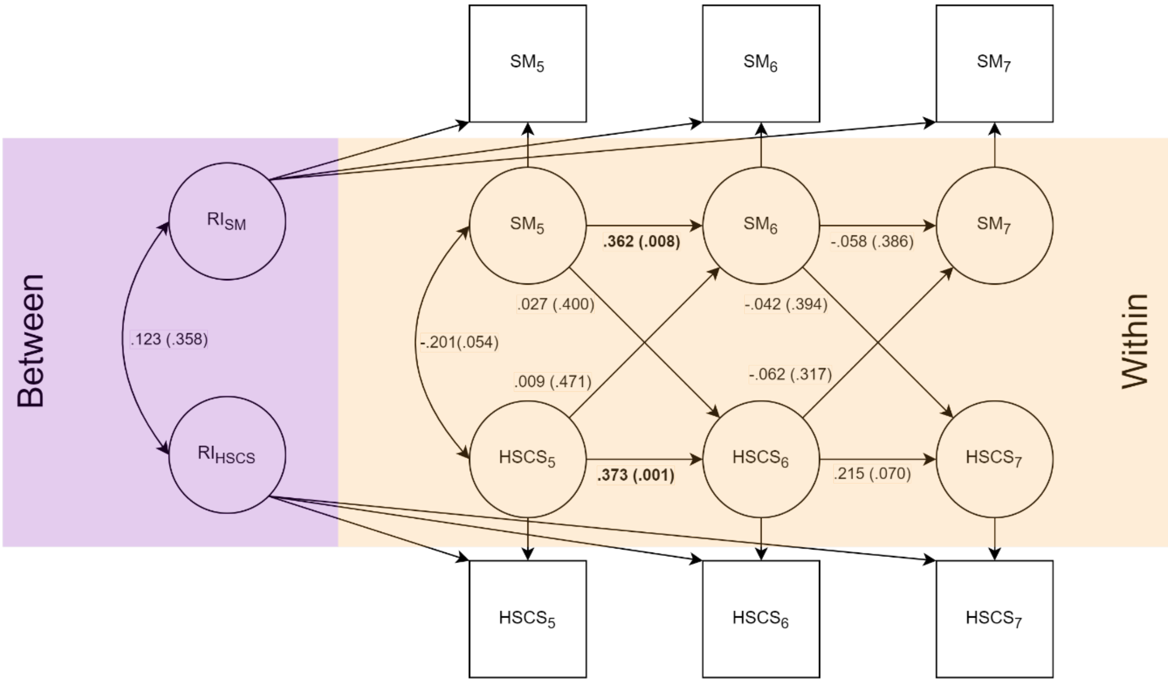

Figure S8a: Baseline RI-CLPM for HSCS. *Displayed are the standardised regression coefficients and the corresponding p-values.*

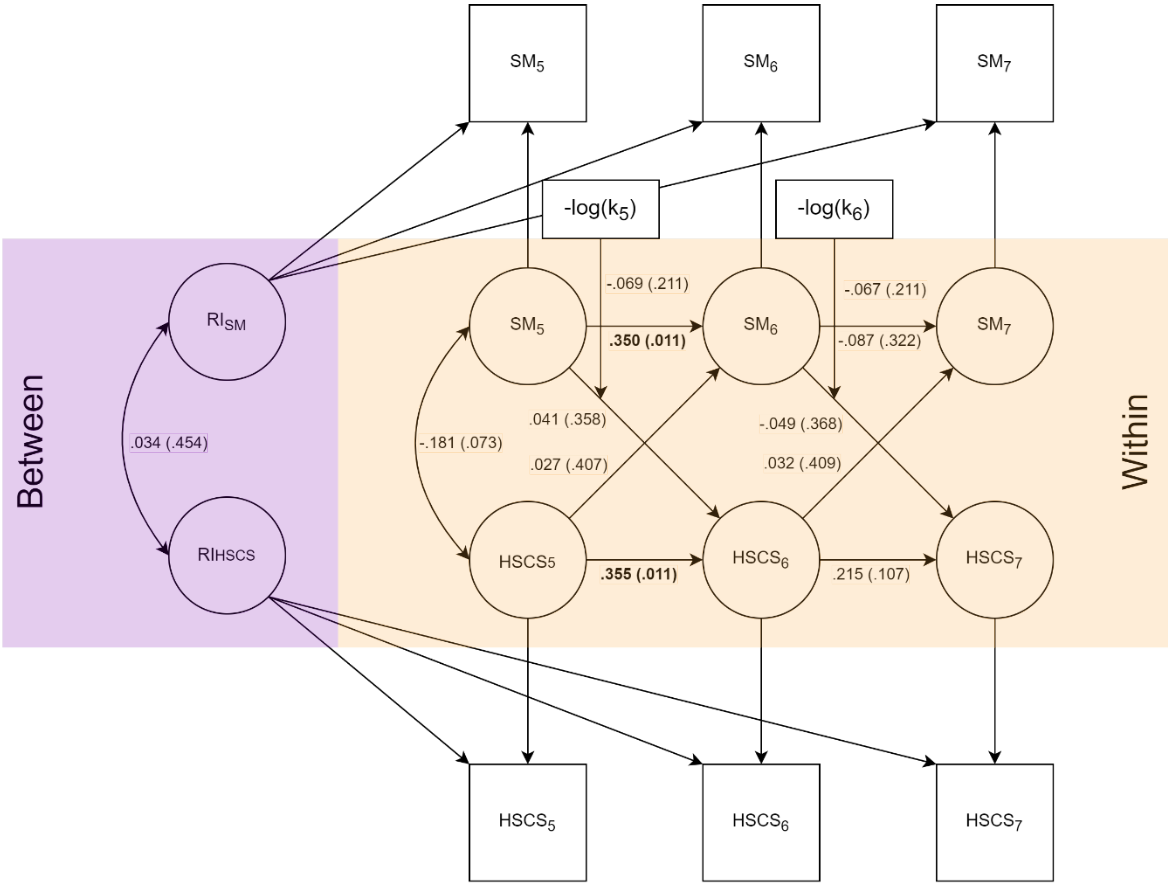

Figure S8b: Within x Between Interaction RI-CLPM for HSCS. *Displayed are the standardised regression coefficients and the corresponding p-values.*

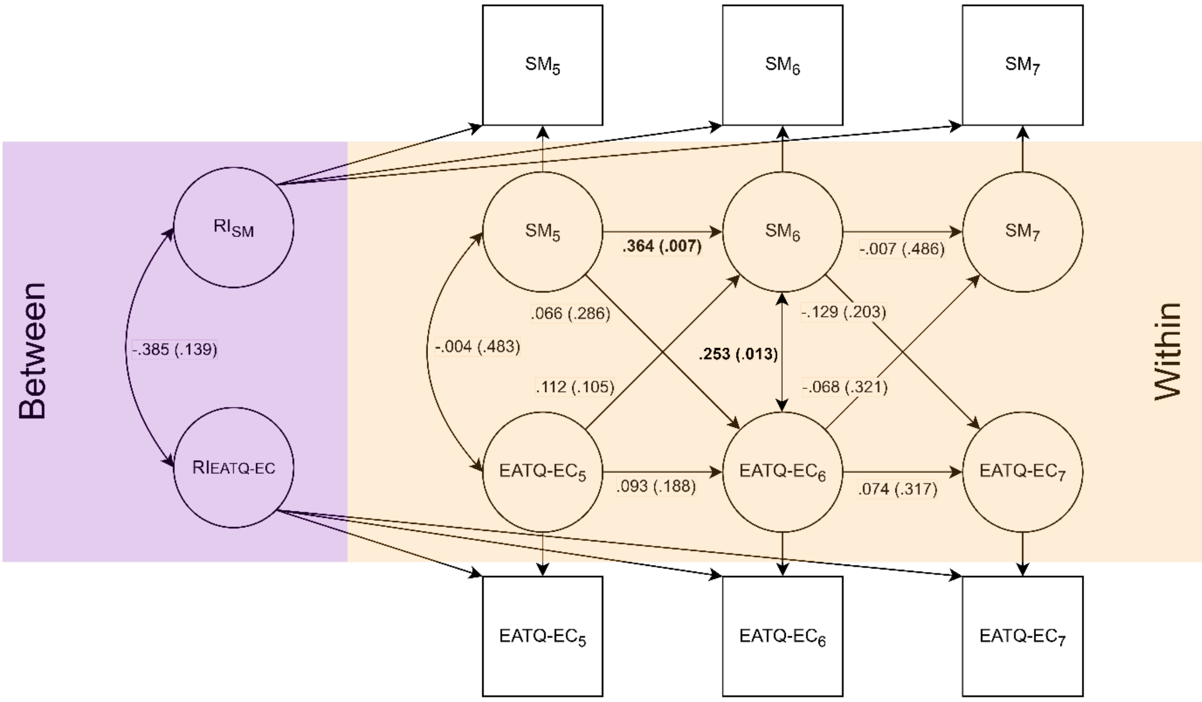

Figure S9a: Baseline RI-CLPM for EATQ-EC. *Displayed are the standardised regression coefficients and the corresponding p-values.*

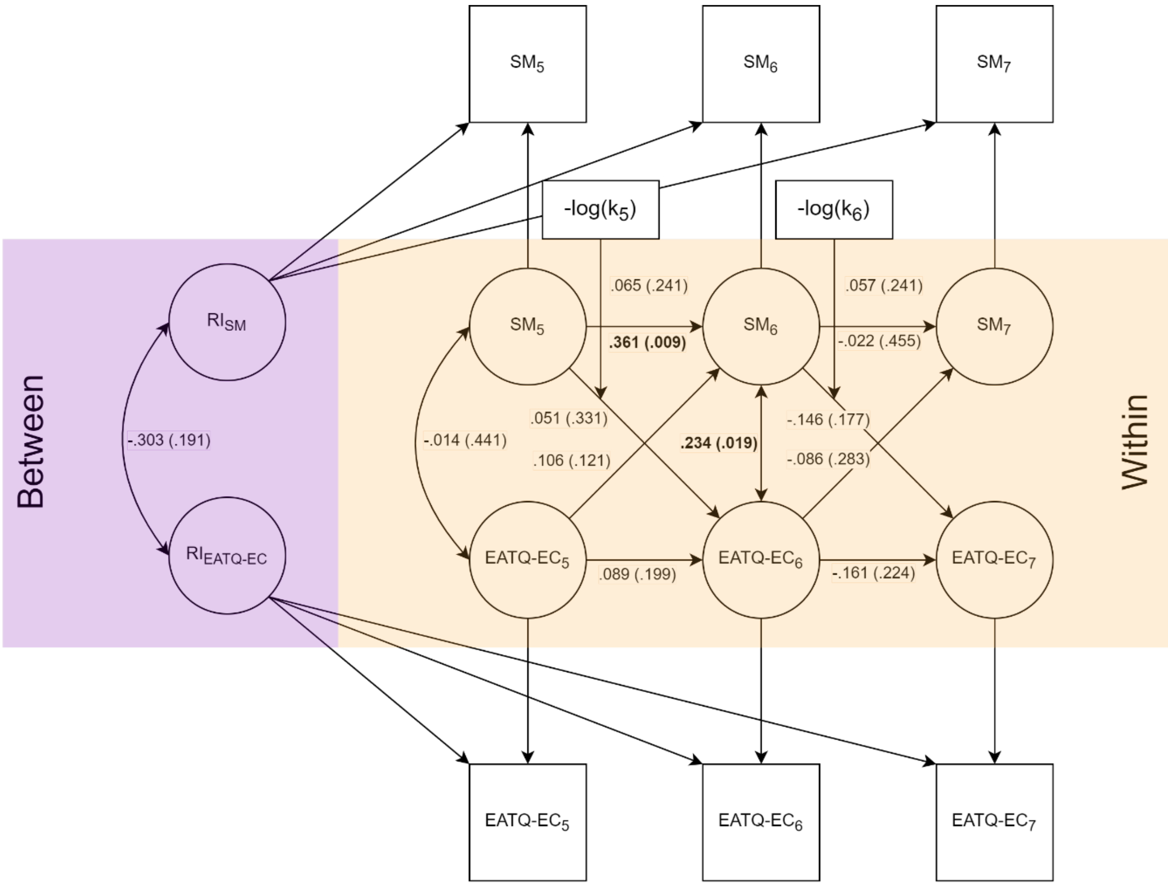

Figure S9b: Within x Between Interaction RI-CLPM for EATQ-EC. *Displayed are the standardised regression coefficients and the corresponding p-values.*

Supplementary Method 6: Developmental Trajectories

Social media use

To examine developmental trends in the time children spend on social media, we examined their responses to the question "on an average weekday, how much time do you spend on your phone or iPad/ tablet viewing public messages, photos, or videos (e.g., Facebook, Instagram, TikTok, Snapchat, Twitter)" across the measurement occasions where it was added to the research protocol (waves five to seven).

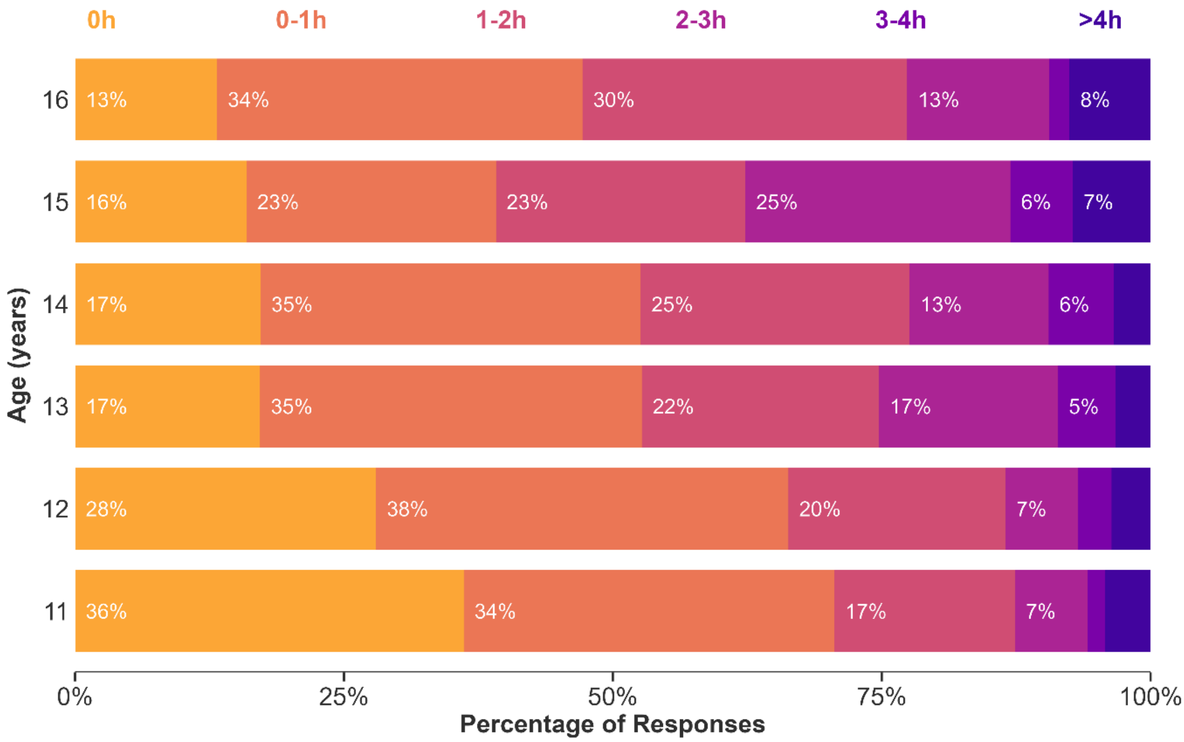

Figure S10: Time Spent Posting and Scrolling on Social Media. *Participants rated the question "on an average weekday, how much time do you spend on your phone or iPad/ tablet viewing public messages, photos, or videos (e.g., Facebook, Instagram, TikTok, Snapchat, Twitter)" at each measurement wave. The proportion of each Likert response option selected is presented unless the option was selected by fewer than 5% of all participants of a given age group. The answer options ranged from zero to more than four hours.*

Mental health indicators

Participants completed mental health questionnaires at timepoints five, six, and seven. Below, we present the distributions of the questionnaire scores across these measurement occasions.

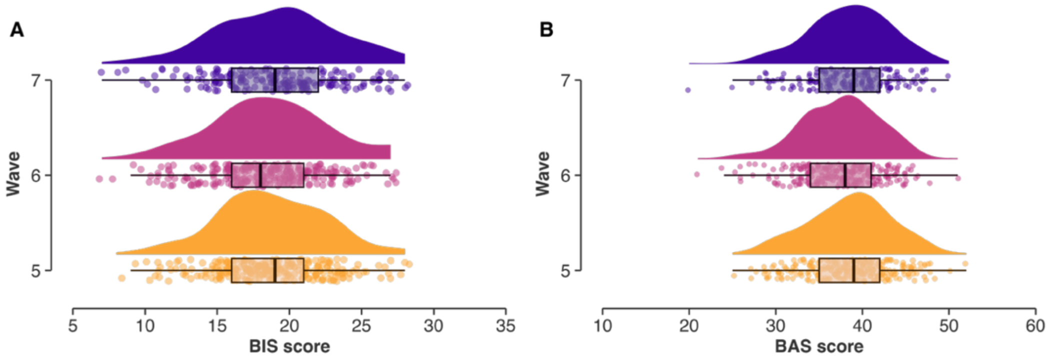

Figure S11: Behavioural Inhibition Behavioural Activation Scale Scores. A: *BISBAS behavioural inhibition subscale score.* **B:** *BISBAS behavioural activation subscale score.*

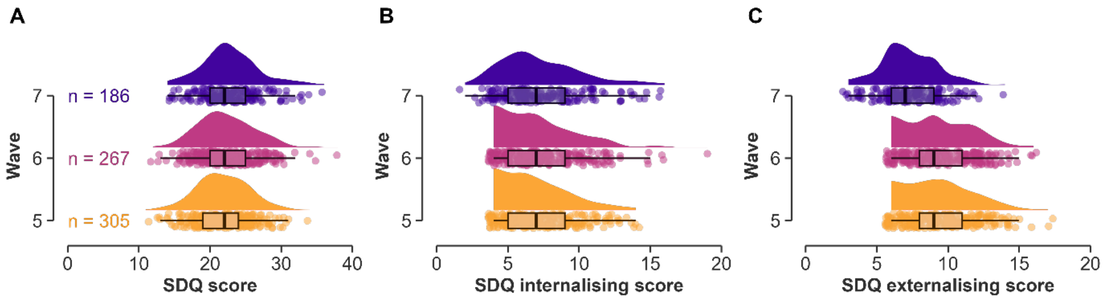

Figure S12: Strengths and Difficulties Questionnaire scores. A: *BISBAS total score.* **B:** *BISBAS behavioural inhibition subscale score.* **C:** *BISBAS behavioural activation subscale score.*

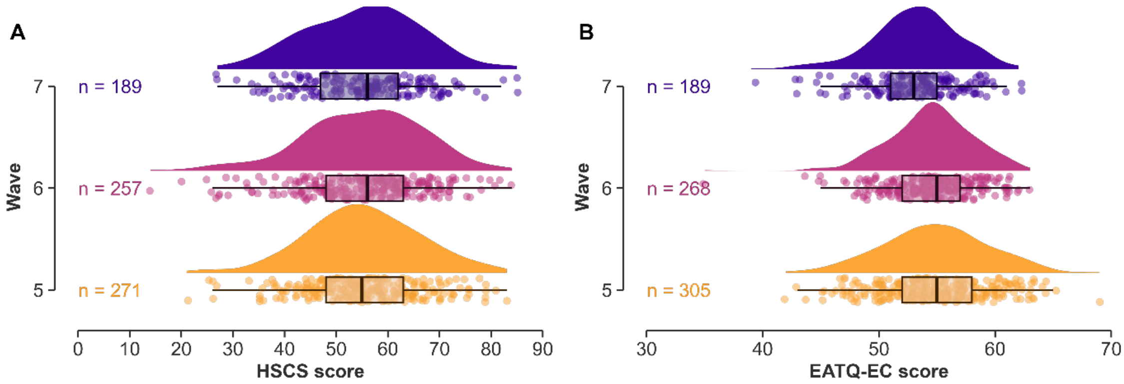

Figure S13: Highly Sensitive Child Scale and Early Adolescent Temperament Questionnaire, Effortful Control Subscale. A: *Highly Sensitive Child questionnaire total scores for each measurement wave.* ***B:*** *Early Adolescent Temperament questionnaire Effortful Control subscale score for each measurement wave.*

Supplementary Method 7: Delay Discounting and Social Media Use

We examined the relationship between the estimated delay discounting parameters and the five mental health indicators included in the analyses. The data for each indicator are visualised in Figure S7. We estimated separate linear mixed effects models for each mental health indicator, including a fixed effect for the delay discounting parameter estimates and random intercepts for participants. The discounting parameter remained non-significant across the linear mixed models for BIS, BAS, SDQ, and HSCS (all *p*s > 0.197) but showed a significant positive association between delay discounting parameter (-logk) and EATQ-EC score ($t(373.687) = 2.021$, $\beta = 0.105$, 95% CI [0.003, 0.207], $p = 0.044$), indicating that participants who discount future rewards less tend to have higher scores on the EATQ-EC subscale.

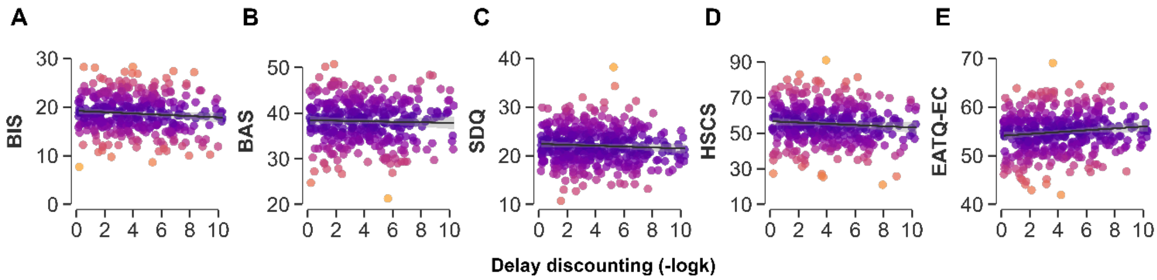

Figure S14: Associations Between Delay Discounting Parameters and Wellbeing Indicators Across Participants. *The presented data points collate data across all participants, and lighter values indicate larger residuals. The correlations are presented for the delay discounting parameter (-logk) and* ***A:*** *BIS subscale score,* ***B:*** *BAS subscale,* ***C:*** *SDQ,* ***D:*** *HSCS score, and* ***E:*** *EATQ-EC subscale score.*

Supplementary Method 8: Multigroup Confirmatory Factor Analysis

The mental health questionnaires may assess overlapping symptoms, and dimensionality reduction approaches have been used to deal with this. However, these approaches require measurement invariance across repeated measurement occasions, which we examined using multigroup confirmatory factor analyses for the five included self-report questionnaire measures. The results are summarised in Supplementary Table S3.

Measure	Model	χ^2 ($\Delta\chi^2$)	Df (ΔDf)	*p*	CFI (ΔCFI)
BIS	Configural	127.159	42	/	0.927
	Weak	(30.172)	(12)	0.003	(0.015)
	Strong	(36.171)	(12)	<0.001	(0.021)
	Strict	(9.097)	(2)	0.011	(0.006)
BAS	Configural	554.563	195	/	0.764
	Weak	(48.196)	(24)	0.002	(0.016)
	Strong	(144.979)	(24)	<0.001	(0.079)
	Strict	(8.233)	(2)	0.016	(0.004)
SDQ	Configural	2528.346)	825	/	0.454
	Weak	(107.203)	(48)	<0.001	(0.019)
	Strong	(113.178)	(48)	<0.001	(0.021)
	Strict	(5.223)	(2)	0.073	(0.001)
HSCS	Configural	936.467	195	/	0.665
	Weak	(37.181)	(24)	0.042	(0.006)
	Strong	(100.871)	(24)	<0.001	(0.035)
	Strict	(0.138)	(2)	0.934	(0.001)
EATQ-EC	Configural	954.475	405	/	0.704
	Weak	(45.734)	(34)	0.086	(0.006)
	Strong	(248.399)	(34)	<0.001	(0.115)
	Strict	(17.554)	(2)	<0.001	(0.008)

Table S3. Multigroup Confirmatory Factor Analysis Results. *The results indicate that the assumptions required to use computational factor modelling on longitudinal data are violated.*

Supplementary Method 9: Preregistration Deviation

The subset of task and questionnaire measures available for the present research is outlined in Table 1 and the procedure in Figure 3. Both are described in detail in the sections below. We set out to analyse participants' trial-level data from the temporal delay discounting task and the stop signal reaction time task. However, we report analyses and results for the delay discounting task data only. This deviation from the pre-registration followed concerns surrounding the data quality of the stop signal reaction time trial level data, which was inconsistent and featured errors in coding and data values in the sixth measurement wave. Since it was not possible to recover the error source, we had to exclude the sixth wave of the stop signal reaction time task and subsequently exclude the task altogether, given that the planned longitudinal analyses were not possible with two measurement waves alone.

The second deviation was a change to the analysis plan, after accessing the data but prior to performing the statistical tests. This change involved examining the BIS and BAS subscales of the BISBAS questionnaire separately (Jorm et al., 1998; Maack & Ebesutani, 2018; Quilty & Oakman, 2004). This change was made to ensure the interpretability of resulting outcomes, as the subscales are considered orthogonal constructs (Carver & White, 1994; Jorm et al., 1998; Maack & Ebesutani, 2018), the combination of which would have resulted in theoretically and psychometrically invalid conclusions. Both deviations are described in detail in the Preregistration Deviation Table S4 (Willroth & Atherton, 2024).

Deviations

#	Details		Original Wording	Deviation Description	Reader Impact
1	Type	Variables	“We will model the stop signal task data using a hybrid racing-diffusion ex-Gaussian stop-signal model (RDEX; Tanis et al., 2024)." "We will fit separate RI-CLPMs for each combination of mental health outcome (n = 4) and cognitive parameters (p = 2) to estimate the bidirectional relationships between social media use and mental health, as well as the moderating effect of the cognitive parameters (p) on this relationship."	We set out to test the presented cognitive framework to studying the relationship between social media use and mental health across multiple cognitive processes. Specifically, we intended to model the cognitive task data of the monetary delay discounting task and the stop signal reaction time task. Upon getting access to the stop signal reaction time data, we noticed errors in the coding of the data of the 6th measurement occasion (which was administered online). The datafiles of participants from this wave features numerous instances where the observed/coded behaviour of participants did not align with the coding of the responses. For example, in some instances the saved response of a participant and the correct response matched, but were evaluated as false. Given the frequency of	This deviation constrains the scope of our findings to a single cognitive process. The exclusion of the stop signal reaction time task due to irresolvable data quality issues reduces the range of cognitive processes we were able to examine, which would have allowed us to test the framework more comprehensively. Readers should interpret the present findings as testing the cognitive framework for the cognitive process of delay discounting, which warrants further investigation across a range of cognitive processes that may be of interest to the relationship between social media use and mental health.
	Reason	Plan not possible			
	Timing	After data access			

				these inconsistencies as well as their non-systematic nature, we were unable to recover the source of the errors. Subsequent attempts to resolve the data quality concerns with the people who had been involved with the collection of the data were unsuccessful and we eventually had to exclude the sixth wave of the stop signal reaction time task for data quality concerns. The planned random intercept cross-lagged panel models were not possible with two measurement waves, and we considered a measurement interval of two years too substantial to yield meaningful insights into the bidirectional relationships between social media use and mental health outcomes.	
2	Type	Analysis	"We will examine the relationship between social media use and mental health using a combination of questionnaire measures. We will examine the total scores of […] the Behavioural Inhibition and Activation	We preregistered the analyses for the total score of the Behavioural Inhibition Behavioural Activation Scale as a single outcome variable. This decision was based on a limited understanding of the measure's psychometric structure at the time of preregistration.	This deviation represents a correction of a psychometric error in the preregistration. Failure to correct this mistake would have compromised the validity and interpretability of the results. Importantly, the decision to analyse the subscales of the
	Reason	New knowledge			
	Timing	After data access			

			System (BIS/BAS; Carver & White, 1994) […]." "Therefore, we will include four outcome variables, EATQ-EC total score, BIS/BAS total score, HSC total score, and SDQ total score."	Upon closer inspection of the analysis plan, we recognised that analyses of the BIS/BAS total score are inappropriate. BIS and BAS represent theoretically orthogonal motivational systems – not extremes on a single dimension. The behavioural inhibition system reflects sensitivity to punishment and threat, while behavioural activation reflects sensitivity to reward and approach motivation. These systems operate independently, meaning individuals can exhibit high levels of both, neither, or a combination thereof. The combination of the two subscales can obscure meaningful patterns and produce psychometrically invalid and uninterpretable results. This is reflected in numerous studies examining the psychometric properties of the scale (Carver & White, 1994; Jorm et al., 1998; Maack & Ebesutani, 2018), and reflected in the consensus in the extant literature to examine the distinct subscales separately (e.g., Heffer et al., 2021; Kelley et	BIS/BAS scale was made based on theoretical soundness rather than in a data-driven manner. Readers should interpret the separate BIS and BAS results as providing theoretically meaningful test

				al., 2019). The mistake was noticed after gaining access to the data, as data access was granted upon submission of the preregistration. However, the mistake was noticed and corrected prior to inspecting the data or performing any analyses.	

Table S4. Preregistration Deviations. *We follow established guidelines for transparent reporting of deviations from preregistration, providing a comprehensive discussion of the changes, their rationales, and their implications for the interpretation of the results (Willroth & Atherton, 2024). We made two substantive deviations: (1) excluding the stop signal reaction time task due to irresolvable data quality issues in the sixth measurement wave, which prevented meaningful analysis of inhibitory control processes, and (2) analysing the Behavioural Inhibition System (BIS) and Behavioural Activation System (BAS) as separate subscales rather than a combined total score, correcting for a psychometric error in the preregistration. Both deviations were made on theoretical and methodological grounds prior to conducting the analyses, with the aim of ensuring the validity and interpretability of our findings*

References

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. *Computational Psychiatry (Cambridge, Mass.)*, *1*, 24–57. https://doi.org/10.1162/CPSY_a_00002

Ahn, W.-Y., Krawitz, A., Kim, W., Busemeyer, J. R., & Brown, J. W. (2013). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *Decision*, *1*(S), 8–23. https://doi.org/10.1037/2325-9965.1.S.8

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Gabry, J., Češnovar, R., & Johnson, A. (2023). *cmdstanr: R interface to CmdStan* [Computer software]. https://mc-stan.org/cmdstanr/, https://discourse.mc-stan.org

Heffer, T., Lundale, C., Wylie, B. E., & Willoughby, T. (2021). Investigating sensitivity to threat with the Behavioral Inhibition Scale (BIS) among children, adolescents and university students: The role of negatively-phrased questions. *Personality and Individual Differences*, *170*, 110416. https://doi.org/10.1016/j.paid.2020.110416

Jorm, A. F., Christensen, H., Henderson, A. S., Jacomb, P. A., Korten, A. E., & Rodgers, B. (1998). Using the BIS/BAS scales to measure behavioural inhibition and behavioural activation: Factor structure, validity and norms in a large community sample. *Personality and Individual Differences*, *26*(1), 49–58. https://doi.org/10.1016/S0191-8869(98)00143-3

Kelley, N. J., Kramer, A. M., Young, K. S., Echiverri-Cohen, A. M., Chat, I. K.-Y., Bookheimer, S. Y., Nusslock, R., Craske, M. G., & Zinbarg, R. E. (2019). Evidence for a general factor of behavioral activation system sensitivity. *Journal of Research in Personality*, *79*, 30–39. https://doi.org/10.1016/j.jrp.2019.01.002

Maack, D. J., & Ebesutani, C. (2018). A re-examination of the BIS/BAS scales: Evidence for BIS and BAS as unidimensional scales. *International Journal of*

Methods in Psychiatric Research, *27*(2), e1612. https://doi.org/10.1002/mpr.1612

Piray, P., Dezfouli, A., Heskes, T., Frank, M. J., & Daw, N. D. (2019). Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. *PLOS Computational Biology*, *15*(6), e1007043. https://doi.org/10.1371/journal.pcbi.1007043

Quilty, L. C., & Oakman, J. M. (2004). The assessment of behavioural activation—the relationship between impulsivity and behavioural activation. *Personality and Individual Differences*, *37*(2), 429–442. https://doi.org/10.1016/j.paid.2003.09.014

Rubinstein, A. (2003). “Economics and Psychology”? The Case of Hyperbolic Discounting. *International Economic Review*, *44*(4), 1207–1216. https://doi.org/10.1111/1468-2354.t01-1-00106

Story, G. W., Moutoussis, M., & Dolan, R. J. (2016). A Computational Analysis of Aberrant Delay Discounting in Psychiatric Disorders. *Frontiers in Psychology*, *6*, 1948. https://doi.org/10.3389/fpsyg.2015.01948

van der Pol, M., & Cairns, J. (2002). A comparison of the discounted utility model and hyperbolic discounting models in the case of social and private intertemporal preferences for health. *Journal of Economic Behavior & Organization*, *49*(1), 79–96. https://doi.org/10.1016/S0167-2681(02)00059-8

Weinsztok, S., Brassard, S., Balodis, I., Martin, L. E., & Amlung, M. (2021). Delay Discounting in Established and Proposed Behavioral Addictions: A Systematic Review and Meta-Analysis. *Frontiers in Behavioral Neuroscience*, *15*. https://doi.org/10.3389/fnbeh.2021.786358

Willroth, E. C., & Atherton, O. E. (2024). Best Laid Plans: A Guide to Reporting Preregistration Deviations. *Advances in Methods and Practices in Psychological Science*, *7*(1), 25152459231213802. https://doi.org/10.1177/25152459231213802