ASYMMETRIES BETWEEN GAINS AND LOSSES IN MOOD AND DECISION-MAKING

Rachel Louise Bedder

Max Planck UCL Centre for Computational Psychiatry and Ageing Research
Wellcome Centre for Human Neuroimaging

March 2021

A dissertation submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy at University College London

i. Declaration

I, Rachel Louise Bedder, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I can confirm that has been indicated in the thesis.

1st March 2021

ii. Abstract

The thesis begins by exploring a large-scale data set from the smartphone application *The Great Brain Experiment*. I leverage this sample size to show that gambling for prospective losses (but not gains) increases throughout the day. I introduce the question of how exploring asymmetries between attitudes and responses to gains and losses may provide useful insights in the field of *Computational Psychiatry*.

The next section of the thesis concerns mood and affective states, and their connections to decision-making. I introduce a novel paradigm: the *Future Prospects Task*, which allows for a comparison between how people feel about choosing between prospective gains and prospective losses, and how they feel about such prospects in the future. Computational modelling reveals that affective responses to losses are greater than responses to gains, demonstrating an *affective negativity bias*. It also demonstrates that the valence of future prospects has an impact on affective state, and that risky decision-making increases with proximity to positive futures, and conversely decreases in proximity to negative futures.

This novel paradigm was adapted for a new smartphone application *The Happiness Project* and for fMRI. Some of the early pilot results for the smartphone application are presented, and their feasibility for future longitudinal testing discussed. The fMRI paradigm and hypotheses are described in the discussion chapter, as data collection was disrupted due to COVID-19.

I also endeavour in the thesis to further extend our understanding of models of affective dynamics, which have become popular in the last decade. I include analyses of robustness, and highlight the statistical issues that should be taken into account with their usage.

iii. Impact Statement

The final year of this thesis work took place during the COVID-19 global pandemic. During this time, as a society we have become acutely aware of the importance of mental wellbeing, and how our mood can affect our activities (and vice versa). The science of mood has made strong steps in the last decade, due particularly to the inclusion of computational modelling and smartphone-based tools, both of which utilise dense sampling of emotional states.

This thesis explores asymmetries between positive and negatively valenced events, and how they affect our mood and behaviour. I have developed a novel paradigm, which allows for comparison between positive and negative events in both the present and near future. The findings contribute to our understanding of how negative events are prioritised in attention and learning compared to positive events– by showing stronger emotional responses for negative events than positive. I have also demonstrated that when we have the opportunity to be optimistic about our future, we increase the risks we take in the present, whilst pessimistic futures decrease the risks we take. This paradigm has been adapted for use in both a smartphone application and fMRI. This allows the findings to be explored on a longitudinal and neural level. The conclusions drawn from this work will contribute to new experimental efforts to understand the mechanisms through which optimistic and pessimistic attitudes towards our future may affect our day-to-day mood and decisions.

Some of the effect sizes presented in this work are small. However, effects of this size have implications when considering behaviour at the level of society. In a large-scale smartphone based cognitive task, I show that decision-making for prospective losses increases throughout the day. Regarding how changes in decision-making are caused by changes in future prospects; this has implications for how future events are communicated when behavioural compliance on the level of society is required. For example, the promising development of a vaccine during a global pandemic may lead to reduced compliance with current rules by means of increased risk taking. The findings from both these studies are being prepared for submission to peer-reviewed journals.

Over the course of my PhD I also developed the smartphone application *'The Happiness Project'* with my laboratory cohort. This smartphone application is designed to collect longitudinal data from the general public about their mood fluctuations over time, to be used in conjunction with clinical studies. Smartphone based tools engage the public in research, and provide an insight into the work done within academic settings in a fun and engaging way. Throughout my PhD, I have been an advocate for developing accessible ways for the public to understand Computational Psychiatry. One of these projects was designing and facilitating a significant public engagement project – *Dear World Project* to this end.

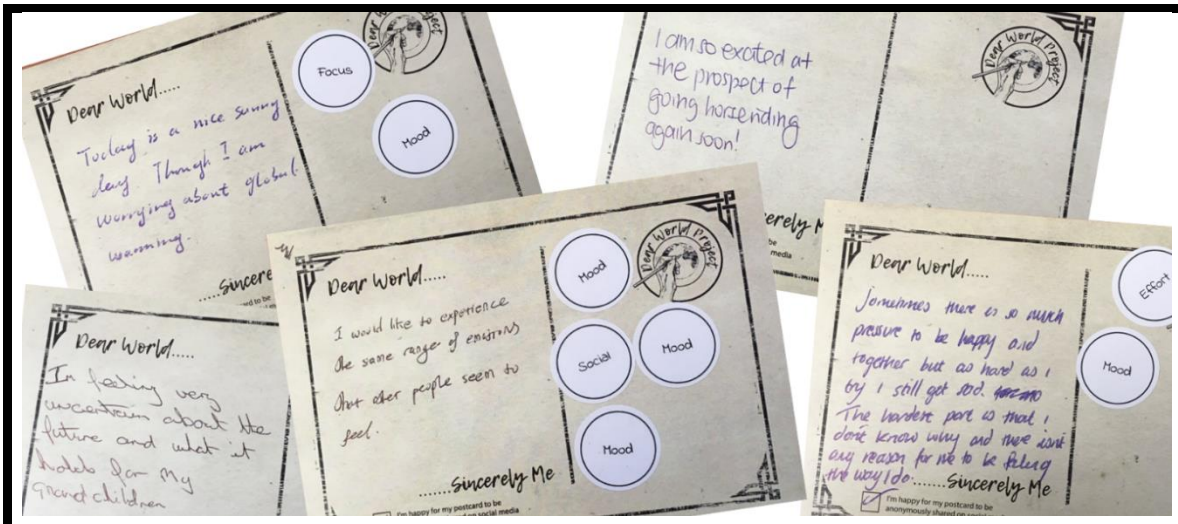


Figure 1-1 Postcards from Dear World Project

Dear World Project was a cross-disciplinary public engagement project developed by myself, Cassie Hugill and a cohort of students and staff at the Max Planck UCL Centre & Wellcome Centre for Human Neuroimaging from 2017-2019. We invited the public to write postcards about their current thoughts and feelings. They then explored postcards written by other members of the public, and had to 'categorise' them based on different labels that corresponding to different research themes (e.g. mood, confidence, effort, focus). We used this as a springboard to discuss how, as researchers, we look to understand how people think and feel, using categorical labels and continuous scales to discuss themes in Computational Psychiatry. This figure contains a selection of postcards where members of the public expressed their concerns and excitement about their future prospects.

iv. Acknowledgements

The work in this thesis was made possible thanks to the following people…

My PhD supervisors **Robb Rutledge** and **Raymond Dolan**, thank you for choosing me for the amazing opportunity, and providing the intellectual environment to learn and flourish.

The two wonderful postdocs I have worked closely with, **Bastien Blain** & **Matilde Vaghi.** I could not have done it without you both!

My Master's thesis supervisor **Dan Bush** for teaching me how use MATLAB with unimaginable patience. I don't think I would be where I am if you hadn't taken a chance on me!

The brilliant masters students I had the privilege of supervising, **Millie Lowther** and **Calum Guinea**. This has been one of the most enjoyable parts of my PhD, and I am so excited to see the amazing things you are going on to do. Thank you both for training me as a supervisor! **Suzanne Buss**, **Amie Okuma** & **Millie Lawrence** who were wonderful undergraduates to work with during their placements in the Rutledge Lab.

Those who have provided support in the form of encouragement. These include (but not limited to!) **Akshay Nair, Rani Moran, Daniel McNamee, Jochen Michely, Quentin Huys, Steve Fleming, Mehdi Keramati, Andrea Reiter, Peter Zeidman, Camilla Nord** and **Rebecca Lawson**. Some of you have repeatedly encouraged my self-belief along the way, others have provided just the right comment at just the right time that was squirreled away in my brain for darker moments. Both of these are invaluable.

The amazing cohort of PhD students I had the absolute pleasure of working with: **Magda Dubois, Alex Hopkins, Alisa Loosen, Yunzhe Liu, Jolanda Malamud, Matthew Nour, Yuki Shimura, Max Rollwage, Ilinca Ureche-Angelescu** and our **Berlin counterparts in the IMPRS**. Extra special thanks to my desk buddies **Benjamin Chew** and **Sam Ereira** for being two literal pillars of knowledge and ostensible serenity. Thank you for tolerating my hundreds of questions and perpetual outrage at trivia annoyances.

My upgrade examiners **Jon Roiser** and **James Bisby** – for providing insightful feedback and for being the first to read my attempt at writing everything down.

Those who have indulged me with long-winded conversations about what on earth mood is, how to talk about it and how to measure it. These include **Paul Sharp, Liam Mason, Daniel Bennet & the rest of the Niv lab,** and **Argyris Stringaris & lab.** I still don't know, but I feel closer!

Xiaosi Gu & **Rick Adams** who graciously let me speak at the Computational Psychiatry London Summer School and to **Vincenzo Fiore** who came along for the ride too. Also to **Peter Dayan** who supported my work on this.

Jennifer Agustus & **Sue Watson**, from the Institute of Neurology Mentoring scheme and ADAPT to Grow respectively. Jen, thank you for being the cheerleader I didn't know I needed. Sue, thank you for the invaluable advice that I still use every day in my professional and personal life.

Cassandra Hugill and the rest of the public engagement team at the FIL. I have enjoyed working on our public engagement projects so much! And of course the artist I had the great pleasure of working with, **Alban Low**.

The admin, IT & facilities team at the MPC & FIL, especially **Toyah Perkins** & **Susan Fisher**.

The imaging team, especially **Megan Creasey,** for supporting my fMRI study, and the scramble to collect as much data as possible before the scanner shut down!

To all those who practically supported me during the final year of finishing a PhD during a global pandemic and in various states of lockdown. These include **Daniel Brown** and **Greg Seago-Curl** of '***Did You Knowing Gaming?'*** for generously providing me with office space to work from. For **Jolanda Malamud, Andrea Reiter, Lorenz Deserno, Juri Deserno** for providing accommodation during this time. For the ***Silent Working Zoom Group*** – what a gift! Thank you all for making this group an amazing reality.

My friends **Frances Kennedy** and **Hannah Moase**. We come from the same world, so when you say you are proud of me it means more than anything.

My parents **Hilary Bedder** and **Stephen Bedder**, for teaching me the only things that matter are, doing something that makes you happy, working hard at it, and treating others well. These

sentiments guide my choices when it matters. Equally important, thank you to **Hilary Bedder** for years of proof reading.

Finally to **Alex Hopkins** and **Evan Russek**, for uncompromising support of every kind and variety. If this thesis is dedicated to anyone, it is to you.

v. Chapter Acknowledgements

1. Concepts, Definitions & Models

Feedback on content from Calum Guinea & Jolanda Malamud.

2. Introduction

Feedback on content from Juliana Sporrer.

3. Risk Taking for Potential Losses, but not Gains, Increases with Time of Day

This chapter is based on an article submitted for publication, in which the writing was contributed to by Matilde Vaghi, Robb Rutledge & Raymond Dolan. The *Great Brain Experiment* was developed by Neil Millstone of White Bat games. I am grateful to the help of Peter Zeidman for assistance with the use of the development backend.

4. Gain and Loss Asymmetry for Short and Long Term Prospects

This chapter is based on an article prepared for publication, where the writing was contributed to by Bastien Blain. Millie Lowther assisted with developing the paradigm and data collection. Feedback on content from Evan Russek.

5. Future Prospects of a New Smartphone Application

The smartphone application, *The Happiness Project,* was designed by the Rutledge lab and developed by *Crysberry*.

6. General Discussion

Some of the content of this chapter was given feedback by Paul Sharp and Evan Russek. The fMRI paradigm presented was developed with Bastien Blain.

vi. Table of Contents

vii. Table of Figures

viii. Table of Tables

ix. Table of Equations

1. Concepts, Definitions & Models

Here, I provide key concepts and definitions that are referred to throughout the thesis. They should be read through now if they are unfamiliar, though their use may not be completely apparent until they are situated in context. Models included in the main text expand upon the models shown here.

1.1 Mood, Emotions, Feelings & Affect

Feelings refer to the distinct conscious element of an affective experience or state. *Emotions* are categorised definitions of affective states (e.g. happiness). *Mood* is a long lasting affective state, which does not directly respond to the environment, but integrates the effects of events over time. *Affect* is a difficult property to define, but refers to the property of an event which had made it salient in a valanced way. These are all further defined and explored in Section 2.2.

Conceptually the work in thesis subscribes to the definitions as given above. Through this thesis, I endeavour to use *affective reactivity* or *affective response* when referring to a consciously experienced valenced change in state (or *affective state*). *Mood* will be used to describe the ongoing affective experience that is largely low intensity and not defined in relation to one particular event. Research, which explores mood, emotions, feelings and affect, will often be described under the heading of 'mood'.

1.2 Decisions, Value and Utility

Value and *Utility* are single scalars which can be used to compare multiple options in one dimension, and terms that are often used interchangeably within neuroeconomics literature (particularly *subjective value* and *utility*). Work which uses value to deliberate between choices considers it axiomatic that decisions are implemented in the brain by an internal neural value signal which incorporates all relative dimensions of the choice (reward, cost, risk, time delays). *Value* and *Utility* are essentially unitless, and are only understandable in comparison to each

other. *Value* is not typically measured directly, but inferred from choice, approximated by computational models.

1.3 Risk in Gain, Loss and Mixed Trial Types

All the decisions made by participants in the experimental work are two alternative forced choices, and are one of three types of trials (Figure 1-1). In gain trials, the choice is between two positive magnitudes with associated probabilities. Loss trials use only negative magnitudes. Mixed prospect trials represent a choice between a certain option of zero and a gamble with a positive and negative magnitude. In all cases where the participant does not receive the magnitude, they receive zero. Tasks that use these kinds of trials are often referred to as exploring *risky decision-making*, or as part of the field of *behavioural economics,* or *neuroeconomics.*

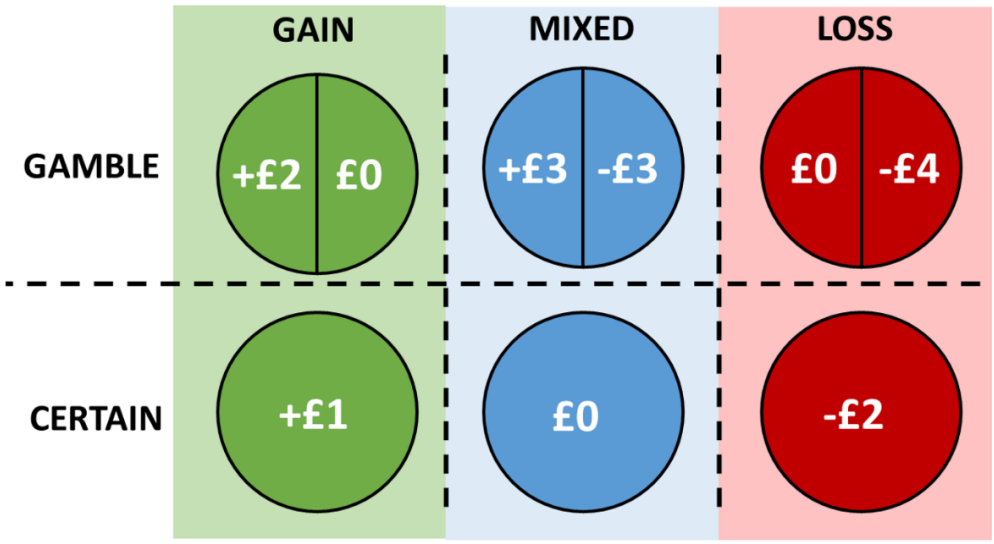


Figure 1-1 Trial Types used in Behavioural Economics Experiments

Three kinds of probabilistic decisions are typically used to measure attitudes to risk, and the relative weighting of gains vs losses in prospect theory. These are gain (green), loss (red) and mixed (blue) trials, an example of which is shown here. Each decision contains the options of a gamble (top row) where there is a 50% chance of receiving each of the amounts in the circle, or a certain option (bottom row) where there is a 100% chance of receiving the amount in the circle. In gain trials (green circles) the choice is between a certain gain of £1, or a 50% chance of winning £2. In a mixed trial (blue box) the participant chooses between a certain option of receiving nothing, or a 50% a chance of winning and a 50% chance of losing £3. In loss trials (red box) the participant chooses between a certain option of losing -£2, and a 50% chance of losing -£4.

1.4 Expected Value and Prediction Error

The *expected value* of a choice is given by the probability multiplied by the magnitude of the outcome, summed across all possible outcomes for a given choice. A *prediction error* (often called a *reward prediction* error) is the difference between the *expected value* and the outcome magnitude. Importantly *prediction errors* give a signed outcome depending on the deviation from what is expected. For example an outcome can be positive (e.g. a free vanilla ice cream), but the prediction error may be negative if a greater option was also potentially available that

was missed out on (e.g. a free double chocolate triple scoop ice cream with extra rainbow sprinkles). The expected value in this example would be the magnitude of the basic ice cream [+1] with a 50% chance of receiving it, plus the magnitude of the chocolate ice cream [+7] with a 50% chance of receiving it. An expected value of 0.5 + 3.5 is greater than an outcome of +1, thus giving a negative prediction error.

1.5 Reinforcement Learning

Reinforcement learning describes behaviour that is driven by the goal of maximizing reward. This is a vast oversimplification, but put briefly, the paradigms described in this thesis which use reinforcement learning involve participants receiving feedback on a choice they made in terms of how good or bad that action was (typically in terms of a reward or punishment; this is the *reinforcement* part). The degree to which participant updates their actions based upon a prediction error (driven by a surprising rewarding or punishing outcome) is called the *learning rate*. Learning rates can be applied to all outcomes, or separated for different conditions (e.g. separate learning rates can be defined for reward and punishments, or wins and losses). The thesis does not contain any experimental work which uses reinforcement learning models, but it is built upon inferences made from work which does.

1.6 What is a Computational Model?

Computational modelling uses three broad elements. The *generative model,* the *observed data* and the *generated data*. The *observed data* is essentially the data collected by the experimenter- this could be choices, ratings, reaction times, or even records of medical diagnoses. The *generative model* is the algorithm which we use to try and capture the variance in the *observed data*. *Rescorla Wagner, Temporal Difference Learning, Prospect Theory Models* and *Bayesian Perceptual* Models are all examples of *generative models.* The *generative model* is passed a series of task events that coincided with the collection of the *observed data,* such as the choices made by a participant, the valence of the stimuli encountered or the time limit given to make a choice. The output of the *generative model* is the *generated data (*or sometimes called *predicted data*) – a data series of the same dimensions as the *observed data.* The aim is

to minimise the mathematical difference between the *observed data* and the *generated data*. This is broadly done in two ways. Firstly by trying various *generative models* to see if they give a better 'fit'. A fit describes the mathematical distance between the *observed* and *generated data* – the smaller the distance, the 'better' the fit[1]. Secondly, each *generative model* typically includes *latent variables* or *free parameters* which can be adjusted to better fit an individual's *observed data*. Statistical tests can be performed on estimates of *free parameters*.

If this sounds novel – it is hardly the case. You can swap *observed data* for *dependent variable*, *generative model* for *general linear model*; *free parameters* for *slope* and *intercept.* Both a *generative model* and *general linear model* generate data – the key difference being the optimal parameter estimates of the *general linear model* can be calculated, whereas the parameters of the generative model are fit using an *optimisation procedure* (see Section 1.9.1 Parameter Estimation). All computational models in this thesis involve fitting *generative models* to *observed data* using the raw data on a trial-by-trial basis – no aggregates or averages are generated from the participants' data prior to the in the fitting process. The *Prospect theory Model* and the *Affective Dynamics Model* described next are both *generative models*.

1.7 Prospect Theory as a Computational Model

Prospect theory converts the objective value into subjective value based on the preferences of the agent making the choices. The choices are those which are described in Figure 1-1. The generative model consists of two parts; the equations which convert objective to subjective utility (Equation 1-1), and an equation which compares subjective utilities and transforms them into probability space (Equation 1-2).

1.7.1 From Expected Value to Subjective Utility

The following equations convert the magnitude (M_{gain} and M_{loss}) and probabilities (here 0.5 for simplicity) presented in the trial choices to a subjective utility (U_{gamble} and $U_{certain}$). The subjective utility is determined by the magnitude of the choice multiplied by the probability of

[1] I will revisit the concept of best fit shortly, but for simplicity will briefly assume the aim is to have the observed data mimic the generated data as closely as possible.

the choice and a power function. The first equation shows the utility of a choice between two magnitudes and associated probabilities. The second equation shows the utility of a certain option, where there is only one outcome. Note these equations are identical, but the utility of the certain option is presented in a simplified form. There are three parameter estimates present; α_{loss}, α_{gain}, and λ. α_{loss}, α_{gain} are often described as *risk sensitivity* and λ as *loss aversion*. λ indexes the relative weighing of gain or loss magnitudes, here is described as *loss aversion* as typically it has been shown to have a value >1, meaning losses are weighted as relatively greater (Kahneman & Tversky, 1979; Sokol-Hessner & Rutledge, 2019).

$$U_{gamble} = 0.5\left(M_{gain}\right)^{a_{gain}} - 0.5\lambda(M_{loss})^{a_{loss}}$$

$$U_{certain} = (M_{certain})^{a_{gain}} \quad \text{if } M_{certain} \geq 0$$

$$U_{certain} = -\lambda(-M_{certain})^{a_{loss}} \quad \text{if } M_{certain} < 0$$

Equation 1-1 Prospect Theory Model

Converting an objective to subjective utility using a power function (i.e. α^{Loss} and α^{Gain}) means that when the risk sensitivity parameter is <1, the subjective utility increases at a lower rate as the objective value increases. Opposing when the parameter is >1, subjective utility increases at a faster rate than the objective value. For example, if you take two magnitudes of [certain option, gamble option; +5, +10], they have a difference of +5. If risk sensitivity is set at 0.8, the subjective utilities are [+3.6 +6.3] and the difference in utility is +2.7. If risk sensitivity is set at 1.2, this gives subjective utilities of [+6.9 +15.8] with a difference of +8.9. When risk sensitivity is described as increasing, what this means is the difference in subjective utilities is increasing. While this in itself does not change the ordering of the utilities in terms of their attractiveness, when this is applied to risky decisions used in experimental tasks, the greater magnitude is typically associated with a lower probability. Thus, if the greater magnitude [+10] only has a 50% chance of receipt if chosen, and the lower magnitude of [+5] has a 100% chance of receipt, the subjective utilities for a risk sensitivity of 1.2 becomes [+6.9 +7.9] with a difference of +1. However for a risk sensitivity of 0.8, the utilities become [+3.6 +3.2] and the difference is now

a negative value of -0.4. Therefore a risk sensitivity of 0.8 appears as *risk aversive* behaviour, as the certain option becomes more attractive as the utility is greater. Conversely, for a risk sensitivity of 1.2, behaviour appears *risk seeking* as the risky option appears more attractive. The behaviour depends upon the sign of the magnitude, thus for loss trials with a greater value α^{Loss} the behaviour become *risk aversive* and a lower value becomes *risk seeking* as the greater magnitude is more aversive.

After subjective utilities are determined they are required to be converted into choice probabilities using a softmax rule (Equation 1-2). The softmax includes an inverse temperature parameter (μ) which quantifies choice stochasticity (Figure 1-2).

$$P_{gamble} = \frac{1}{1 + \mathrm{e}^{-\mu(U_{gamble} - U_{certain})}}$$

Equation 1-2 Softmax Rule

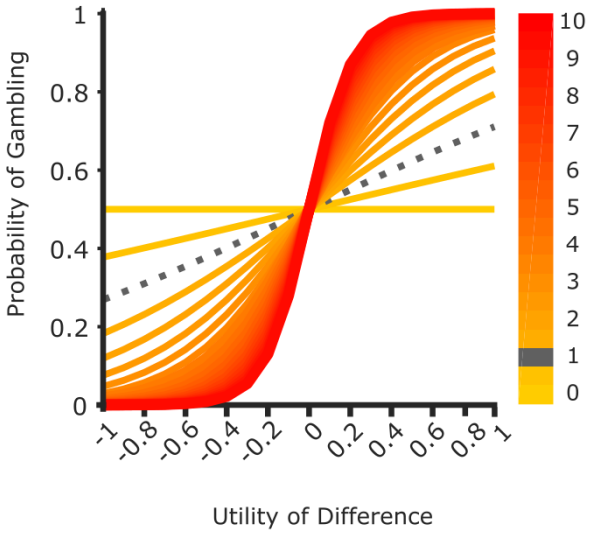


Figure 1-2 Softmax Rule Simulation

The softmax rule converts the difference in value for each of the two options (utility of the difference) to a probability of choosing one of the options (values closer to 1 predict choosing the gamble option, and values closer to 0 predict choosing the safe option). The free parameter μ dictates the degree to which the utility of the difference effects the probability of choice. The colour bar on the right indicates values of μ. When μ is high, the behaviour is more deterministically based on the difference in subjective value of the choices. When μ is low, the subjective value has less influence over the choice as the subjective probability reaches 0.5 for all choices and behaviour becomes stochastic and more difficult to predict. Importantly changes in μ do not allow the choices to be inverted.

1.8 A Computational Model of Affective Dynamics

Affective ratings that are recorded throughout an experiment can be fit using a model often branded as the the *Happiness Model* as the data is recorded in response to asking participants being asked "*how happy are you at this moment?"* or a similar question (Rutledge et al., 2014). Here *happiness*, each time it is rated by a participant, is determined by free parameters including a baseline constant C (free parameter) and *affective dynamics* parameters, which are parametric weights on different parts of the design matrix depending on the choice made by the participant. One example of an affective dynamic free parameter is a weighting on the prediction error of the choices outcome, shown below as wPE. This parameter describes the

degree to which reward prediction errors affect happiness ratings. Parameter τ captures how many trials in the past influence the current rating with an exponential decay. (Figure 1-3). An example of this model, is shown in Equation 1-3. In this example both *prediction errors* and the *expected value* of the chosen option have parametric weighted influence on the happiness ratings. The model has a highly mutable form, where any number of task events can be weighted as alternatives (or in addition) including outcomes, or separating prediction errors into different trial types (e.g. gain and loss).

$$Happiness_t = C + w_{EV} \sum_{j=1}^{t} \tau^{t-j}\ EV_j + w_{PE} \sum_{j=1}^{t} \tau^{t-j}\ PE_j$$

Equation 1-3 Affective Dynamics Model

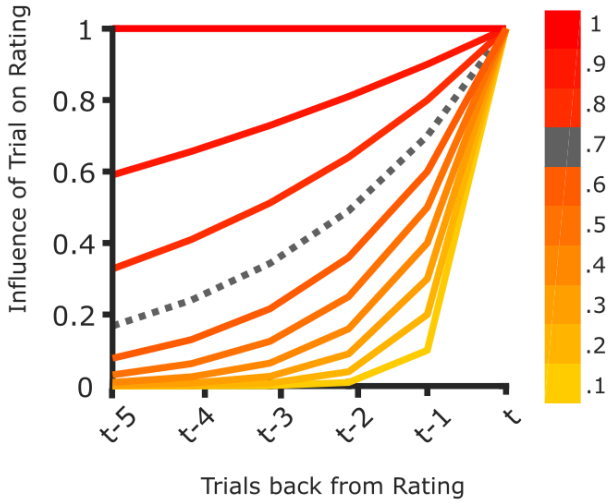


Figure 1-3 Simulation of Affective Decay

The Affective Dynamics Model captures the influence of the proceeding trials on the current rating by using a parameter which captures the decaying influence of trials further into the past. The colour bar on the right indicates values of τ. Higher values of τ indicate that trials preceding the happiness rating have more of an influence on the current rating than o. Lower values of τ indicate the preceding trials have less of an influence on the ratings. For example for τ = 0.8 trial *t-2* would have 64% of the influence as trial *t*. For τ = 0.5 trial *t-1* would have 25% of the influence of trial *t.*

1.9 Model Fitting & Comparison

Fitting models to data and comparing between them can be driven by two (partially overlapping) goals. Firstly, models can be analytical tools for analysing the observed data, similar to a statistical test of parameters in a general linear model. Secondly, they can be expressions of hypotheses about cognition that causes behaviour, or other observed data (Palminteri et al., 2017).

1.9.1 Parameter Estimation

The generative models within this thesis are fitted using two different techniques depending upon whether the observed data is choices or ratings. Choice data (i.e. binary decisions) was fit with *negative log likelihood* (nLL; Equation 1-5a) and continuous data (e.g. happiness ratings) was fit with *sum of squared errors* (SSE; Equation 1-4a). nLL and SSE both act as quantities that are smallest when the observed data and generated data are more similar, these values are called *the objective function*. The *objective function* is minimised by iterating through a series of free parameter estimates, until we find the minimum objective function we are able to sample using this process. The iteration is performed through a function in MATLAB called *fmincon* (**F**ind the **MIN**imum of **C**onstrained n**ON**linear multivariable function). Once fmincon establishes it has found the minimal objective function, it terminates the process and the model has been *fit* to the data.

Continuous ratings were fit by minimising the SSE. SSE describes the amount of unexplained variance in the observed data by the generative model (Equation 1-4a). When this is divided by the total variance in the data (the residual squared difference between the data and its mean) it gives a value which can be used as a proportion of how much variance in the data the model fits (r^2). An r^2 of 0.35 would mean the model can explain 35% of the total variance in the data (Equation 1-4b).

a. Estimating Sum of Squares Errors

$$SSE = \sum_{i=1}^{i=n} (Y - M\theta)^2$$

b. Calculating R^2

$$r^2 = 1 - \frac{SSE}{\sum_{i=1}^{i=n} \left(Y - \hat{Y}\right)^2}$$

Equation 1-4 Calculating a Measure of Fit for Continuous Data

Binary choice data were fit by minimising the nLL. Log likelihood is the likelihood of the data given the model under its current parameter estimates. A high value of log likelihood indicates a better fit, thus we use the negative log likelihood as a function to minimise (Equation 1-5a). From the log likelihood we can calculate a *pseudo* r^2 which can be interpreted similarly to an r^2 as the proportion of variance in the observed data accounted for by the model (Equation 1-5b).

a. Estimating Log Likelihood

$$Log\ Likelihood = \log(p(d \mid M\theta)$$

b. Calculating Pseudo r^2

$$Pseudo\ r^2 = 1 + \frac{Log\ Likelihood}{\log(0.5) * \text{No of trials}}$$

Equation 1-5 Calculating a Measure of Fit for Binary Decisions

1.9.2 Comparison of Model Fits

The models fit to both kinds of data were compared using *Bayesian Information Criterion* (Equation 1-6). The first term in the equation uses the *objective function*. In the second term the model guards against overfitting by penalising for the addition of each parameter. N is the number of data points being fit, K is the number of parameters in the model. BIC penalises for the addition of each parameter as a guard against overfitting the data by adding increasing

complexity. When two models have an equivalent likelihood (and are fit to the same number of data points) BIC will prefer the model with fewer parameters.

a. Choice Data

$$BIC = -2lnL(\theta|y, M) + KlnN$$

b. Continuous Data

$$BIC = N\ ln\left(\frac{SSE}{N}\right) + KlnN$$

Equation 1-6 Bayesian Information Criterion

The BIC of a model can return extremely large values where interpretability is non-trivial. One way to easily compare BIC scores is to calculate the difference between the BIC for the model with the lowest BIC score and the BIC of other candidate models. Here the difference in values can be interpreted according to a particular criterion where Model A has a lower BIC than Model B (Table 1-1). It is worth noting that in the vast majority of circumstances including an additional parameter in the model will increase the amount of variance captured. To guard against capturing small amounts of variance which can be deemed as noise in the data, the BIC model comparison trades off between generalisability (*how well this model may be able to explain other data sets*) and complexity (*the number of parameters*)[2]. The model can be described as *overfitting* when fitting a model to a particular data set is prioritised over fitting the ability to fit to new hypothetical data sets collected in the same manner.

This method of model comparison tells us how well the generated data fits the observed data. However there are other metrics of fit that can be factored in when comparing models. Briefly, these include the following (non-independent) criteria:

[2] It is interesting to note complexity does not refer to the structure of the model. A model that uses addition is no less complex than a model that utilizes multiplication, power functions and multiple transformations – as long as the number of free parameters is the same.

Identifiability: *can you refit the model to the generated data and have it prefer the same model (also called model recoverability)?*

Recoverability: *can you refit the model to the generated data and have it return the same parameter estimation values (also called parameter recovery)?*

Falsifiability: *can we define a feature of the data which would distinguish between models?*

These principles of model comparison have received considerably less attention than formal comparisons of quality of fit like BIC. This is likely to be because there is not a formalised criterion for assessing them.

Table 1-1 Bayesian Information Criterion

ΔBIC = < 2	Little difference to distinguish Models A and B.
ΔBIC = 4 -7	Considerable less support for Model B over Model A
ΔBIC = 8 - 9	Moderately less support for Model B over Model A.
ΔBIC > 10	Essentially no support for Model B.

ΔBIC is the difference in BIC values for two models. This can be for one participant, or an average taken across all participants.

1.10 A Primer on Computational Psychiatry

The work in this thesis was conceptualised and developed in an environment which self-defines itself as *Computational Psychiatry.* The field of *Computational Psychiatry* applies computational mechanisms to clinically orientated problems with the goal of understanding more about psychiatric disorders (Rutledge & Adams, 2017). The word *computational* is used differentially in two alternative sub-approaches; theory driven and data driven (Huys et al., 2016). Data-driven approaches capitalise on computational power to analyse large psychiatrically relevant data sets in order to predict clinical outcomes or find clinically relevant clusters of symptoms or behaviours. An example of this is work by Adam Chekroud and colleagues who used clinical trial data of participants' self-reported depressive symptoms (N = 4,4041) to train a machine

learning algorithm to predict clinical remission based on the success of different SSRIs (Chekroud et al., 2016). Theory-driven approaches typically involve describing experimental hypotheses as computational models and constructing experimental tasks which attempt to verify those models and falsify competing hypotheses (the proceeding sections describe this approach). One such example is a study by Robb Rutledge and colleagues where the *Affective Dynamics Model* (Equation 1-3) was fitted to participants' self-reported happiness when taking part in a probabilistic gambling task. They found that changes in happiness ratings and neural activation in response to task prediction errors did not differ in patients with major depressive disorder (MDD) compared to healthy controls, but did differ according to the baseline around which their happiness fluctuated. Those with higher depressive scores had lower baselines (Rutledge et al., 2017). This was interpreted as suggestive of prediction error signals in the dopaminergic system in people with depression may be intact, which is contrary to studies exploring prediction errors in learning tasks (e.g. Kumar et al., 2008). These two approaches complement each other: theory driven approaches are useful for trying to infer mechanisms in behavioural and neural data; data-driven approaches can capitalise on known mechanisms to explore large data sets in a more directed way (Browning et al., 2020).

Happiness is a specific. Misery is a generalisation. People usually know why they are happy. They very rarely know they [why] they are miserable.

Jeanette Winterson, *Written on the Body* (2013)

2. Introduction

2.1 Summary

This section of the thesis introduces the literature that describes how gain and loss have been thought about within the fields of mood and decision-making. It begins by addressing how we will talk about mood, its proposed function and how to engage with it experimentally. It summarises the relevant literature in computational modelling and experience sampling methods. It also expands upon the concept of *reward sensitivity* as it is addressed in behavioural studies and is informed by functional magnetic resonance imaging (fMRI). This is expanded upon by considering each stage of the decision process and how affective experience may influence, and be influenced. There is an emphasis upon how beliefs about future prospects (i.e. optimism and pessimism) can be understood as anticipated utility. Finally, the introduction focusses upon asymmetries in gain and loss and considers open questions in this area.

2.2 Defining & Measuring Mood, Emotions and Feeling

2.2.1 Definitions

The three terms used in the literature to describe different facets and dimensions of the affective experience are: *'mood', 'emotions', and 'feelings'*. Whilst we may use these words interchangeably in our colloquial language (i.e. the same event can elicit the following verbal responses: emotion: "*I am really sad about it."*; feeling: "*That made me feel awful.";* mood: "*that has put me in a really bad mood all day*"), scientific accounts tend to define these terms as distinct and with moderate convergence. The term *'emotions',* heralding from the Latin '*movere'* meaning to move, is typically associated with the categories we use to understand and

communicate our affective state: *happiness, sadness, contentment, rage, kaukokaipuu*[3]. In the past, emotions have been considered to be made up of a discrete universal set (Ekman, 1992), with more contemporary views describing emotions as being clusters of bodily sensations and environmental contexts (Barrett, 2016). The term *'feeling'* refers more loosely to the phenomenal experience of being in an affective state and thus acts as a conscious dimension of an *emotion*. The etymology of the term *'feeling'* is connected with the conscious physical sensation of a tactile stimulus, demonstrating an implicit assumption that *'feeling'* is a term which necessarily prescribes embodiment. '*Mood'* is often described as a lasting affective state with a time course that can last hours, days, or even weeks. The origin of the term '*mood'* is less clear but may relate to Old Saxon and Norse terms for 'mind' and Germanic and Gothic terms for *courage*. *'Affect'* itself defines the property of some experience or motivation that has some element that connects to our mood, emotions, and feelings. For example, an *affective state* describes an agent's position where their behaviour may be changed in some particularly way by their emotions or feelings.

The literature has begun to converge on how these three affective terms are used in relation to one another (Bennett et al., 2020; Clark et al., 2018; Marinier & Laird, 2009). Emotions act as the immediate result of some affective event. Feelings are what the agent consciously perceives, and mood operates as a summary of past affective experience. Some theories marginally differ in how they define emotions vs. feelings, or are less specific. However, an important distinction is made between emotions/feelings and mood in terms of time scales and their relationship to individual events. While emotions/feelings are responses to discrete events, mood has a *non-intentional property*: it does not relate to any particular subject (Beedie et al., 2005) and is not necessarily causally linked to the present environment (Bennett et al., 2020). For example, this distinction is made in the Free Energy Principal theory of mood, which describes emotional states as reflecting the brain's precision on prior beliefs (about the consequence of actions) and mood acting as hyper-priors that constrain the range of emotional responses (Clark et al., 2018).

[3] A Finnish word, meaning to be homesick or long for a place you have never been.

2.2.2 What is the Function of Mood & Affective Reactivity?

The etymological root of *emotion* – to move – is consistent with what it is generally agreed affective properties are for; to motivate behaviour that is beneficial to the agent experiencing the affect. However, the exact mechanism which implements this, or the exact behaviour it is designed to motivate, is of continued debate. There are two broad questions here: what is the purpose of a conscious experience of affect, and what is the purpose of an ongoing evolving affective state (i.e. mood)?

Historically science of affect has focused on emotions, their experience and expression. Emotions have been considered discrete and automatic responses, with distinct physiological characteristics and facial expressions that evolved in all cultures to improve fitness by adapting the body for reacting to the environment and communicating one's current state to others (Ekman, 1992). More contemporary views have expanded on this, describing emotions as statistically consistent clusters of bodily sensations and environmental context, which help us prioritise context-appropriate action selection (Barrett, 2016).

Understanding the function of mood is much more challenging. *What use is a diffuse, non-intentional long time scaled affective experience?* Mood is frequently described as a *latent state*, which causes changes in our behaviour and feelings but may not be directly observable. Another question of perhaps equal importance asks: *how are moment-to-moment fluctuations in affective experience integrated into mood?* Prominent computational approaches based on reinforcement learning principles have proposed mood as representing the rate of change in reward levels in an environment ('*Mood as Momentum':* Eldar et al., 2016; Eldar & Niv, 2015). This theory proposes that a high mood (positive or negative) biases the perception of outcomes in the direction of the current mood, which then updates expectations about future outcomes and encourages the agent to continue or discontinue *foraging*[4]. More recently, it has been proposed that mood integrates the current advantage[5] of performed actions in order to facilitate the process of learning the best actions in each current circumstance or state in dynamic environments (*Mood as Integrated Advantage:* Bennett et al., 2020). Importantly both

[4] *Foraging* is a term from Behavioral Ecology which describes the searching for rewards in the environment.

[5] *Advantage* is a term from Reinforcement Learning which describes how much more successful an action is compared to some other standard e.g. all other possible actions in the current circumstance.

these computational theories parametrize the integration of mood fluctuations over time using a parameter that scales the degree to which affective state at time point *t-1*, influences state at time point *t* and propose that this may vary in value in different individuals. Other models deriving from a *Resource Rationality Framework*[6]*,* consider mood/emotions to provide a meta-reasoning strategy for reducing the amount of computational strategies from which to choose (Huys & Renz, 2017; Russek et al., 2020) . All these theories emphasise the evolutionary adaptive role of mood, both in the environments in which we evolved, and perhaps also in the modern world.

2.2.3 From Day-to-Day, to Trial –to-Trial

There are three broad ways mood is quantified and explored in the literature: Clinical report (Self or Observer), *Experience* Sampling Methods, and *Experimental* Sampling Methods. The clinical report comprises a variety of questionnaires that can be completed by the participant, or by a training observer. These typically return one score per participant (or more if the questionnaire can be divided into validated sub scores). Many studies use more than one questionnaire depending on the psychological constructs of interest. Measures that comprise questions about day-to-day mood include *Beck's Depression Inventory II,* which asks for a rating of how often the subject has felt sad (amongst other questions, Beck et al., 1996)*,* and the *Mood and Anxiety Symptom Questionnaire* that asks how cheerful the responder has been feeling (amongst other questions, Watson et al., 1995). Both questionnaires ask behaviour and feelings in the last two weeks. Studies focusing on changes in symptoms, such as after an experimental intervention (e.g. Browning et al., 2012), typically collect these questionnaires in more than one instance to examine any potential change in symptoms.

While such questionnaires provide an implicitly agreed-upon standard[7] between researchers across studies, they can only be repeatedly deployed on an infrequent basis and thus cannot tell us much about how mood evolves over shorter timescales. To this end, experience sampling methods such as *Ecological Momentary Assessment* or *Experience Sampling Methods*

[6] *Resource Rationality Framework* argues that the cognitive constraints of humans need to be given more consideration in algorithm based descriptions of behaviour.

[7] Recently proposed by the National Institute of Mental Health (US) to be an *explicit* standard.

(EMA/ESM, EMA for brevity) can be used to capture the time series of an individual's mood when they are not necessarily primed for such queries (as they may be within the lab). Original methods for this include using a digital wristwatch with a timer to 'beep' at variable intervals throughout the day, and reports made in a collated daily booklet (e.g. Wichers et al., 2009). With the increased ubiquity of personal smartphones and reduced cost of accessing the internet, methods that require frequent and brief sampling of current affective states on smartphone devices have become possible, increasing the popularity of EMA as a tool of clinical interest. Study participants can now download smartphone applications to their personal devices, which prompt them at a set schedule to fill in a brief set of questions about their current affective state and/or activities in which they are partaking. With a continued repeated set of samples (each sample prompted by a push *notification*[8] on the device) for each participant, a time series can be produced of how affective states evolve, and their relationship to day-to-day activities can be probed. Whilst the work in this thesis does not include EMA, the principles upon which it is based have motivated the development of the app *The Happiness Project* (Chapter 5), and therefore I will briefly explore some of its interesting results.

Many EMA studies have used metrics of Positive Affect (PA) and Negative Affect (NA), typically generated using a dimensionality reduction method on EMA items with positive or negative valence (Figure 2-1). In many cases PA and NA have been shown to have differential predictive power. For example, an increase in PA (but no change in NA) in the first week of a new drug treatment was shown to be associated with a reduction in depressive scores after six weeks of treatment (Geschwind et al., 2011). Although requiring frequent interactions, EMA methods can be considered a relatively low burden for participants (compared with repeated visits to a laboratory setting or questionnaires gathered by phone calls). This may explain the high study retention rates that are often observed (e.g. Tsanas et al., 2016) and make it possible to continue studying patients for considerable periods of time, allowing data to be collected about individuals who have remittent episodes (e.g. Wichers et al., 2012).

[8] *Push Notifications* are messages that pop up on the home screen of a mobile device, such as the short message to alert you that you have a text message. You do not need to have the application open in order to receive its push notifications

Figure 2-1 Example of Experience Sampling Method Question

Mood Zoom as featured in Tsanas et al (2016), where participants can quickly report their current mood on various dimensions. Reproduced from Tsanas et al (2016) Journal of Affective Disorders from under Creative Commons License.

Further to exploring PA/NA at various time points, their evolution can also be capitalized on by more sophisticated time scale analysis. Metrics have been used to differentiate between the different conditions characterized by mood instability (Dejonckheere et al., 2018) particularly Bipolar Depression (BD) and Borderline Personality Disorder (BPD). One such example using *Mood Zoom* (Figure 2-1) analysed the variance of a time series of EMA with daily monitoring for over a year with a study adherence of over 80% (Tsanas et al., 2016). This generated four different metrics for variability (e.g. root mean squared successive difference) and compared whether these metrics of the EMA could differentiate between BPD, BD, and healthy controls (HC). They demonstrated that BPD and BD diagnoses were better differentiated by variability metrics from the EMA than weekly standardized questionnaire scores (e.g. GAD-7 and ASRM).

Many EMA studies have proposed alternative metrics of affective dynamics and heralded their differential prediction power. However, a recent meta-analysis of 15 EMA studies (N = 1,777) demonstrated that many newly proposed unique dynamic affective features do not predict variance to a more significant degree than the mean and standard deviation of PA and NA time

series (Dejonckheere et al., 2019). Dejonckheere and colleagues included seven proposed metrics including: *Emotional Inertia* (the degree to which PA/NA carry over from one moment to the other); *Emotional Instability* (the average change in intensity between two successive measurement occasions for PA/NA); *Emotional Interdependency across time* (the degree to which various PA and NA predict one another over time) and *Bipolarity* (the degree to which PA and NA are experienced independently). Very few of the seven new affective dynamics included in the 15 studies captured more variance than the mean or standard deviation of PA or NA. Only *Emotional Inertia* and *Bipolarity* captured additional variance of interest, with *Bipolarity* capturing additional variance in life satisfaction and depressive scores (but not borderline symptoms). This paper highlights that while time series analysis of mood can have good clinical predictive power, we should be sceptical of work proposing more complex dynamics which claim to represent distinct emotional properties.

EMA examines affective time scales with high granularity over multiple days, weeks, or months. However, to investigate the influence of experimental manipulation on a participant's affective state, experimentalists have used a similar repeated sampling method within the laboratory experimental setting. These studies typically involve multiple choice trials where decisions may vary within features such as risk, uncertainty, and magnitude. These choice variables provide varying quantifiable amounts that can be compared to participants' ratings of their current affective state on a trial-by-trial basis (many studies using this method are covered in later sections). For example, in work by Robb Rutledge and colleagues, participants are asked throughout the task (between a set number of trials) to respond to the question '*how happy are you at this moment?*' and asked to rate their current happiness on a continuous scale (Figure 2-2). Sampling affective state within an experiment was used in behavioural and fMRI studies, as well as in a gamified cognitive task on a smartphone platform (Rutledge et al., 2014). It is useful to compare and contrast these methods, particularly in terms of their capacity to shed light on psychiatric disorders characterised by mood, and their ability to advance our understanding of mood and affective dynamics.

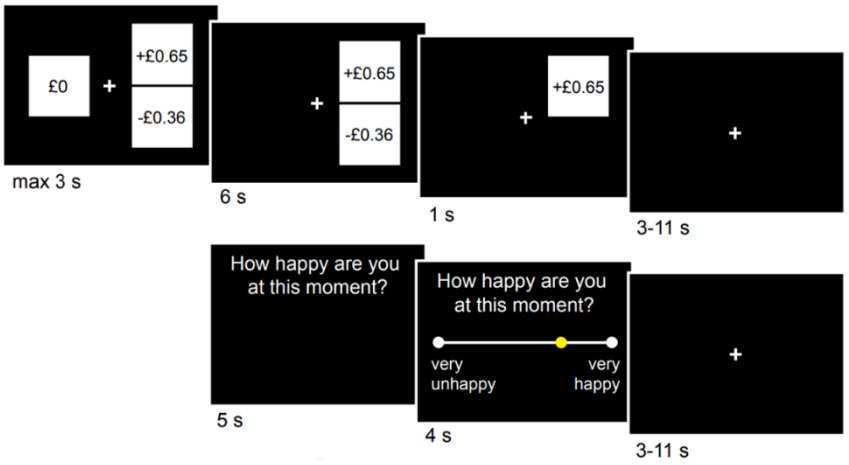


Figure 2-2 Repeated Within Experiment Sampling

Many studies utilise sampling affective state, or *happiness* repeatedly within the experiment. Shown here is an example from Rutledge et al (2014) where participants made choices in a probabilistic gambling task (example shows a mixed prospects trial), and their *happiness* was sampled after outcomes every 2-4 trials. Happiness rating was reported by adjusting a yellow cursor on a line using the computer mouse.

EMA methods allow for the assessment of reaction to particular events. For example, the degree to which positive events boost mood, or how stressors reduce mood and the transience of these effects. Daily activities, such as sleep or exercise to which researchers have a priori hypotheses, can be explored in terms of their relationship to affect. For example, one study found that how long participants slept had a differential effect on PA and NA for adolescents compared to adults (Cousins et al., 2011). Physical activity has been shown to have a larger effect on PA in patients than controls in a non-remitted group (Mata et al., 2012). Daily events can also be explored more generally. One study found that patients made more negative and less positive appraisals of day-to-day activities than healthy controls (Wichers et al., 2009). Additionally, patients were more negatively affected by negatively appraised events. However, only the magnitude of change in PA in response to positively appraised activities was a predictor of treatment success.

Humans show inaccuracies when asked to report their mood in the past (Wilson et al., 2003) or predict their mood in the future (Dunn et al., 2016). EMA methods often reduce the possibility for interference from such biases by asking people to report how they feel in the current moment (Verhagen et al., 2016). However, EMA studies have also asked participants to report their perception of recent events (e.g. last day or hour), reports which may be vulnerable to interference in terms of evaluating their valence, and therefore whether they should be included in the EMA assessment at all. For example, in Wichers et al (2009) participants were asked to appraise the current situation using Likert scales in response to questions such as *"does this activity require effort?'* or *'are you skilled at doing this activity?'* Importantly, depressive symptoms are known to affect perception of effort in an experimental setting (e.g. Cléry-Melin et al (2011); Silvia et al (2014)) and it is likely to bias these reports in a similar way. Examining an affective time series in the absence of an objective assessment of the respondent's surroundings and events arising from it may mean the data lacks valuable context of how affective responses may be sensitive to the same events in quantitatively different ways between individuals.

Another key distinction between EMA methods and lab-based mood sampling is that typically in lab-based studies only one question is asked that usually refers to positive affect (e.g. Blain & Rutledge, 2020; Keren et al., 2019; Vinckier et al., 2018). Whilst this decreases the dimensionality and thus the richness of the data, it may provide a more naturalistic assessment of someone's current emotional state; we are not frequently asked to assess our current state across multiple dimensions simultaneously. These studies implicitly assume that asking about *happiness* acts as a proxy for how much positive or negative affect is felt and therefore provides a more general assessment of current subjective wellbeing. This assumption is not unfounded. The concept of happiness occupies the sole position as the only purely positive emotion in Ekman's basic emotions theory (Ekman, 1992). Furthermore, the term *happiness* has a broadly general use beyond its concrete emotional concept (Barrett, 2016) as exemplified by its colloquial use as an evaluative term e.g. *are you happy with how that went?*.

In summation, these two methods provide different avenues for the understanding of affective dynamics. Broadly the distinction between these two methods can be defined as follows: EMA describes how individuals respond to uncontrolled natural events, and sampling within experiments describes how individuals respond to controlled unnatural events. Further work

exploring a combination of EMA and experimental tasks may find these naturally complement one another. Sampling within experiments benefits from the affective events being defined by the experimenter, allowing for interpretation of the subjective affective response to the event without confounding the subjective description of the event itself. However, by constraining the affective events, participants can respond to in such a way that the emotional salience of events is necessarily reduced as they are not relevant to a participant's nature day-to-day choices and goals. Work in Chapter 5 describes a smartphone experimental design which seeks to explore these methods in combination.

2.3 The Recursive Nature of Mood and Decision-Making

2.3.1 Reward Sensitivity

Mood and decision-making are most typically studied in the context of reward sensitivity. Rewards can be divided into primary and secondary categories. Primary rewards are those that provide direct nourishment of the body and satisfy survival needs (e.g. food). Secondary rewards are those that can be used to procure primary rewards (e.g. currency). Experimentally, primary rewards are typically understood as consumable when given and cannot be taken back by the experimenter. These include taste and smell (e.g. Amsterdam et al., 1987) , and seeing an attractive picture of a member of the gender or sex of your preference (e.g. Iigaya et al., 2016). Secondary rewards are not typically given until the experiment ends (e.g. cumulative points scored are converted to prize money; Rutledge et al (2014)). Financial loss tends to be understood under the reward framework, as it constitutes a *lack* or *reduction* of reward from a given endowment. This is distinct from punishment which involves the presence of an aversive event. These include electric shocks (e.g. Berns et al., 2006), tastes and smells (Swiecicki et al., 2009). It is worth noting that there is inconsistency in the literature as to whether losing money is described within the reward framework (as a negative reward; e.g. Maddox et al., 2012; Rutledge et al., 2015;Sokol-Hessner et al., 2015), or as part of a punishment framework (Beevers et al., 2013; Guitart-Masip et al., 2012; Pulcu & Browning, 2017). Most of the variance here may be explained by whether a reinforcement learning framework is used (punishment), or a risky decision-making (negative reward).

Our current understanding of reward sensitivity is largely driven by the activation seen in areas of the brain during neuroimaging studies (particularly fMRI) when participants are making decisions and receiving rewards (e.g. Rutledge et al., 2015) Many neuroimaging studies from the neuroeconomic and reinforcement learning literature focus on the neural correlates of subjective value and describe areas of correlated activation as *value-encoding*. This implicitly assumes that the brain encodes for a single scalar value and can be used to compare between any multiple number of choices[9].

Areas of the brain where neural activity correlates with subjective utility include the *ventral striatum*, the *ventral medial prefrontal cortex*, and the *orbitofrontal cortex* (Levy & Glimcher, 2012). Activity in these areas show increases based on the value of gain prospects (and outcomes) and decreases for the value of loss prospects (and outcomes). In an exemplar study by Levy & Glimcher (2011), participants were deprived of food and water and were then required to make risky decisions where rewards corresponded to food and water (and money). They found that individual differences in risk preferences were correlated across reward type (i.e. those who showed a preference for a small certain financial gain over a riskier large gain also showed the same preference when the prospective gains were food). Furthermore, whilst distinct neural regions were found to encode for the subjective value of food, the subjective value of all potential rewards correlated with activation in the vmPFC, suggesting that the value of both food and money lies on a common scale (with allowance for re-scaling based on context). This is an important assumption for computational psychiatry; risk preference is consistent across domains and is described by the same brain mechanism.

Reward Sensitivity occupies an important place within the RDoC (Research Domain Criteria) framework developed by the National Institute of Mental Health in the USA (NIMH; Cuthbert & Insel, 2013). The perspective of the RDoC framework is that mental illnesses are brain disorders and thus should be addressed at brain circuit level. Following this, dysfunction in brain circuits can be identified using the tools of neuroscience (e.g. functional neuroimaging) and this will

[9] A recent controversial paper has suggested that the concept of *value* is not what the mammalian brain has evolved to represent and manipulate for the making of choices (Hayden & Niv, 2020) but rather, the concept of *action weights* (i.e. the probability of performing a particular choice in a given state/situation). The *marrian* goal is *correct behaviour* rather than *computing value.* Both perspectives agree that *value* is at least a correlate to what the brain actually uses to compute and is thus a sufficient understanding for the neuroimaging studies included in this section. Furthermore, the paper concedes that the brain may compute value on occasion, but it is not what it typically uses day to day.

yield *biosignatures* that can be used for clinical management (e.g. Clementz et al., 2015). The RDoC framework specifies six major domains which reflect current thinking about the major systems at play in cognition, motivation and behaviour: *Cognitive; Arousal/Regulatory; Sensorimotor; Positive Valence; Negative Valence; Social Processes* (Insel et al., 2010). Reward sensitivity is defined under *Positive Valence Systems* which subdivides further: *Reward Responsiveness; Reward Learning,* and *Reward Valuation* (and others). The contrasting *Negative Valence System* includes: responses to threat; reactions to the withdrawal/prevention of reward; states of deprivation from a significant motivator (e.g. shelter, companionship, status). The prospective of losing reward is included in the *Positive,* and not the *Negative, Valence System*[10].

While RDoC has a biological focus, it by no means discounts the role of behavioural tasks and self-reporting. Here, behavioural tasks aid the identification of neural circuits relevant to symptoms of interest, and can be used to test hypotheses about how they may work to do this. For example, increased negative activity with strength of punishment was seen in the *habenula* (a small structure with considerable connections to sources of both dopamine and serotonin neurons in the brain; Yang et al., 2018) has been observed in patients with MDD while completing a conditioned shock task (Lawson et al., 2017). This is in contrast to an increasing positive activity seen in healthy volunteers, and no significant activation present in response to financial wins or losses. Anhedonia scores were also shown to have a negative relationship with the volume of the habenula in both healthy volunteers and patients with MDD. Such observations allow for the advancement of mechanistic discussion about the association between depression and active avoidance.

Importantly, the goals of RDoC are to both identify pathophysiological mechanisms that are present in multiple diagnosed psychiatric disorders (e.g. anhedonia), and identify mechanisms that are unique to specific psychiatric symptoms that can be used to reflect risk of developing

[10] One may also argue that this perspective is driven by the privilege of wealth where financial loss does not destabilize an individual's status or create a state of deprivation. For example a loss of £7 to an affluent individual is disappointing but does not affect their status. A loss of £7 to a family in a low income background can mean a parent going without a meal in order for their children to eat well. Mental illness is known to be more present in lower socio-economics groups (Hudson, 2005).

these symptom profiles. This approach explicitly embraces heterogeneity in psychiatric disorders (Drysdale et al., 2017; Merikangas et al., 1994). The idea is to yield new classifications based on discoveries that cut across multiple levels of analysis (behaviour, neuroscience, genomics) that can predict a degree of treatment response (Insel et al., 2010)[11]. Currently there is no strong evidence to suggest that diagnostic categories based on DSM style diagnostics relate to consistent underlying neurobiology (Hyman, 2007).

Anhedonia is a symptom that has received considerable attention in RDoC perspectives and Computational Psychiatry. It is defined as *the loss or lack of reactivity to pleasurable stimuli* (APA, 2013) and is typically understood as an extreme dimension of reward sensitivity (*reward hyposensitivity*) characterized by low approach motivation to rewarding stimuli (Auerbach et al., 2019). The other end of the scale, *reward hypersensitivity* characterized by strong approach behaviour to rewards, is observed in patients with bipolar disorder and other manic like symptoms (Nusslock & Alloy, 2017). Anhedonia is observed in a wide range of diagnostic categories including major depressive disorder (MDD; Pizzagalli, 2014), schizophrenia (Gard et al., 2007, but for a counter perspective see Strauss & Gold, 2012) , obsessive compulsive disorder (OCD; Pushkarskaya et al., 2019), and even physical illness (Eisenberger et al., 2010). It is also associated with poor clinical outcomes (e.g. Uher et al., 2012; Vrieze et al., 2013). In one such study an anhedonia symptom dimension was shown to be the only dimension that predicted longer time before remission and fewer depression free days in a trial of adolescents with treatment resistant depression randomized to a medication switch (with or without Cognitive Behavioural Therapy in parallel; McMakin et al., 2012). Dimensions that were not predictive were: reported depressed mood; somatic symptoms; morbid thoughts and observed depression.

Although anhedonia is seen as a trans-diagnostic symptom dimension there is much debate as to how it breaks down into further dimensions itself, such as anticipated, consummatory (deficits in enjoying reward), and motivational (deficit in wanting to get rewards) anhedonia (Auerbach et al., 2019). Delineating the respective contributions of anhedonia to the anatomy

[11] Akin to computational modelling, a category is *useful* given its ability to predict outcomes and we assume it will be more *useful* if it more closely reflects reality.

of a decision by fractioning the contribution of reward sensitivity at each stage can lead to better phenotyping of possibly distinct syndromes from similar symptomology (i.e. apathy, (Husain & Roiser, 2018). The decision-making process can be divided into three stages where reward sensitivity may be examined in terms of affective responses. Firstly, there is '*decision utility*' which describes the weighing of associated outcomes prior to a decision being made (Kahneman et al., 1997). Following this is '*anticipated utility';* the experience of waiting for an outcome after the decision has been registered (Loewenstein, 1987). Finally, when an outcome is revealed there is '*experienced utility'* (Kahneman et al., 1997) which describes how we feel about the outcome. All three of these stages may be affected by an individual's current mood which may influence the perception and evaluation of a decision (Hartley & Sokol-Hessner , 2018). I will next review the literature for each of these in terms of reward sensitivity and affective responses.

2.3.2 Incidental Emotional Effects

Mood can have an incidental effect on decision-making. For example, when we are in a good mood we might be more amenable to a colleague's suggestion to go for an after work drink, whereas when we are in a bad mood we may be more tempted by an early night in our pyjamas. Experiments have utilised *mood inductions* to test how incidental mood may affect decision utility. Positive mood, as induced by happiness-inducing musical extracts (Schulreich et al., 2014), sunny weather (Bassi et al., 2013), and local sports team wins (Otto et al., 2016), have all been shown to increase propensity to take risks (however this has not been observed unilaterally, Treffers et al., 2016). It is suggested that a good mood makes us feel more optimistic about the chances of our next decision (Sharot, 2011). Other work has shown that a mood induction caused by a surprise windfall of points or a significant loss of points from a 'wheel of fortune' midway through an experiment, increase a preference for stimuli encountered after a windfall (when a self-reported mood was more positive) and an aversion for those encountered after a loss (when a self-reported mood was more negative) compared

to stimuli they encountered before the surprising wheel (Eldar & Niv, 2015, however this result was only seen in participants who had self-reported high mood instability)[12].

Mechanistic explanations for how mood inductions modulate risk taking include priming to think about the possible outcomes in the direction of the induction (e.g. create a pessimistic or optimistic outlook about the decision in hand (Isen, 1993). However not all studies have consistently found that positive mood increases risk taking, and the evidence for how negative mood affects risk taking is less clear still. For example, Yuen & Lee (2003) used a mood induction of happy, sad and neutral film clips (from 22 to 26 minutes long) and then asked participants to complete the *Choice Dilemmas Questionnaire,* where participants had to respond to their willingness to take part in real life situations. They found that sad movie clips reduced risk taking (compared to neutral and happy clips), but happy clips did not have a significant effect on risk taking from neutral clips. Some studies have found the opposite effects. Yen and Chuang (2008) found that when asked to write about emotional content, sad-related content increased risk taking (by means of reducing loss aversion) and happiness-related content decreased risk taking (by means of increased loss aversion). Here loss aversion was quantified by using choice task where participants had to choose between 'status quo' products (e.g. apartments, vacation hotels and help clubs) and new alternatives. One reason for this inconsistency may be the differences in task design and the type and duration of the mood induction paradigm. Mood inductions such as film clips and writing about emotional content may induce mood in a way that is personal and idiosyncratic. It would be useful for the field to try to isolate the mechanisms by which incidental mood may change decision-making.

fMRI work focusing on mood induction has suggested that mood inductions, specifically those induced by prediction errors, cause an increase in activity in value encoding areas which drives the perception of subsequent decisions (Abitbol et al., 2015; Eldar & Niv, 2015; Iigaya et al., 2019; Vinckier et al., 2018b). This is consistent with work showing that simple endogenous fluctuations in value encoding areas bias choices in risky decision tasks (Chew et al., 2019; Huang et al., 2014). One example showing the effect of a mood induction on neural activity and risk taking is a study by Fabien Vinckier and colleagues where quiz questions of various difficulty

[12] Only probability was varied between choices and not magnitude, thus mood induced changes in risk taking were not tested.

with biased feedback were used to create episodes of high or low correct choice rate which acted as a mood induction. The choice task consisted of whether to accept or reject a motor precision task which varied in level of difficulty and prospective gain or loss framing. Rejecting the trial meant being presented with a much easier motor precision task with a smaller prospective loss or gain. Using the data generated from their computational mood model, they found that mood after the quiz task (prior to motor choice onset) correlated with positive activity in the ventral medial prefrontal cortex (vmPFC) and negative activity in the anterior insula. They extended their mood model by including each participant's individual activation in these two regions which modulated the weights on prospective gains and losses in the choice task, for the vmPFC and anterior insula respectively. Behaviourally, this means that propensity to make risky choices for prospective gains was associated with greater mood, and reduced propensity to make risky prospective loss choices was associated with lower mood. Importantly the activity in vmPFC and anterior insula was not a reflection of each other, both explaining unique variance in the mood ratings when included in the same regression model.

2.3.3 Decision utility

Work on the contribution of emotions to decision utility can be divided into two broad ideas; being presented with a decision itself can have affective properties which drive choice, and the *affective forecasts* we make about how we will feel given a particular choice outcome may also drive choice (e.g. how happy we might be if we win a prize, how disappointed we will be if we do not, or the regret we will feel if we did not take the risk). This has been hotly debated with regard to the behavioral phenomenon of *loss aversion* where *'losses loom larger than gains'* (Kahneman & Tversky, 1979). Is loss-aversion driven by the presence of negative affect at the time of decision, or anticipated increased negative affect at losing (compared to anticipated positive affect at winning)?

Theories postulating how feelings about decisions drive choice are described within various *risk as feelings* hypotheses where affective responses to a prospective decision may differ from cognitive evaluations (Loewenstein et al., 2001). In one prominent theory, the *somatic marker hypothesis* (Bechara & Damasio, 2005) feelings and their associated physiological components (e.g. increased heart-rate, nausea in the stomach) provide valuable information about the

stimulus that has been encountered which can lead to more optimal decision-making (Bechara & Damasio, 2005). For example, when a stimulus associated with a positive outcome is encountered, such as an ice-cream, this induces a positive affective state which drives you to approach the stimuli (and ideally eat it). Conversely, if the last time you ate ice cream it made you sick, the feeling of nausea when you encounter it again may lead you to sensibly avoid it. However, it is unclear as to the degree to which the effect of emotions in decision utility may scale to simple repetitive probabilistic gambling choices as often used in a computational modelling approach.

While *risk as feeling* focuses on emotions at decision time, other theories have focused on the role of anticipated feelings. In *affective forecasting* the agent uses information about how they believe they will feel about the prospective outcomes to drive their decision-making, however, this information is not experienced as affect, rather it is cognitively evaluated (Wilson & Gilbert, 2005). This provides an avenue for many cognitive biases and individual beliefs to affect judgement as future feelings must be estimated in terms of their valence, intensity and duration. While humans (and other mammals) tend to be very good at predicting the valence of future events, we are not as good at estimating the duration and intensity of the emotional response, tending to overestimate both of these (termed the *impact bias* Gilbert et al., 1998). For example college students were found to overestimate their future happiness if they were assigned to a desirable dormitory, and underestimate their happiness if assigned to a undesirable dormitory (Dunn et al., 2016). The authors of the original *affective forcasting theory* explored this by asking participants how they would feel immediately after they won or lost a fair coin toss which would result in winning $5 or losing $3 and how they would feel 10 minutes after. After finding out the result, they rated their happiness a second time after 10 minutes of a filler task. The results demonstrated a loss aversion in the participant's affective forecasts where they predicted losing would have a greater emotional impact than winning, both immediately and 10 minutes after, where they predicted that their happiness would not have returned as closely to their baseline as it in fact did (Kermer et al., 2006). These results were also present when participants completed 44 trials and were divided into *experiencers* where they actually won or lost money or *forecasters* where they watched the computer play and predicted how they would have felt. Interestingly, it has also been suggested that the *impact bias* may work retrospectively, where Americans rated their emotional responses as more

extreme four months after the 2000 Bush vs Gore election than they rated them at the time (Wilson et al., 2003).

Whilst the original affective forecasting study provided a convincing argument for the existence of an affective loss aversion at population level, they did not correlate each individual's results with behavioural loss aversion. To this end, Caroline Charpentier and colleagues utilised a model comparison based approach to explore whether feelings at outcome, or predicted feelings at outcome, were a better explanation of behavioural loss aversion. Participants made predictions and gave ratings at outcome, and both were used as inputs into a computational model predicting choice in a separate choice task. The best fitting model was one that used their *expected feelings* as opposed to *experienced feeling,* and another model that used a feeling ambivalent value function alone, suggesting that loss aversion is driven by affective forecasts with an impact bias for losses. Interestingly neither of the studies explicitly asked participants how they felt at the point of decision-making by sampling their current *experienced* emotions. This leaves open the possibility that affective forecasts and confounded with affective properties when they are made, i.e. the experienced negative affect at the time of choice leads me to over predict the negative affect of a prospective loss. A similar argument is made by Levine et al where they critique the questions used to quantify the affective forecast, suggesting they may cause the effect itself (Levine et al., 2012, 2013). They note that when forecasts are made, participants are asked *how they will feel about a particular event* and this is compared to later when they asked about *how they currently feel*. The first question encourages focalism on the event, while the second samples the emotion in the context of a more general affective state. This also highlights the care that any experimental mood sampling must take to ensure that the inferences made are appropriate to the nature of the reports given.

One overlooked feature of this is that if negative affect associated with decision utility motivates the avoidance of certain choices, an overestimation of this effect (as suggested by Charpentier) may lead to negative events being under sampled thereby limiting opportunities to accurately update the decision utility (similar to a prediction error update in a reinforcement learning model).

2.3.4 Anticipated Utility

Anticipated utility refers to the valenced effect of waiting for an event to resolve itself with an outcome. It is not difficult to think of many emotional terms referring to the waiting period before an outcome: *dread, excitement, sanguinity, suspense, hopefulness.* Many of us experience *dread* before a dental appointment or *excitement* when planning an upcoming holiday. Interesting perspectives on anticipated utility come from work on early information preference in paradigms where participants are given the option to learn about future deterministic outcomes in advance (Berns et al., 2006; Iigaya et al., 2016, 2019). Here, it is shown that for both rewarding and punishing stimuli, participants prefer to receive knowledge about their occurrence in advance.

Work exploring information seeking for future rewards comes from two studies by Kiyohito Iigaya. In the first, they asked participants whether they would like to receive prior information about whether they will see a rewarding stimuli or not (the reward being a picture of an attractive female model) after a known delay period of between 1 and 40 seconds. The likelihood of seeing the rewarding stimuli was 50% for all trials, and receiving prior information meant a deterministic answer as to whether they would see the rewarding stimuli. They found that the longer the delay the more likely participants were to choose receiving prior information about the delayed outcome. The authors argue that new information constitutes a *positive prediction error* from receiving the information itself, starting the process of *savouring* the *anticipation* of the future reward (Iigaya et al., 2016). In a follow up study using a similar paradigm in fMRI, they replicated that participants would prefer to know in advance if they would see a rewarding stimuli (this time a more general pleasant image), and that their preference increased when the delay was longer. Interestingly, they found preferences for information were heterogeneous whether the probability of the rewarding stimuli was higher or lower. They proposed a choice model which includes an advanced information prediction error (aRPE) and an anticipatory signal which ramps up in strength from the onset of the aRPE until the rewarding image is seen. They found the ramping up anticipatory utility signal from the model correlated with an increase in activity in the vmPFC throughout the delay period, and as expected, aRPE strength correlated with activity in the dopaminergic midbrain (VTA/SN). RPE however are phasic signals; interestingly an aRPE signal was also present in the hippocampus along with prolonged activity throughout the delay period. This hippocampal activity was found

to mediate functional coupling between the vmPFC and dopaminergic midbrain – the authors suggest a role for the hippocampus in *imagining* the future reward and thus playing a key role in anticipation.

Iigaya and colleagues speculate whether they would be able to generalize their conclusions to events of a negative valence and feelings of *dread.* The rewards they used did not have a ranking of how rewarding they are, which meant that they could not test whether anticipation had any value that could be offset against the value of the final reward (i.e. would they prefer early knowledge of a possible smaller reward, or no early knowledge of a probabilistic larger reward?). These ideas have been explored in work featuring punishing stimuli (Berns et al., 2006). Berns and colleagues found that when given the choice some participants preferred to receive slightly larger shocks with a shorter delay over smaller shocks with a longer delay. Participants with this preference (referred to as *extreme dreaders*) rated the shocks after a longer waiting period as more painful than the rest of the cohort – demonstrating that the aversive nature of waiting is on par with the pain of a shock. During fMRI, the authors observed similar temporal dynamics to Iigaya's work in areas of the brain known to respond to pain (i.e. *right posterior insula* and *caudate ACC*). Where activity from the cue predicting the punishment to the time of the painful shock showed a similar ramping effect as in the vmPFC for anticipated rewarding stimuli. Those characterized as *extreme dreaders* showed earlier and more sustained activity than the *mild dreaders*.

These studies in concert suggest that people have a tendency to prefer prior information about future rewards and punishment. I will revisit this observation when we consider the mechanisms behind optimism and pessimism in Section 2.3.6 (*Optimism and Pessimism*). It is worth noting that work on anticipation has focused on the future receipt of primary rewards. There is currently no work concerning anticipation of future reward prediction errors in secondary reward (i.e. you will get chances to win money soon!).

2.3.5 Experienced Utility

Work on how decisions affect our mood has largely been driven by our understanding of experienced utility; how we feel about receiving an outcome. Dominant in this area is the idea that mood is largely driven by surprises, or *prediction errors*; deviations between what we

expected and what we received. In one frequently cited study, real world unexpected events (quantified from historical trends) such as sunny days or a local sports team win boosted happiness at city wide level and increased lottery ticket purchases. Similarly in an EMA study of university students with ratings taken after receiving of exam grades, deviations from expected grades drove ratings of PA and NA (Villano et al., 2020). Here PA/NA, were densely sampled every 45 minutes over the eight hours following grade receipt. A computational model which used both the grades and student's expectations about them (after taking the exam and before receiving the grades) better predicted the time course of PA and NA than simply the grades themselves. Another interesting example is a study looking at the introduction of new technologies in the classroom (e.g. tablets), which found lower expectations about the technology intervention resulted in higher approval ratings after their implementation (Hassan & Geys, 2016).

Laboratory work has demonstrated that even transient emotional fluctuations are driven by deviations from expectations on a trial-by-trial time scale. In 2014 Rutledge and colleagues sampled *happiness* throughout a probabilistic gambling task after the outcome of the decision was received every 2-3 trials (Figure 2-2). A model comparison approach revealed that happiness ratings were better explained by weights for the influence of expected value and prediction errors (Equation 1-3 *Affective Dynamics Model*), compared to outcomes alone, a result which was consistent across behavioural and fMRI, and could be replicated multiple times in a large scale smartphone data set (Rutledge et al., 2014). Another laboratory study used the same method of asking about happiness every 2-3 trials but looked at how it varied across three differently structured environments (Keren et al., 2019). They used a probabilistic gambling task where trials of gain, loss and mixed prospects are pseudorandomised (as did Rutledge et al., 2014), and a *structured* task the trials were organised into blocks of very high or very low prediction error values. They also included a *structured adaptive* task where the reward prediction errors for a subsequent trial were calculated using a *closed-loop control*[13] in order to deliver participants the trials they needed to maximally boost their happiness ratings as close to the top or the bottom of the scale as possible. In this experiment, the also found that happiness ratings were driven by expectations and prediction errors, but defined these

[13] Closed Loop is a control system where the operations are regulated through feedback.

differently. The expected value was defined as the average of all previous outcomes, and prediction errors were the difference in the outcome and the expected value.

The studies presented here all find a role for expectations and predictions in describing the happiness ratings after an event. While each of these studies presents a subtly different model to draw inferences, this may reflect differences in the experimental paradigm, highlighting that caution that should be applied when generalizing these results to *what makes us happy*. For example, in the study of university student's exam grades by Villiano and colleagues, they found the best fitting model included outcomes in addition to expectations and prediction errors. This may be due to the personal salience of the event and its relevance to future rewards; getting an A in an exam is a positive event whether you expected it or not, as it is a good indicator of future exam success and even good job prospects[14]. Laboratory based studies lack personal salience in their mood inducing events and include multiple similar events in sequence so outcomes may play less of a role. Differences in how the expectations and prediction errors are defined between Rutledge at al and Keren et al may be due to the way environments are structured. In Keren's study, as environmental changes occurs in the *structured* and *structured adaptive* tasks, causing there to be a greater role for previous events on current affective ratings. These models also differ in their conclusions about whether recent events or primary events (i.e. those at the beginning of sessions) have a greater influence on mood. While these suggestions are speculative, it may be that in the interpretation of any proposed model of mood, we need to consider how the model may mirror the paradigm in a way that makes it less generalizable to day-to-day fluctuations.

2.3.6 Optimism and Pessimism

I have briefly dissected the contributions of affective responses throughout the decision process. These studies focus largely on events that are happening to us right now. Specifically, studies exploring anticipation and dread limit themselves to events happening after a delay of less than a minute. Thus the field has yet to ask whether optimistic and pessimistic beliefs about the future may in themselves have affective qualities which may impact upon decision-making.

[14] There is no guarantee it will be linked to future happiness. It could also lead to doing a PhD.

We will briefly review the literature on optimism and pessimism with regards to mood, behaviour and major depressive order (MDD) in order to assess which models of affective anticipation (and dread) may be extended to explore further future prospects.

A multitude of studies have shown that overall we have an *optimism bias*; a tendency to believe events in the future will be more positive than negative, and update our beliefs when presented with more information in a positive direction (for reviews, Sharot, 2011, 2012). For example, in a community wide sample it was found that respondents to a mailed questionnaire had a tendency to rate themselves as less at risk than their peers for susceptibility to various kinds of physical and mental harm (Weinstein, 1987), [15]. In another example, heavy smokers were shown to consider themselves less likely to be at risk of premature mortality (Schoenbaum, 1997). Despite these examples showing how optimistic beliefs about risk may encourage less desirable behaviour, the association of optimism with positive benefits means it is often considered a desirable bias to have. Observational studies have shown that trait optimism is associated with better mental and physical health outcomes (Conversano et al., 2010; Scheier & Carver, 1987). A possible suggestion for this is that holding optimistic beliefs promotes behaviours that increase the potential for rewarding outcomes by increasing motivation and self-confidence (Bouchard et al., 2018).

Another perspective is that good subjective wellbeing requires accurate perception about self and the environment (e.g. Dillard et al., 2009). One paper suggests that both optimistic and pessimistic beliefs lead to reduced long term wellbeing (de Meza & Dawson, 2020). Utilizing a survey of British households across 18 years (N = 1,601), they found that those who predicted themselves to be financially better or worse off both experienced significant reductions in their long term subjective well-being compared to those who held realistic beliefs about their future financial situations. This presents a contrast to the infamous concept of *depressive realism* where depressive symptoms are associated with an accurate perception of future prospects and how well things are going (Alloy & Abramson, 1979). However, it is worth noting two

[15] However it is worth noting that the methodology that supports the *optimism bias* in particular *optimistic belief updating* has been criticized (Shah et al., 2016), highlighting the methodological complexities of using analogue scales for reporting and assuming linearity throughout the scale.

reviews that conclude that *depressive realism* may be a small effect which is very dependent upon study methodology (particularly against the objective standard of reality that is used to compare, Ackermann & DeRubeis, 1991; Moore & Fresco, 2012) and most work focuses upon events that will be resolved quickly within the experiment (i.e. not in the future). It may be that not all situations benefit from an optimistic disposition, particularly those with such an importance to real life environments as financial situations (Dawson et al., 2015). Biases about the valence of our future are likely to strongly interact with current environmental statistics.

When considering real life events further into the future, MDD has been associated with a more pessimistic perspective of the future. In an online study, participants were asked to estimate the probability of events of different valence occurring in their lives (e.g. having a serious headache, being yelled at by a stranger, having your work praised). Those with higher (subclinical) depressive scores predicted positive events to be less likely than those with lower scores compared to their actual incidence over a 30 day period (Strunk et al., 2006). It is not clear whether trait pessimism is a risk factor or a symptom of MDD.

2.4 Loss and Gain Asymmetry

In the preceding review of work on the affective properties of decision-making, it is clear that understanding both responses to positive and negative stimuli in both rewarding and punishing environments needs to be explored. One ongoing implicit debate in the field is whether responses to gain and loss may be best described as a mirror of each other. For example, one infamous behavioural principle from *prospect* theory is the *reflection effect* (Daniel Kahneman & Tversky, 1979). The *reflection effect* refers to the reversal of preferences when shifting between prospect choices in the reward domain as gains or losses, which can be captured by a single free parameter (e.g. Arnett, 1992; Ruggeri et al., 2020; This is the α parameter discussed in Section 1.7 *Prospect Theory as a Computational Model*). In a similar vein, many studies in reinforcement learning capture positive and negative prediction errors in the same parameter (e.g. Behrens et al., 2007; Gershman et al., 2009). Some studies have employed separate learning rates for positive and negative prediction errors (Gershman, 2015; Niv et al., 2012). In both cases, studies looking across lifespans have found differential development patterns for

behaviours towards gains and loss (Christakou et al., 2013; W. van den Bos & Hertwig, 2017), motivating dual process biological accounts of behaviour (Bossaerts & Murawski, 2015).

The term *affective gradient* has been used to describe individual differences in preferential processing of positive vs negative stimuli (Pulcu & Browning, 2017). Identifying whether *affective gradients* are driven by a single or separate processes for positive and negative stimuli is an important question for computational psychiatry, particularly in terms of treatment approaches. For example, if preferential processing is governed by one operation only one side of the spectrum need be approached (i.e. increasing attention to positive things will decrease attention to negative things). Whereas if these are partially independent processes, one valence may be a better treatment target (i.e. we can focus on increasing attention to positive things OR decreasing attention to negative things). Overall it seems people hold a *negativity bias*, where negative information seems preferentially processed in terms of our attention and behaviour (Baumeister et al., 2001; Rozin & Royzman, 2001). This has been observed in learning, where learning from negative prediction errors exceeds learning from positive prediction errors (Gershman, 2015; Niv et al., 2012). Negativity biases have also been observed in other modalities, for example losses show increase physiological arousal in terms of pupil diameter and heart rate compared to equivalent gains (Hochman & Yechiam, 2011). To this author's knowledge, there is no work in the affect domain where negativity biases are observed in affective responses driven by prediction errors.

The roles of the neurotransmitters dopamine and serotonin have been implicated in possible opponent roles in decision-making in terms of their contribution to positive and negative stimuli (Cools et al., 2011). Briefly, classic work by Schultz, Dayan and Montague showed in midbrain single cell recordings from dopaminergic neurons, a spike in activity for unexpected reward, a depression in activity for the absence of an expected reward, and no activity change for an expected reward (Schultz et al., 1997). Since this, the role of phasic dopamine has been well established as a reward prediction error signal (Glimcher, 2011) and associated with increased approach bias for rewarding stimuli (Rutledge et al., 2015). While the role of serotonin has been implicated in depression (Cowen & Browning, 2015) its direct contribution to behaviour has been less clearly delineated. The considerably larger number of serotonin receptors (compared to dopamine) suggest that its influence and role in psychiatric disorders is highly complex (Naughton et al., 2000). Specifically, serotonin is associated with promoting withdrawal from

aversive stimuli through behavioural inhibition (Cools et al., 2011; Crockett et al., 2012). However, isolating serotonin's contribution to punishment compared to loss is not clear, due to negative experimental outcomes not always being defined as one or the other. Recently Jochen Michely and colleagues found that during a week-long daily dosage of SSRIs (placebo controlled, in a non-depressed population), learning from losses, but not wins, was enhanced in a reinforcement learning task compared to a session prior to SSRI or placebo to the task session days afterwards (Michely et al., 2020). Here they postulate SSRIs may boost mood over time by means of increasing the amount of positive surprises available (by slowing down reward learning) and reduce the number of disappointments (by speeding up punishment learning).

Irrespective of the precise implementation of responses to positive and negative stimuli, it is largely suggested that dissociable decision processes represent an adaptive process for responding to differential environmental statistics (Dayan & Huys, 2009). Different environmental statistics can include changes in the presence of food and predators across the day, or in terms of their autocorrelation (e.g. the presence of one fruit may indicate the presence of a fruit tree, however the presence of one tiger does not suggest more tigers[16]). It has been suggested that negative affective gradients observed at population level are a rational response to the assumption that losses have a higher information content then gains. Indeed it has been shown that participants adapt their learning rates separately to gains and losses depending on the volatility of the environment (a stimuli's volatility acting as a proxy for its information content about future occurrence (Pulcu & Browning, 2017). Conversely, simulation work from Eran Eldar and colleague's *mood as momentum* model suggests that increased learning from positive events (compared to negative) increases overly positive expectations regarding rewards in one's surroundings leading to disappointment when expectations increasingly exceed the value of outcomes. Thus symmetrical learning rates are best primed for stable mood fluctuations (Eldar et al., 2016).

Reducing a negativity bias has been suggested as a target for novel behavioural interventions (Browning et al., 2010). The use of *attentional bias modificiation* has been tested in this context; whereby computerized training is used to encourage focus on positive stimuli compared to negative (by means of the positive stimuli being more reliability associated with a probe).

[16] 80-95% of carnivores hunt alone.

Browning and colleagues found in a placebo controlled study modification, training using valenced faces (but not words) lasting two weeks was associated with a reduction in BDI scores after four weeks (Browning et al., 2012). Strikingly, the scores associated with training was seen two weeks after the training had finished, and not immediately preceding the training itself, consistent with the delay in therapeutic effects often reported for antidepressant medication.

An interesting counter perspective is provided by work comparing reward and punishment domains. Here, depressed individuals have been found to perform better in tasks involving the minimization of punishment compared to their non-depressed counterparts in tasks where immediate benefits are traded off against longer term pay offs (Beevers et al., 2013; Maddox et al., 2012) with the inverse being seen in reward maximization conditions in one of the two studies (Beevers et al., 2013). The authors of both papers argue that the idea of depressive disorder being associated with decision-deficits per se may be due to the literature's focus on reward based tasks. They argue elsewhere that when a task framework matches an individual's 'regulatory style' they are more likely to make optimal decisions (Maddox & Markman, 2010). It is worth noting that the *mood as momentum* model (Eldar et al., 2016; Mason et al., 2017) all takes place within the rewards domain. A more recent study has suggested that mood related reward prediction errors are no different between healthy controls and patients with MDD, however they do not test prediction errors for wins and losses separately (Rutledge et al., 2017). These results highlights the need to explore both rewards and punishments, and consider the environmental stimuli typically present in an individual's real life in order to evaluate the utility of asymmetrical learning.

2.4.1 Negativity Bias & Optimism Bias

One intriguing unanswered question is why the literature on the whole, observes a *negativity bias* for learning about current events, and an *optimism bias* for our beliefs about future events? Why do we learn more from bad things, but believe things will turn out better? One explanation could be that optimism pertains to higher-level abstract beliefs, whereas a negativity bias relates to low-level reinforcement learning processes. Germain Lefebvre and colleagues, found greater learning rates for positive prediction errors than negative prediction errors, leading them to conclude that the *optimism bias* may, in fact, be explained by a low level learning

mechanism (Lefebvre et al., 2017). It is not clear how this result fits with the field's norm of observing a *negativity bias* (the authors suggest that differences may be seen due to whether paradigms are instrumental or pavlovian) but it invites the question of whether an individual's position on an affective gradient should pertain to both beliefs about the future and their current choices. The field has yet to study both of these measures in the same task, to understand whether those who believe their future will be positive, learn more from positive events in the present.

3. Risk Taking For Potential Losses But Not Gains Increases With Time Of Day

3.1 Abstract

Humans exhibit distinct risk preferences when facing choices involving potential gains and losses. These preferences are subject to neuromodulatory influence, particularly from dopamine and serotonin. As these neuromodulators manifest distinct circadian rhythms, this suggests decision-making under risk might be affected by time of day. Here, in a large subject sample (N = 26,720), we tested the hypothesis of a diurnal modulation in risk taking for gains and losses. We found that risky options with potential losses were increasingly chosen over the course of the day. Using a computational modelling approach to obtain a more fine-grained account, we show this diurnal change in risk preference reflects a decrease in sensitivity to increasing losses, but no change was observed in the relative impacts of gains and losses on choice. This diurnal sensitivity, present across two different task designs, was robust to between- and within-subject analysis, age, and gender. Thus, our findings reveal a striking diurnal modulation in human decision-making, a pattern with potential importance for real-life preferences that include voting, medical decision-making and global stock market investments.

3.2 Introductions

Everyday decisions are driven by risk attitudes that vary across individuals (Levin & Hart, 2003; Weber et al., 2002; Yechiam & Ert, 2011). Where one person might wait for a walk signal to turn green before crossing the road, another may take a cursory glance at the oncoming traffic and hurry on their way. Equally one person may be adventurous in trying a new ice cream flavour, while another may be more inclined to stick with an old favourite. People can exhibit opposing preferences in gain frames with prospective rewards and in loss frames with prospective losses (Tversky & Kahneman, 1981). Thus, the same person who makes a safe choice when deciding between potential rewards (i.e., ice cream) may make a risky choice in the face of potential losses (i.e. jaywalking).

The *reflection effect* refers to a reversal of preferences when shifting between gain and loss frames. This effect is a central feature within prospect theory (Daniel Kahneman & Tversky, 1979) and is repeatedly observed at the population level (Liu et al., 2014; Tymula et al., 2013). It is also robust across multiple demographics including, for the most part, a sample of 4,098 participants spanning 19 countries and 13 different languages (Ruggeri et al., 2020). This effect is also observed in non-human primates (Lakshminarayanan et al., 2011). Behaviourally the reflection effect manifests as *risk aversion* for gains (i.e., smaller certain reward are preferred over larger risky rewards) and *risk seeking* for losses (i.e., larger risky losses are preferred over smaller certain losses (Kahneman & Tversky., 1979).

Biological accounts link risk-taking behaviour to underlying neural circuitry (Bossaerts & Murawski, 2015) and to processes that change across the lifespan (reviewed in development: (Rosenbaum & Hartley, 2019) and in ageing (Samanez-Larkin & Knutson, 2015)). For example, epidemiological evidence shows that adolescents engage in greater risky behaviours (Arnett, 1992) which is often attributed to developmental changes in dopaminergic reward circuitry (Blakemore & Robbins, 2012; Fareri et al., 2008). However, the majority of such studies focus on situations in which every risky option includes a potential gain, without considering decisions that involve only potential losses. Many studies implicitly assume that risk taking for gains and losses is governed, at least partially, by shared processes (Arnett, 1992). However, natural aging is associated with a decline in the integrity of the dopamine system and this has been linked to a parallel decline in risk taking for potential gains but not potential losses (Rutledge et al., 2016). Furthermore, risk taking for losses shows a linear decrease while risk taking for gains shows a

quadratic relationship between the ages of 8 to 22 (van den Bos & Hertwig, 2017). In a broader sample between the ages of 12 to 90, risk attitudes for gains did not relate to risk attitudes for losses in the same individuals (Tymula et al., 2013). The overall pattern of findings from these studies suggests that risk taking for gains and losses might be controlled by distinct processes.

Similarly, computational modelling in studies of risky decision-making often focuses on a single loss aversion parameter. It is assumed to govern the relative impact of potential gain and losses on choice ('*losses loom larger than gains':* Charpentier et al., 2016; Kahneman & Tversky, 1979), which could influence risk taking related to both gains and losses (reviewed in Sokol-Hessner & Rutledge, 2019). However, many studies do not consider situations in which only potential losses are involved. It is an open question whether any shared process influences attitudes towards risk for both gains and losses. In line with dual-process models of decision-making, dopamine is associated more with reward-related behaviours (Rutledge et al., 2015) while serotonin is associated more with loss-related behaviours (Dayan & Huys, 2008). An environmental asymmetry in rewarding and punishing stimuli has been proposed as providing an adaptive account for dissociable decision processes that could vary independently (Dayan & Huys, 2009).

Whether circadian rhythms influence risk taking in humans is unknown. Regular changes in environmental statistics are associated with changes in behaviour in animals across the course of a day. For example, the presence of foraging opportunities and predators change on a regular diurnal cycle (Chiavacci et al., 2015). Circadian changes in risk taking are observed in male dark-eyed juncos deciding whether to sing or forage (McNamara et al., 1987), and increased risk taking during foraging is observed later in the day in sparrows (Caraco, 1981). The regularity of these diurnal changes is entrained in brain circuitry and might be explained by regular fluctuations in neuromodulators throughout the day (Fuller et al., 2006).

We tested risk attitudes for gains and losses in humans as a function of time of day. To this end, we measured propensity to take different types of risks at different time of the day in a sample of more than 25,000 people playing a gamified risky decision task on a smartphone platform (*The Great Brain Experiment,* www.thegreatbrainexperiment.com). We implemented computational models based on prospect theory (Sokol-Hessner et al., 2009; Tversky & Kahneman, 1992) to test whether risk taking for gains and losses show differential diurnal patterns that might reflect a differential mechanistic basis.

3.3 Experimental Procedures

3.3.1 Participants

We included the data from 26,720 participants from the UK and US aged 18-69 (18,215 18-39; 13,168 Female; 19,156 UK) who completed a risky decision-making task between March 1, 2013 and September 30, 2014 on the gamified cognitive task platform *The Great Brain Experiment* (freely available on Apple iOS and Google Android). Participants played the task at whatever time they liked with no prompting. Results from part of this data set have been previously published elsewhere (Rutledge et al., 2014; Rutledge et al., 2016). Both gender, age and location were defined as participants self-selecting their demographics when they downloaded the app. Participants selected their age from the given age brackets of; 18-24, 25-29, 30-39, 40-49, 50-59, 60-69 and 70+. All participants gave informed consent within the smartphone platform. The Research Ethics Committee of University College London approved this study.

3.3.2 Within-Subjects Sample

We identified a subset of participants (N = 2,599) to utilize for within-subject analyses. Eligibility for this subset was defined as having completed at least two game plays on different days. All plays included were between 8am and 10pm, time windows utilized in previous studies (Correa et al., 2020; Orban et al., 2020). All plays were required to be using the same design matrix.

3.3.3 The Great Brain Experiment

The Great Brain Experiment is freely available on Apple iOS and Google Android. Participants downloaded the app under a *Citizen Science* endeavor to contribute to neuroscientific research. There were two significant download peaks, firstly when the app was released in March 2013, and secondly when the paper contained the *Happiness Equation* (Equation 1-3 *Affective Dynamics Model*) received considerable press attention in August 2014.

3.3.4 Smartphone-Based Experiment

The task involved 30 trials where participants chose between safe and risky options. Participants also rated their happiness 12 times and the results of those analyses are reported elsewhere (Rutledge et al., 2014; Rutledge et al., 2017). The majority of participants took between 3 and 5 minutes to complete each game. Participants began with 500 points. When each game is finished the participants are told the score they achieved, what percentage of all other plays from all players it was higher than, and their all-time high score. On each trial participants chose between a safe option (where the points were guaranteed if they chose the option) or a risky option where they had a 50% chance of two potential outcomes. The risky choice was represented on a spinner where an arrow moved around until it landed on either of the two potential outcomes. Where the safe option was chosen the outcome was resolved immediately. There was no time constraint on making a decision on each trial. The task contained three types of trials: (1) gain trials, where they could choose between a certain gain and a gamble with a greater potential gain or a zero; (2) loss trials, where they chose between a certain loss and a gamble with a greater potential loss or a zero or (3) mixed trials where they chose between a certain option of zero or a gamble with a potential loss and a potential gain (Figure 3-1).

3.3.5 Design Matrices

The data included was from two different design matrices, with the second design matrix added with an update of the smartphone application to test the robustness of our findings. We ran between-subjects analyses separately for the ratio and uncorrelated designs. For the within-subjects analyses we only included participants who had completed plays that were both ratio or both uncorrelated designs.

Each design contained 11 gain, 11 loss, and 8 mixed trials in each 30 trial game. The ratio design had a set of 60 gain, 60 loss, and 30 mixed trials to choose from. The loss trial set was identical to the gain trial set except with negative values. The ratio design comprised all downloads or updates between March 6, 2013 and July 16, 2013. In this design the gain (and loss) trials had possible 4 certain amounts (30, 35, 45, 55) and 15 multipliers (1.64, 1.7, 1.76, 1.82, 1.88, 1.94, 2, 2.06, 2.12, 2.18, 2.26, 2.4, 2.7, 3.2, 4). Mixed trials had 3 prospective gains (40, 44, 75) with

10 prospective losses generated from multipliers (0.2, 0.34, 0.6, 0.54, 0.77, 0.89, 1, 1.1, 1.35, 2).

The uncorrelated design had a set of 45 gain, 45 loss, and 30 mixed trials to choose from. In the uncorrelated design (comprised of all downloads or updates after July 17, 2013) gain trials had 3 certain amounts (35, 45, 55) and 15 gamble amounts (59, 66, 72, 79, 85, 92, 98, 105, 111, 118, 124, 131, 137, 144, 150). Mixed trials has 3 prospective gains (40, 44, 75) and 10 prospective losses (-10, -19, -28, -37, -46, -54, -63, -72, -81, -90).

3.3.6 Time of day

All data from the smartphone platform was time stamped with Greenwich Mean Time (GMT). Thus 6.5 hours was subtracted from timestamps for data collected in the US to correct for the time difference for the average participant in the US.

3.3.7 Statistical Procedures

We report Pearson correlation coefficients for effect sizes of relationships between task measures (i.e., gambling for losses) and time of day of the game's completion. Additionally, we computed the difference between pairs of effect sizes (i.e., effect size of the relationship between gambling for losses and time of day, effect size of the relationship between gambling for gains and time of day) to test if time of day had different effects on each parameter. All p values were computed based on permutation tests using 10,000 random shuffles of the time of play to determine null distributions (MATLAB, Version 2018a). We also included Bayes Factor tests in support for the null hypothesis (BF_{01}) for all instances where the permutated p values were less than 0.1 (JASP, Version 0.14.1). A BF_{01} of less than 3 offers mild evidence for the null hypothesis, a BF_{01} between 3 and 10 offers moderate evidence for the null hypothesis, and values exceeding 10 offer strong evidence. We computed these analyses separately for both genders, and for older and younger players, and for US and UK participants.

3.4 Results

On each choice trial, participants were presented with choices between risky and safe options, and had to activate a spinner to make a risky choice (Figure 3-1).The outcome of chosen gambles was revealed after a brief delay. Participants received no prompt as to when they should complete the task. To detect the presence of diurnal variations in risk taking, we examined how preference for risk taking in gain, loss, and mixed trials changed according to the time of day that the task was completed. We examined plays from the two countries with the largest sample sizes in our data set (UK, N = 18,977; US, N= 7,743), involving a total of 26,720 participants. For between-subject analysis, we examined first plays alone so that all participants would contribute the same amount of data. For all analyses, we considered the day to begin at 06:00. The median time the task was completed was 18:48 with 70% of plays between 08:00 and 22:00. For within-subject analysis, we evaluated a subset of players (N = 2,599) who had completed plays on two different days between 08:00 and 22:00.

Figure 3-1 The Great Brain Experiment Task Design

Gain trials (green outline added for description purposes as every trial had the same outlook to the participants) have potential gains and no potential losses. In an example gain trial, a participant chooses between a risky option (here, 50% probability of 59 points) and a safe option (here, 100% probability of 35 points). Loss trials (red) have potential losses and no potential gains. Mixed trials (blue) have both potential gains and potential losses with a safe option that is always worth 0 points.

3.4.1 Risk Taking for Potential Losses Increases with Time of Day

We found that in gain trials there was no significant relationship between risk taking and time of day (Pearson's r = 0.0018, p = 0.77). Here, we found evidence for the null hypothesis over the alternative hypothesis with a Bayes Factor (BF_{01}) of 124.94. In contrast, in loss trials (i.e., those with no potential gains), we observed a strong positive correlation between risk taking and time of day (r = 0.037, p < 0.0001). A similar increase was also present in risk taking in mixed trials, which feature both potential gains and losses (r = 0.027, p < 0.0001) (Figure 3-2). Time of day effects for loss and mixed trials did not differ significantly from each other (p = 0.068), but effect sizes for mixed and loss trials were both significantly greater than for gain trials (both p < 0.0001). Hence, risk taking for potential losses, but not potential gains, increases with time of day, with people on average choosing more risky options with potential losses in the evening

compared to in the morning. Participants on average chose the risky option more than 50% of the time (gain trials: 70.2% ± 26%; loss trials: 53.7% ± 29.4%; mixed trials: 66.3% ± 26%). As risk taking appeared greater for gain trials than loss, we tested whether the lack of effect observed in gain trials could be due to a ceiling effect. In this analysis, we only included participants who choose the riskier option in gains, losses and mixed less than 100% and more than 0% of the time in all trial types (N = 16,322). We found risk taking and time of day was correlated for loss trials ($r = 0.028$, $p = 0.0004$), but not gain ($r = 0.0063$, $p = 0.424$) or mixed trials ($r = 0.0066$, $p = 0.400$).

This diurnal effect on risk taking was robust to gender in loss trials (female: $r = 0.040$ $p < 0.0001$; male: $r = 0.032$, $p = 0.0002$). In mixed trials the effect was statistically significant for female ($r = 0.038$, $p < 0.0001$) but not male participants ($r = 0.011$, $p = 0.20$, $BF_{01} = 40.54$). Positive effect sizes were seen for younger and older age groups for loss trials (18-39: $r = 0.039$, $p < 0.0001$; 40+: $r = 0.036$, $p = 0.0014$ and mixed trials (18-39: $r = 0.025$, $p = 0.0006$; 40+: $r = 0.037$, $p = 0.0007$). Overall, increased risk taking in loss but not gain frames was a highly consistent result present in 7 out of 8 subsamples tested (Table 3-1).

Although increased risk taking in mixed trials (most closely associated with loss aversion in the literature) was also robust to age and task design, the effect was present in females but no significant effect was found in males. We further tested this effect due to their being unequal splits between genders for the two design matrices. The first play for male participants was 77.2% the uncorrelated design (N = 10,543) and 22.9% the ratio design (N = 3,123). The first play for female participants was 64.6% the uncorrelated design (N = 8,427) and 35.5% the ratio design (N = 4,627). As the time of day effect appeared greater in the ratio design and female participants had a greater percentage of plays with the ratio design, we confirmed the differences in risk taking for mixed trials was present in each design. Where the effect of time of day in risk taking for mixed trials was seen to be greater in female and males this could be observed when looking at uncorrelated design first plays ($p = 0.017$) but not in ratio design first plays ($p = 0.244$). Thus, the time of day difference observed could not be put down to differences in task design.

We also considered other factors that may vary throughout the day. We first tested whether mood changed across the day. Participants rated their current mood by answering the question *how happy are you at this moment*? when the game began, but before they had been presented

with any trials. Participants mood did not change linearly throughout the day ($r = -0.013$, $p = 0.033$, $BF_{01} = 13.17$), however mood was higher in plays in the morning (6am – 12pm) compared to plays completed late at night (12am till 6am) ($t(7948) = 5.35$, $p < 0.0001$, Figure 3-3). Therefore, we included this as a covariate in a partial correlation test for risk taking for loss trials and time of day to test whether changes in choice behaviour could be partially explained by changes in mood. In this test we observed the same positive correlation between risk taking for losses and time of day ($r = 0.037$, $p < 0.0001$).

We next tested whether the time of day was related to the response times for each trial type (Figure 3-3). Median response times were on average faster when participants chose the safe option compared to the risky option in loss trials (safe option: 3.78 ± 1.94; risky option: 4.30 ± 2.18; $t(48{,}450) = 9.01$, $p < 0.0001$), and faster to choose the risky option than the safe option in gain trials (safe option: 3.42 ± 2.02 ; risky option: 3.010 ± 1.38; $t(47{,}507) = 2.72$, $p < 0.0066$). Participants had the same response times for when they choose the safe or risky option in mixed trials (safe option: 4.05 ± 2.24; risky option: 3.69 ± 1.73; $t(47{,}784) = 1.76$, $p = 0.078$). Thus, we tested separately effects of response times across the time of day for each trial and choice type. In each response times analysis we removed all participants with an average median response time of over 30 seconds.

Participants were faster when they choose the risky option later in the day in loss trials ($r = -0.023$, $p < 0.0001$), but not when they chose the safe option ($r = 0.0078$, $p = 0.226$, $BF_{01} = 59.25$). Response times did not change for gain or mixed trials in safe or risky options (safe option: $r = 0.0014$, $p = 0.844$, $BF_{01} = 115.08$; risky option: $r = -0.013$, $p = 0.029$, $BF_{01} = 11.93$). We used a partial correlation test to determine whether the relationship between risk taking and time of day could be partially explained by changes in response times for these trial types. When median response time for loss trials for safe or risky options were included as a covariate we still observed a positive correlation between time of day and risk taking (safe option: $r = 0.0340$, $p < 0.0001$; risky option: $r = 0.0328$, $p < 0.0001$).

Table 3-1 Effect Sizes for Time of Day on Risk Taking for each Demographic Split

Data	Sample size	Gain trials		Loss trials		Mixed trials	
		Effect size	p value BF_{01}	Effect size	p value BF_{01}	Effect size	p value BF_{01}
All	27,720	0.0018	0.77 124.94	0.037	<0.0001	0.027	<0.0001
UK	18,977	-0.00035	0.96 109.79	0.040	<0.0001	0.029	<0.0001
US	7,743	0.011	0.34 44.00	0.027	0.016 4.12	0.021	0.066 12.80
Male	13,666	-0.0038	0.65 84.38	0.032	0.0002	0.011	0.20 40.54
Female	13,054	0.0052	0.56 76.50	0.040	<0.0001	0.038	<0.0001
Younger	18,106	-0.0068	0.39 70.54	0.039	<0.0001	0.025	0.0006
Older	8,614	0.0045	0.68 67.83	0.036	0.0014	0.037	0.0007
Ratio	7,750	-0.0026	0.82 68.48	0.046	<0.0001	0.029	0.010 2.71
Uncorrelated	18,970	0.0073	0.31 66.52	0.028	<0.0001	0.022	0.0031

Effect sizes (Pearson's r) for each demographic split of the data (country, gender, age, ratio or uncorrelated task designs). Demographic pairs (where N = 27,720) are every pair of rows after 'All'. Bayes Factor (BF_{01}) tests for evidence for the null hypothesis are included in all instances where the permutated p value > 0.01. BF_{01} of > 1 indicate support for the null hypothesis.

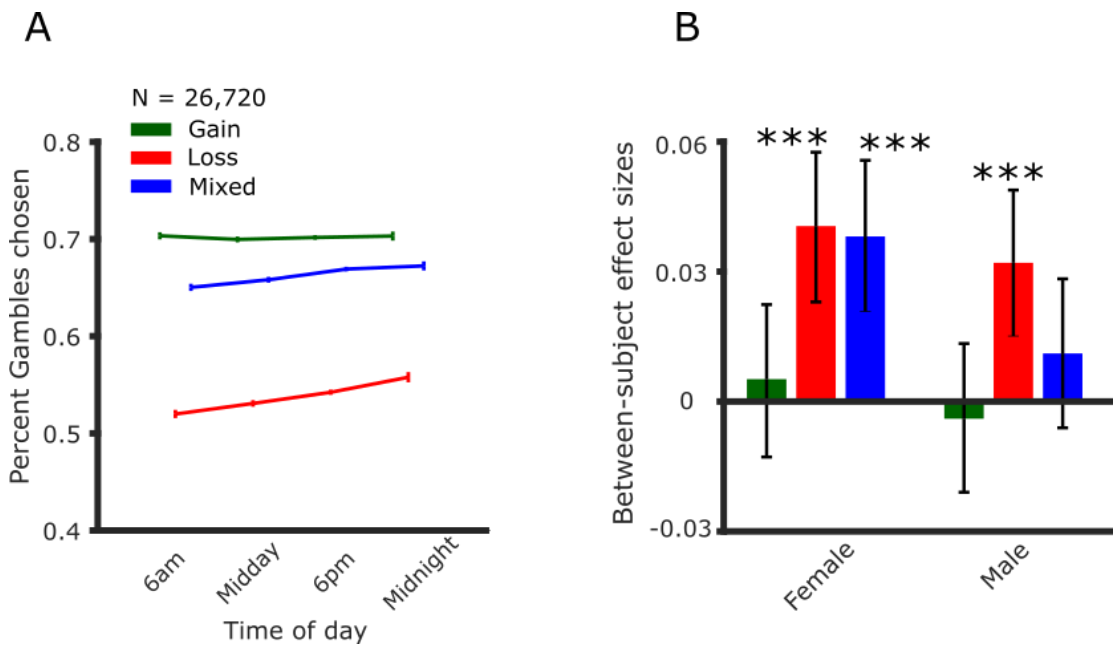


Figure 3-2 Risk Taking for Potential Losses Increases with Time of Day

A) Risk taking in choice trials with potential losses increased with time of day (N = 26,720). This was true in loss trials featuring risky options with equal probabilities of zero or a potential loss, and in mixed trials featuring risky options with equal probabilities of potential gains and losses. The frequency of choosing risky options with only potential gains was unaffected by time of day. Each time bin includes data collected in the interval starting at the time indicated (i.e., the 6am bin includes all data collected from 06:00 until 11:59). Error bars represent the SEM.

B) Risk taking in loss trials increased with time of day in both female (N = 13,054) and male participants (N = 13,666). Risk taking in mixed trials increased with time of day in females but not males. Error bars represent bootstrapped 95% confidence intervals.

[* $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$]

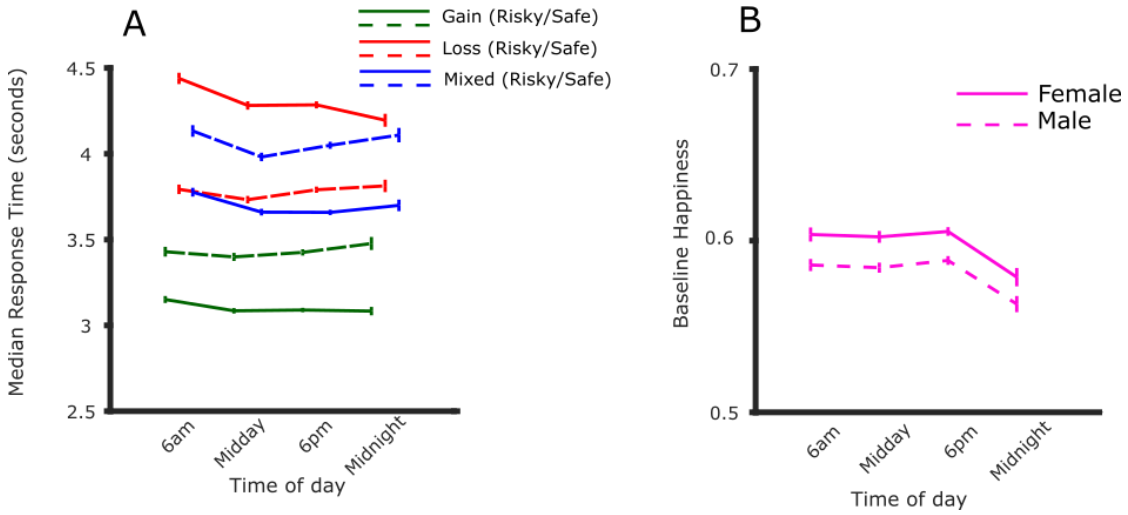


Figure 3-3 Responses Times and Mood Across the Day

A) Median response times decreased across the day for when risky options were chosen in loss trials, but not for any of trial and choice type combination. For each trial and choice type we removed participants who either did not ever choose that trial choice type combination or had a median response time of over 30 seconds. Thus, we removed the following number of participants from each analysis (gain risky: 2878, gain safe: 5113, loss risky: 420; loss safe; 5052; mixed risky: 320; mixed safe 1060). Response times for safe trials are indicated by a dashed line and risky trials by a filled line. Each time bin includes data collected in the interval starting at the time indicated (i.e., the 6am bin includes all data collected from 06:00 until 11:59). Error bars represent the SEM.

B) Baseline happiness was lower between 12am and 6am than between 6am and 12pm for both female (N = 13,054, $t(3684) = 3.74$, $p < 0.0001$) and male participants (N = 13,666, $t(4262) = 3.91$, $p < 0.0001$).

3.4.2 Computational Modelling using Prospect Theory

Choice behaviour in risky decision tasks is well described using a parametric computational model based on prospect theory (Equation 1-1 Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Standard decision models include parameters for risk sensitivity (α), loss aversion (λ), and choice stochasticity (inverse temperature, μ). This model can be extended to allow risk sensitivity to vary for gains and losses separately (α_{gain} and α_{loss}) (Sokol-Hessner et al., 2009). Adopting a model-based approach to our behavioural data allows us to delineate between alternative mechanisms which could explain the time of day effect on risk taking for losses but not gains, asking whether time of day affects a single model parameter. An increase in risk taking for trials that include losses could be explained by a decreased loss sensitivity (α_{loss}), reducing the difference in expected utility between potential losses associated with safe and risky options. One effect of a decrease in loss sensitivity is to render the safe option subjectively less attractive relative to the risky option. Loss aversion is perhaps the best known of the phenomena described in prospect theory. Increased risk taking for losses could also relate to a decrease in loss aversion (λ), a parameter that captures the relative weighting in choice of potential losses and equivalent gains.

We fit a model with a single risk sensitivity parameter (*Single Alpha Model*) and a second model with separate risk parameters for gains and losses (*Dual Alpha Model*). The *Dual Alpha Model* (pseudo-r^2 = 0.37 ± 0.23, mean ± SD) fit the data better than *Single Alpha Model* (pseudo-r^2 = 0.29 ± 0.21) and was preferred according to Bayesian model comparison which penalizes for model complexity (Table 1-1).

Table 3-2 Model Comparison for Individual Plays

Model	Parameters per subject	Mean r^2	Median r^2	Model BIC	BIC-BIC_{dual}
Single Alpha	3	0.29	0.25	1061299	1481
Dual Alpha	4	0.37	0.34	1059817	0

Bayesian Information Criterion (BIC) measures are summed for fits for each participant's first play (N = 26,720). Both models included choice stochasticity (inverse temperature) and loss aversion parameters. The final column is the difference between the model BIC and BIC for the Dual Alpha model.

3.4.3 Loss Sensitivity and not Loss Aversion Decreases with Time of Day

Using the preferred *Dual Alpha Model,* we examined each parameter in relation to the time of day the game was played. As expected, α_{gain} (1.02 ± 0.30, mean ± SD) was not associated with time of day (r = -0.0042, p = 0.50, BF_{01}= 103.49) nor was choice stochasticity μ (0.74 ± 2.76; r = -0.0071, p = 0.24, BF_{01}= 66.84). Instead, time of day was correlated with α_{loss} (0.79 ± 0.34, r = -0.039, p < 0.0001), where a decreased α_{loss} (reduced loss sensitivity) is consistent with the observed increase in risk taking in loss trials. This result was robust to gender (Figure 3-4), age, and task design (all r < -0.029, p < 0.001). If loss aversion parameters decreased with time of day, this could partially explain increased risk taking for losses. Instead we observed a modest increase in loss aversion log(λ) (0.63 ± 0.93, r = 0.013, p = 0.036). Furthermore the data supported the null hypotheses with a BF_{01} of 4.87. Thus, decreased loss aversion does not provide an explanation for the greater risk taking for potential losses we observe as the day goes on. Loss sensitivity α_{loss} decreased significantly more with time of day than loss aversion log(λ) in both females (p < 0.0001) and males (p = 0.0007).

In this novel data set, the values of α_{gain} and α_{loss} diverge from what is typically observed in the literature (Ruggeri et al., 2020), where α is typically on average estimated to be greater than 1, meaning participants are risk seeking for losses (when the expected value of the choices is equal, choosing the safe option) and risk aversive for gains (when the expected value of the choices is equal, choosing the safe option).

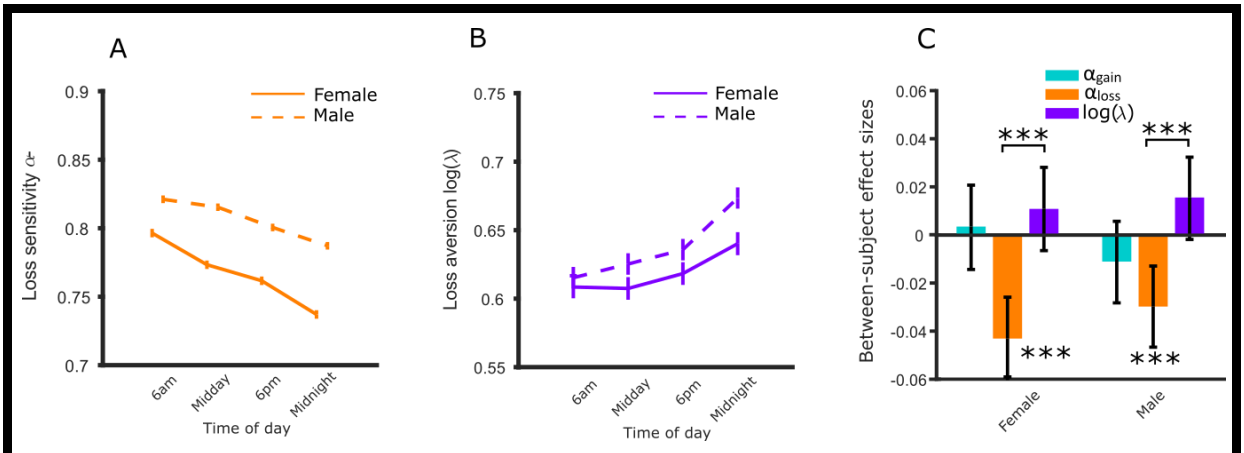


Figure 3-4 Loss Sensitivity and not Loss Aversion Decreases with Time of Day

A) Loss sensitivity parameters (α_{loss}) decreased with time of day in both females and males, consistent with increased risk taking in trials with potential losses. Error bars represent the SEM.

B) Increased risk taking for losses could be partially explained by decreased loss aversion. However, loss aversion parameters ($\log(\lambda)$) did not decrease with time of day in both samples. Error bars represent SEM.

C) Loss sensitivity decreased more than loss aversion in both females and males. Error bars represent bootstrapped 95% confidence intervals.

[* $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$]

3.4.4 Loss Sensitivity Decreases with Time of Day in Individuals

To test for the presence of circadian effects within individuals, we used the first two plays from a subset of players (N = 2,599) who completed plays on two different days between 08:00 and 22:00 (eligibility criteria in Supplementary Information). We confirmed the results of the between-subject sample model comparison within this smaller sample (**Table 3-3**). For each subject, we computed the difference in the time of day between the two plays and asked whether a larger time difference predicted a greater change in risk taking. The mean time of day difference between each play used in the analysis was 4.3 h (SD, 3.4 h). Time of day difference was positively correlated with the difference in risk taking between two plays in loss trials ($r = 0.057$, $p = 0.0038$), but not in mixed ($r = 0.015$, $p = 0.44$, $BF_{01} = 30.40$) or gain trials (r

= -0.0038, p = 0.84, BF_{01} = 39.93) (Figure 3-5a).Additionally, we again found that the difference in time of day between the two plays was negatively correlated with the loss sensitivity parameter α_{loss} (r = -0.061, p = 0.0024) but not with the loss aversion log(λ) (r = 0.0075, p = 0.70, BF_{01} = 37.79) or gain sensitivity parameters α_{gain} (r = -0.011, p = 0.59, BF_{01} = 34.94) (Figure 3-5b & c). We find that the strength of the decline in loss sensitivity over the course of the day was significantly greater than the decline in loss aversion (p = 0.015).

We also tested whether mood or response times changed within individuals with the time of day. The time of day difference was not correlated with the difference in mood ratings (r = -0.0158, p = 0.413, BF_{01} = 29.41). Regarding response times, there was no change in gain or mixed trials for either choice type (all r < 0.02, p > 0.404, BF_{01} > 23). No significant differences were seen for response times in loss trials, but bayes factor tests revealed only mild support for the null hypothesis in both cases (safe: N = 1,815, r = -0.043, p = 0.0649, BF_{01} = 6.32 ; risky: N = 2,488 r = -0.047, p = 0.019, BF_{01} = 2.36).

Table 3-3 Model comparison for Within Subject Data set

Model	Parameters per subject	Mean r^2	Median r^2	Model BIC	BIC-BIC_{dual}
Single Alpha	3	0.31	0.27	201608	1705
Dual Alpha	4	0.40	0.36	199903	0

BIC measures are summed across the within-subject dataset (N = 2,599). Both models included choice stochasticity (inverse temperature) and loss aversion parameters. The final column is the difference between the model BIC and BIC for the Dual Alpha model.

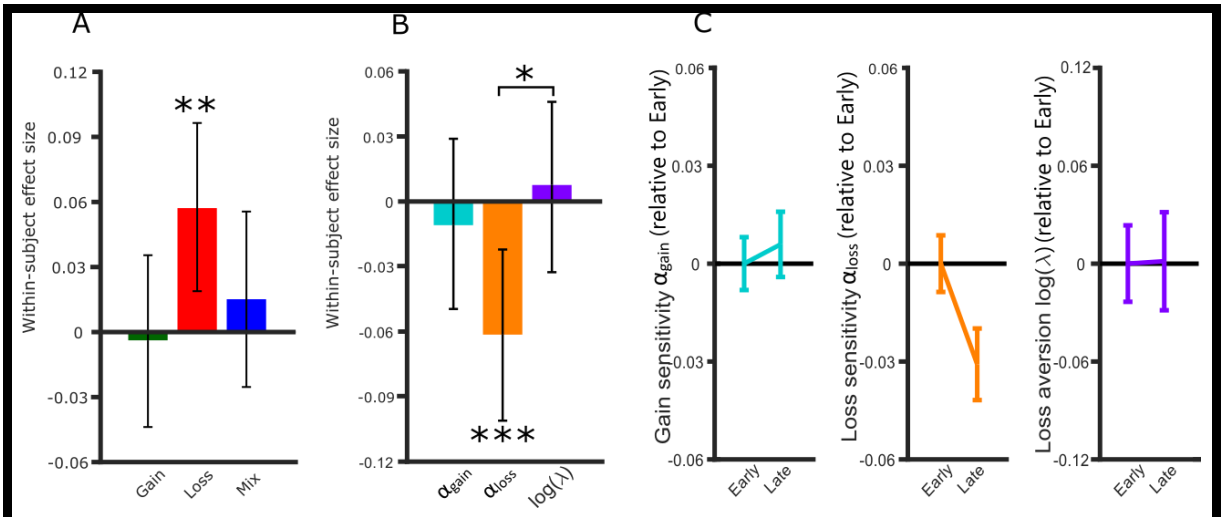


Figure 3-5 Loss Sensitivity Decreases with Time of Day in Individuals

(A) We identified individuals (N = 2,599) who completed the risky decision task on two different days between 08:00 and 22:00. Risk taking in loss trials increased with time of day within individuals. **(B)** Loss sensitivity (α_{loss}), but not gain sensitivity (α_{gain}) or loss aversion (log(λ)), decreased with time of day in this within-subject sample. Error bars represent bootstrapped 95% confidence intervals **(C)** Loss sensitivity decreased significantly with time of day in within-subject analysis. Error bars represent bootstrapped 95% confidence intervals. Data binned on differences less (early) or more (late) than 7 hours are shown for illustration purposes.

[* p < .05, ** p < .01, *** p < .005].

3.5 Discussion

We show that time of day is associated with changes in attitudes towards risk, informing a theoretic account of human behaviour. Time of day, reflecting circadian rhythms, may influence risk attitudes with respect to potential losses but with no impact on risk taking observed for reward. Using computational modelling, we show that changes in risk taking for losses were explained by a reduced sensitivity to increasing losses over the day. The results are consistent with an account for risk taking with distinct processes related to risk taking for gains and losses.

In terms of computational processes, a model with separate value sensitivity parameters for potential gains and losses explained the data better than a model with a single value sensitivity parameter. This was true in both between- and within-subject datasets, consistent with a time

of day effect on one value sensitivity parameter but not the other. One possible mechanism for diminishing loss (but not gain) sensitivity could be related to potential increases in tonic serotonin over the course of the day. Elevated serotonin concentrations late in the day have been reported in the dorsal raphe nucleus (DRN) the primary site in the brain of serotonin neurons (Ågren et al., 1986), and for blood plasma levels in both rodents (Mateos et al., 2009) and humans (Rao et al., 1994). The DRN has reciprocal connections with central neural structures for circadian rhythms like the Superchiasmatic Nucleus, which plays a central role in regulating circadian rhythms (Ciarleglio et al., 2011; Mazuski et al., 2018). However, the role of serotonin in sensitivity to valenced stimuli is unclear. Selective Serotonin Reuptake Inhibitors (SSRIs) are associated with reduced attention to negatively valenced stimuli (Harmer et al., 2004). Studies exploring the propensity to take risky choices with losses have associated taking tryptophan (the precursor of serotonin) with reduced risk taking for losses and a reduced reflection effect (Murphy et al., 2008). However, no change in the reflection effect has been associated with tryptophan depletion (Rogers et al., 2003). In the computational modelling literature serotonin is frequently associated with pavlovian avoidance (Cools et al., 2011; Crockett et al., 2012; Dayan & Huys, 2008).

The phasic response to potential aversive outcomes by serotonin neurons could be highest relative to a background of low tonic serotonin earlier in the day. If phasic serotonin is related to the aversiveness of options under consideration, a reduced dynamic range for serotonin in the evening when tonic serotonin is high could provide a potential mechanistic explanation for our finding of reduced sensitivity to increasing losses. A similar mechanism has been proposed for explaining increased reward seeking when pre-stimulus dopaminergic midbrain BOLD activity is low and phasic responses to potential rewards are elevated (Chew et al., 2019). An established link between phasic dopamine and reward seeking (Rutledge et al., 2015; Stopper et al., 2014) combined with the lack of effect in the present study despite the large sample suggest that phasic dopamine responses to reward-related stimuli do not change throughout the day.

An alternative account is that making decisions about losses increases task attention which may interact with changes in wakefulness through the day (Yechiam & Hochman, 2013). Previous diurnal patterns of behaviour have been observed in the literature when taking into account a participant's chronotype (e.g. Ingram et al., 2016; May, 1999). Ultimatum Game decisions

varied with time of day and did so differently in participants who identify as being a 'Morning Type' compared to those who identify as being an 'Evening Type' (Correa et al., 2020). Such differences are attributable to changes in wakefulness and reduced inhibitory executive function (Correa et al., 2020; McElroy & Dickinson, 2019). Here, we saw no change in choice stochasticity for simple value-based choice despite the large sample size. However, when we analysed response times for each trial type and choice type, differences were only seen in loss trials, but not gain trials. Future work might usefully examine how chronotypes relate to a modulation of diurnal changes in loss sensitivity. Finally, it is of interest that profound changes are reported in resting-state brain network connectivity profiles over the course of the day (Orban et al., 2020). Integrating the latter approach in conjunction with measures of decision-making could help to identify the specific networks that underlie diurnal changes in loss sensitivity.

We observed an association between time of day and risk taking in mixed trials in females but not males. Sex differences in risk taking have been observed in a variety of effects (e.g., under stress (Lighthall et al., 2009; Preston et al., 2007), sensitivity to winning and losing (R. van den Bos et al., 2013) and real life risky behaviours such as drug use and dangerous driving (Byrnes et al., 1999)) but little is known about how these may interact with the circadian rhythms of neuromodulators. Speculatively, one possible reason is human sex hormones interacting with different phases of circadian rhythms. Estrogen has been associated with shortened circadian periods in rodents (i.e. more free running in daylight hours, Krizo & Mintz, 2015) and has been suggested to shorten the circadian period in humans (Leibenluft, 1993). Increased estrogen during ovulation has also been associated with reduced loss aversion in mixed trials in humans (Lazzaro et al., 2016).

Our results demonstrate that diurnal patterns in risk behaviour can be remotely assessed with smartphone. This also provides a non-invasive method to observe circadian rhythm disruption, a key factor in many mental health disorders (Spencer et al., 2013; Teicher et al., 1997). The causal direction of circadian rhythm disruption and psychiatric disorders is unknown (Parekh & McClung, 2016). Further work exploring the interaction of hormones and circadian cycles may also shed light on the increased prevalence of seasonal affective disorder in woman which has been associated with disrupted circadian rhythms (Leibenluft et al., 1995; Magnusson & Partonen, 2005). The advancement of smartphone platform and online testing provides a

powerful methodological framework where frequent sampling of behaviour in the morning and evening can provide an efficient way to measure circadian rhythm disruption in relation to mental health disorders (Brown et al., 2014). Gamification can also make a task more enjoyable, which may be particularly valuable in the study of adolescence when mental health problems peak (Paus et al., 2008), with potential for early interventions. However, researchers designing gamified tasks should maintain an awareness of how gamified mechanisms may affect cognitive or affective biases. In the *Great Brain Experiment,* we saw that participants were on average risk aversive in loss trials, and risk seeking in gain trials, counter to what is typically observed in the literature (Ruggeri et al., 2020). We speculate this may be due to the use of a spinner animation when the risky option is chosen, which takes several seconds to complete before the outcome is shown. This may increase risk taking for gains due to the inclusion of the positive affect of anticipation, and decrease risk taking for losses due to a negative effect of dread. Future work could test this hypothesis by varying the length of time the spinner animation plays for, including resolving the outcome instantly.

In this study, we analysed the relationship between time of day and risk taking without a causal manipulation of when participants played the task (i.e. they were not prompted to play at a particular time). Thus, we cannot statistically conclude that time of day is a causal influence on risk taking. Furthermore, as data was not collected on other possible factors with diurnal rhythms, such as wakefulness or hunger which may affect cognitive biases, we can only speculate on a serotonin based mechanism at this stage. Future studies which prompt participants to play at set times, could collect data on other factors and incorporate them as covariates using a within-subjects approach.

Many factors contribute to human risk taking and it would be surprising for gradual circadian changes to elicit large effects on risk taking. The robust effects on loss sensitivity that we observe could have significant implications at the societal level. International stock markets are simultaneously at different positions in their diurnal patterns (i.e. the New York Stock Exchange (NYSE) opens in the morning around when the Tokyo exchange ends trading for the day). Individuals making decisions about purchasing or selling stock from an international exchange may exhibit differences in loss sensitivity to individuals making purchases in local time zones. Given NYSE opening hours, investors in California may make decisions about NYSE-listed stocks earlier in their day on average than investors in Berlin. Time of day is known to be relevant to

investor behaviour, and it has been observed that buying near the closing of the market and selling at the opening next morning incurs greater returns than buying at opening and selling at closing (Kelly & Clark, 2011). Our finding of lower sensitivity to potential losses late in the day draws attention to one potential factor that may contribute to this effect. Policy changes that allow for more voting in the early morning, or late evening, could also have profound implications particularly if voters view candidates in a loss frame (i.e., they dislike both candidates) where we have shown that risk taking increases throughout the day. Understanding how diurnal biases in risk taking affect behaviour at individual and population level is useful for policy makers and might shape medical decision-making.

4.Gain and Loss Asymmetry for Short and Long Term Prospects

4.1 Abstract

Asymmetries in how we respond to positive and negative events have been observed in learning from reinforcement, and beliefs about the valence of future events. Intriguingly, in the former we often observe a *negativity bias* and in the latter an *optimism bias*. In both cases, these tendencies are typically observed in decision-making behaviour, and the link to differences in affective responses has been implied, but not been explicitly tested. We tested subjects (N = 81) in a novel probabilistic gambling task in which gain and loss trials were performed in short blocks while also displaying the valence of the next two blocks. Participants rated their happiness throughout the task. Firstly, we found using computational modelling, that losing points has a greater affective impact than winning the equivalent amount of points. We found this affective negativity bias is related to an increased propensity to make risky choices in loss trials, and is attenuated in participants reporting greater anhedonic symptoms. Secondly, we found that mood was not only affected by the current and past trial events, but also by the information related to the forthcoming trials. More specifically, participants were happier when the future contained more gain blocks, compared to when it contained more loss blocks. Intriguingly, we also observed increases in risky decision-making as participants approached a future with prospective gains, compared to small decreases in risk taking when a future with prospective losses was approached. This is consistent with work in *anticipation* and *dread* which suggests that the effect of future events on mood increases with temporal proximity.

4.2 Introduction

Expectations are ubiquitous and are key for adaptive behaviour. Moreover, what we expect to happen to us in the next minute, day or hour can influence our current emotional state. For example, the expectation of a soon to be enjoyed ice cream may have a positive effect on our emotions, additionally looking forward to a weekend at the beach several days in advance may

also fill us with delight. Expecting to have to walk home in the rain later in the afternoon, or dreading a difficult exam on Monday can have a negative impact on one's current mood.

The influence of expectations on mood has been primarily explored in terms of *deviations* from them (i.e. prediction errors). Happiness in particular, has been shown to be driven by prediction errors in laboratory settings (Blain & Rutledge, 2020; Eldar & Niv, 2015; Keren et al., 2019; Rutledge et al., 2014; Rutledge et al., 2017) and in the wild; where deviations from expected exam results (Villano et al., 2020), and unexpected sports team victories or sunny days also drive changes in mood (Otto et al., 2016). However, the unique role expectations themselves may themselves play in mood, has been not been fully defined. Studies have explored the role of deterministic future rewards or punishments on decision-making. Specifically, participants show an increased preference for learning about future rewarding outcomes when a delay is longer, suggesting that the experience of *anticipation* itself has a positive utility (Iigaya et al., 2016, 2019; Loewenstein, 1987). In complimentary research on future deterministic electric shocks, many participants preferred to receive a larger shock sooner, than a smaller shock later, suggesting that experiencing *dread* has a negative utility (Berns et al., 2006; Story et al., 2013). While both these studies imply the existence of an effect of future prospects on mood, they did not collect data regarding changes in affective state.

Expectations about longer term futures have been explored in studies looking at abstract beliefs about future prospects (e.g. belief in susceptible to future harmful events; Weinstein, 1987). Generally, an *optimism bias* has been observed, where individuals rate themselves as less likely to befall prospective negative future events (such as smoking related premature morbidity; Schoenbaum, 1997); and mental and physical harm; Weinstein, 1987) and more likely to encounter positive events, and update their beliefs in response to good news rather than bad news (Sharot, 2011; Sharot & Garrett, 2016). Notably trait optimism has been associated with better mental and physical health outcomes (Conversano et al., 2010; Scheier & Carver, 1987), however the mechanism by which this occurs is unclear. It has been suggested that optimistic beliefs promote behaviour that increases the potential of rewarding outcomes through increased confidence and self-belief (Bouchard et al., 2018). An alternative hypothesis is that expectations about the future intrinsically contain affective properties, thus optimistic beliefs boost current mood and future well-being.

In contrast to the *optimism bias* pertaining to abstract future events, a *negativity bias* tends to be observed in more low-level information processing. Negative information has been shown to receive preferential processing in terms of learning from reinforcement (Gershman, 2015; Niv et al., 2012), and increased physiological arousal (Hochman & Yechiam, 2011). For example, it has been shown that learning from negative prediction errors exceeds learning from positive prediction errors (Gershman, 2015). Counterintuitively, a *negativity bias* has been suggested to have a positive effect on mood, specifically due to its association with successful avoidance of negative future events (Michely et al., 2020). Evidence for a general *positivity bias* can be seen on occasion (e.g. Lefebvre et al., 2017).

While both a *negativity bias* and an *optimism bias* may appear contradictory, it has been suggested that both represent an adaptation to environmental statistics, with positive events being more common than negative ones (Pulcu & Browning, 2017). In this study we examine whether there exists both a *negativity bias* and an *optimism bias* in mood by explicitly quantifying the affective consequences of gains and losses in the short and longer term. We designed a novel experimental paradigm where participants performed blocks of trials where they can only gain points (or miss out on gaining), or lose points (or avoid losing). Importantly, participants also see the whether the blocks in the next future will be predominantly opportunities to gain points, or lose points. Regarding the *negativity bias*, we predict that prediction errors for losing points will have a greater impact on mood than prediction errors for gaining points.

Regarding the effect of future prospects, we predict that futures containing gain trials (i.e. positive futures) will increase mood, while futures containing loss trials (i.e. negative futures) will decrease mood. If mood is successfully manipulated by future valence, we will predict there will also be an effect on decision-making: specifically we predict that positive futures will increase risk taking, and negative futures will decrease risk taking. Regarding the *optimism bias* we will explore whether this is present by contrasting the effects of positive futures with negative futures.

4.3 Methods

4.3.1 Participants

81 young adults (18-35) were recruited through University College London Psychology Database. Participants were screened for no history of psychiatric or neurological disorder. The study was approved by the University College London research ethics committee, and all participants gave written informed consent.

4.3.2 Task Design

Stimuli were presented in MATLAB (MathWorks, Inc) using Cogent 2000. All participants performed a 196(±2) trial economic decision-making task including trials were there were prospective gains (i.e. gain trials) and trials with prospective losses only (i.e. loss trials). All trials were independent, and the magnitude and probability of each option was explicit so no learning was required. In the gain trials, participants chose between a reference option (referred to hence forth as the 'safe gamble') with a 80% chance of winning +20 points. The alternative option's magnitude (the 'risky gamble') were drawn from a set of [25 40 65 80] with a probability varying between 10% and 70% in increments of 10. Thus the risky gamble always has a greater magnitude than the safe gamble, and a lower probability. We refer to choosing the risky gamble as "risky decision-making", while acknowledging that choosing the safe gamble is also a risk. The loss trial design matrix was identical to the gains, but with negative magnitudes. The maximum absolute outcome on a trial was 80 (maximum expected value 56) and the minimum was 20 (minimum expected value 4). The risky gamble had a greater absolute expected value for 60% of the gambles. Earnings were cumulative but not shown on screen and participants were paid based on their total points at the end of the task. The trials were presented in a series of blocks, each block containing 2 or 3 trials of the same valence (Figure 4-1). Importantly, on the top half of the screen the next two blocks were shown. Red blocks indicated future blocks would be loss trials, green blocks indicated future trials would be gain trials. Future prospects could be either: positive (two gain blocks), negative (two loss blocks), or mixed (one loss block and one gain block Figure 1-2). Mood was sampled frequently throughout the experience to probe how task events contribute to mood fluctuations. Participants are asked '*How happy are you at this moment?*' on a scale ranging from *very unhappy* to *very happy*. After the happiness rating was made, participants would see a brief animation transitioning them into the next

block. This method of quantifying moment to moment changes in happiness has been used in behavioural and fMRI studies, and used on a smartphone platform with over 200,000 unique users (Rutledge et al., 2014). Importantly, this question makes no reference to current or future task events. Participants were able to practice 20 blocks of the task prior to beginning the main experiment.

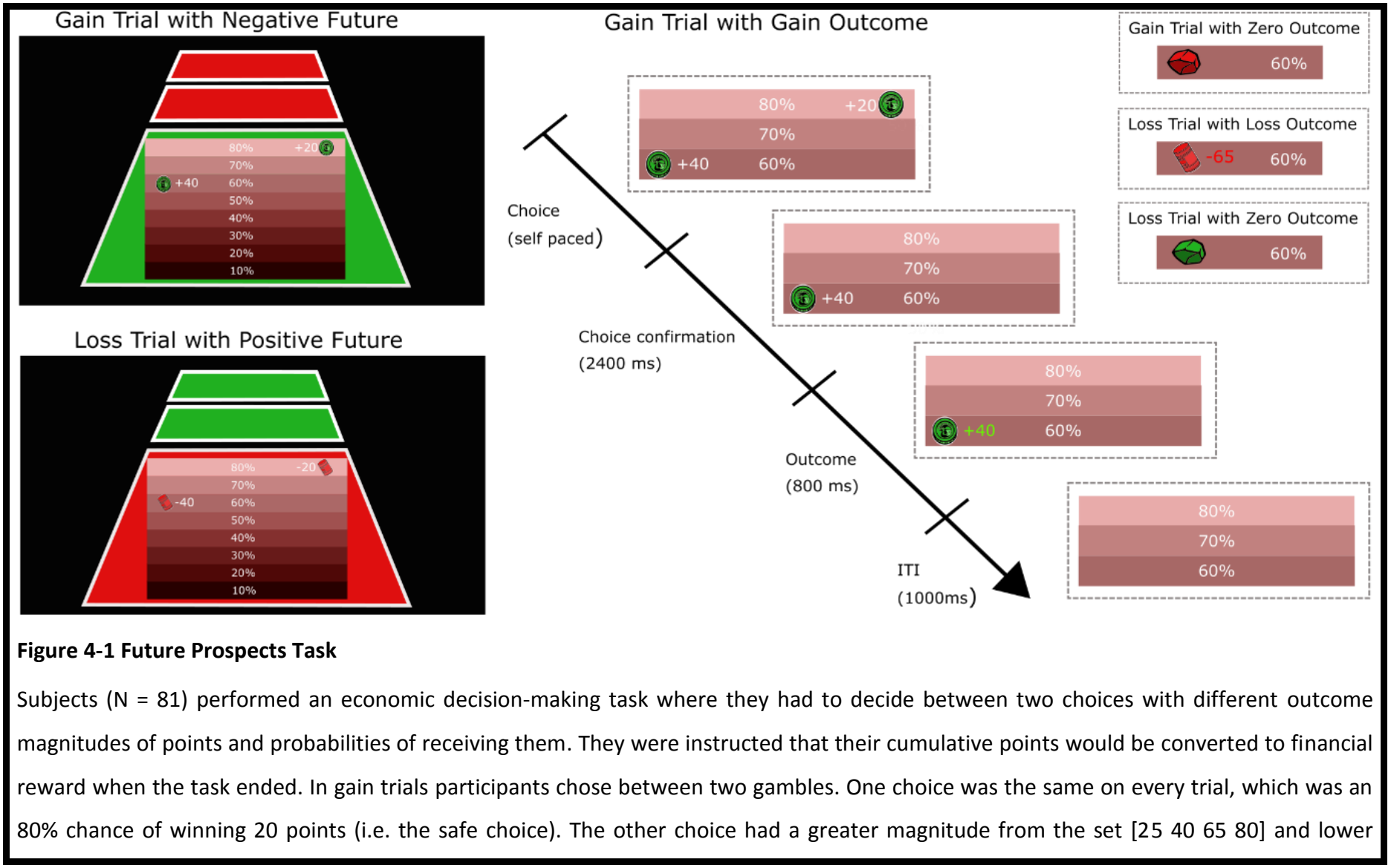


Figure 4-1 Future Prospects Task

Subjects (N = 81) performed an economic decision-making task where they had to decide between two choices with different outcome magnitudes of points and probabilities of receiving them. They were instructed that their cumulative points would be converted to financial reward when the task ended. In gain trials participants chose between two gambles. One choice was the same on every trial, which was an 80% chance of winning 20 points (i.e. the safe choice). The other choice had a greater magnitude from the set [25 40 65 80] and lower

probability (between 70% and 10%), termed the risky choice. All combinations of probability and magnitude were shown for the risky choices. The example shows a risky choice of 60% chance of 40 points. The probabilities were arranged on screen so objects lower down the screen had lower probabilities – this was explained intuitively to participants as items being 'harder to dig up'. In loss trials participants chose between outcomes with negative magnitudes. Participants had explicit knowledge of the block they were currently in (but not the length) by means of a coloured border (green for gain trials, red of loss trials) around the area of the screen where the choice options resided. Each trial began with a self-paced choice between the safe and risky choice. After the decision is made by clicking left or right the choice remains on screen for 2400ms and the unchosen option is removed. If the points are obtained the magnitude text changes colour (green in gain trials, red in loss trials) to indicate that the magnitude were added to the participants cumulative earnings. If they do not obtain the magnitude they are shown a rock indicating no points were added to their score. 38 Participants performed the design shown. The other 43 participants partook in an alternative design which did not include a 10% probability option and there was no colour gradation in the choice area.

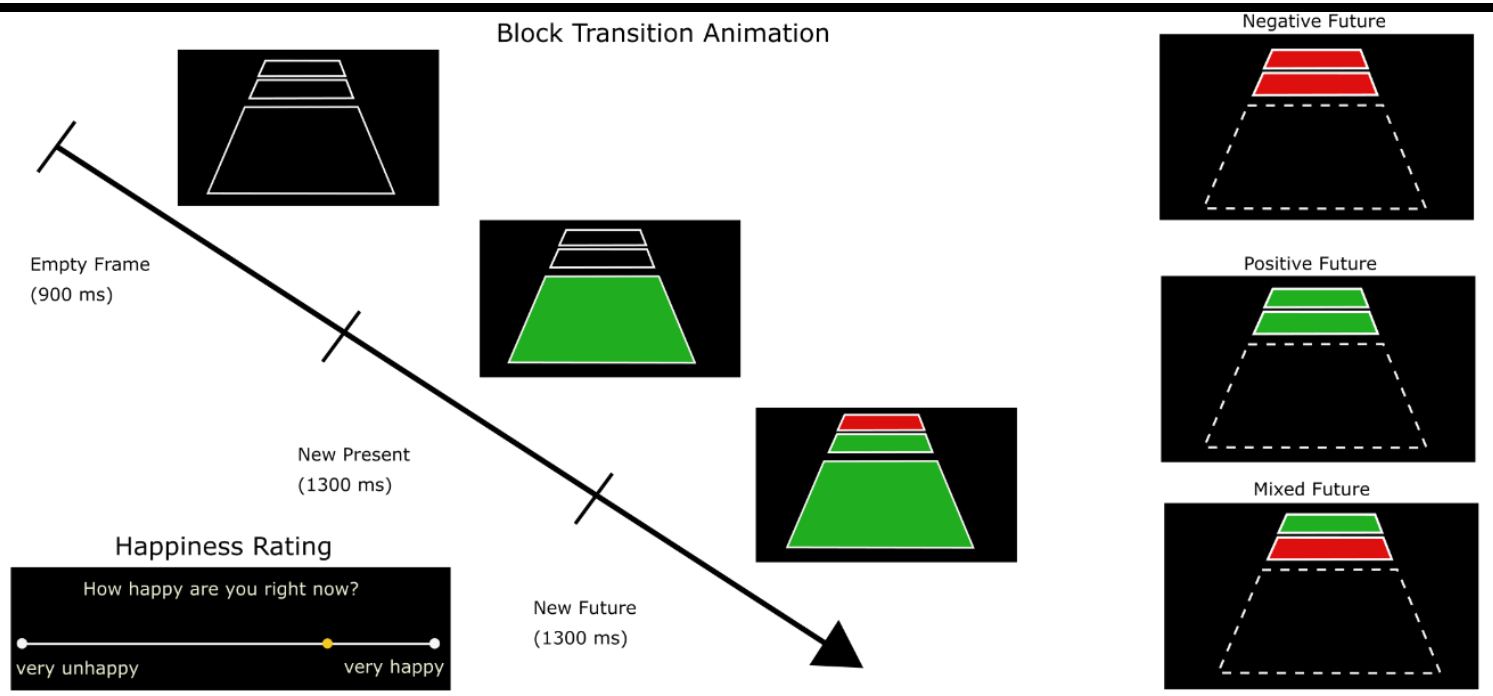


Figure 4-2 Future Block Transition

While participants made choices, they were also able to see the two subsequent blocks which could be either two loss blocks (negative futures), two gain blocks (positive futures), or a gain and loss block (mixed futures). Every 2-3 trials participants were asked to rate their happiness by answering the question *how happy are you right now?* by moving a cursor on a line with the computer mouse. A block transition animation was then shown before the next block began. Participants performed on median of 45 trials with positive futures, and 45 with negative futures (minimum 39 maximum 53) corresponding to between 15 and 21 blocks each of 2-3 trials (median 18). The remaining trials all had mixed futures (median 41 blocks, between 95 and 112 trials).

4.3.3 Questionnaires

We included five questionnaires, Beck's Depression Inventory II (BDI-II; Beck et al., 1996), the trait section of the State Trait Anxiety Inventory (STAI; Spielberger et al., 1970), Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), Rumination Response Style Questionnaire (RRSQ; Nolen-Hoeksema & Morrow, 1991) and the Mood and Anxiety Scale Questionnaire Anhedonia Subscale (MASQ-Anhed; Huys et al., 2013; Watson et al., 1995). Questionnaires were completed prior to the task being performed in the same session. The first 43 participants completed the MASQ and RRSQ online after the laboratory session.

4.3.4 Statistical Procedures

Two-sided Wilcoxon signed rank tests were used to compare parameter estimates in all computational models at group level. Spearman rank correlations across participants were used to test for relationships between parameter estimations and questionnaire scores. All averages were reported as (mean ± standard deviation), unless otherwise stated. All analyses were performed using MATLAB. The minimum p value threshold reported is $p < 0.0001$.

In order to correct for multiple comparisons where the tests are not independent we can use a correction to control for the *false discovery rate* (FDR) i.e. the expected proportion of false positives. Here we use the *Benjamini-Hochberg Procedure* (Benjamini & Hochberg, 1995). In this procedure the hypothesis tests are ranked according to the lowest to the highest p value. The p values are then compared to a threshold determined by the number of tests, desired alpha value (in this case 0.05) and position in the ranking.[17] . The test with the greatest rank where

[17] This test is similar to the *Holm-Bonferroni Method* – where the statistical tests are performed in sequence. The sequence is terminated once a hypothesis test exceeds the threshold for its ranking. Thus this test controls the *family wise error rate* and is more conservative than the *Benjamini-Hochberg Procedure.* We considered a less conservative test more appropriate due to the shared variance in the questionnaire scores and apply conservativeness to the specificity of our interpretations.

the p value does not exceed the threshold determines the rank at which all tests below that rank are deemed significant. Significance of tests will be reported as (*Rho, p^{HB} adjusted threshold, actual p value*).

4.4 Results

The results presented for this task include happiness ratings and decision-making in regards to current state (i.e. performing gain or loss trials) and future states (i.e. positive or negative futures). For ease of reading, much of the modelling work is located after the main results and discussion in Section 4.6 *Extended Analyses & Results*.

4.4.1 Participants Removed

Two criteria for each participant's data were defined to ensure data quality. Firstly they had to choose the riskier gamble at least 50% of the time when it was the best option in either gain or loss trials. This is to ensure we are able to model their behavioural and happiness where we assume expectations are determined by probability and magnitude. Only one participant was removed under this criteria (Figure 4-3a). The second criteria was that participants had to rate their mood not at the centre, or the extreme end of the scale more than 50% of the time. This was to ensure adequate variance in their happiness scores that could be captured by a model. Under this criteria we excluded 2 participants (Figure 4-3b)

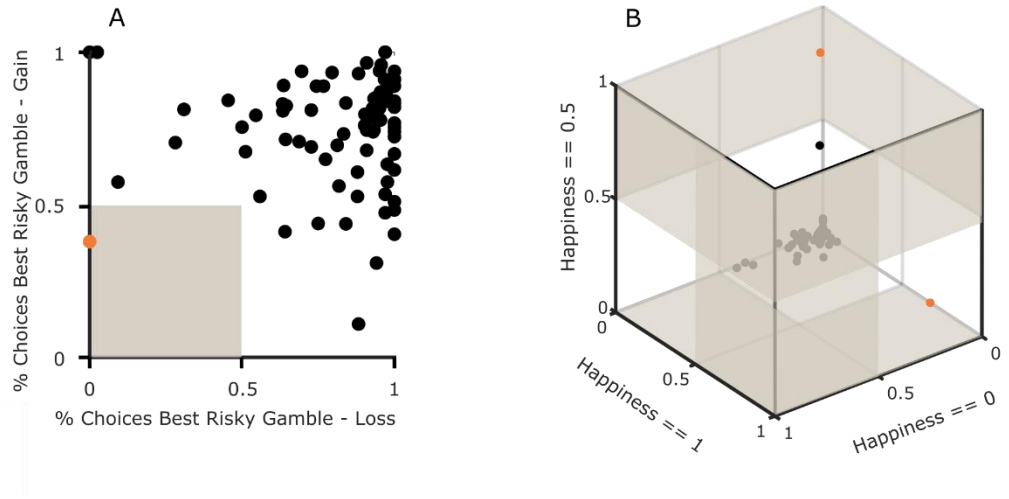


Figure 4-3 Excluded Participants

Grey areas of the plot indicate the exclusion criteria. Orange data points indicate excluded participants. **a)** Participants (N = 81) were excluded from all analyses if they chose the risky gamble less than 50% of the time when it was the option with the greatest expected value in both loss and gain trials. One participant was excluded for choosing the risky option in gain trials only 38% of the time, and none of the time in loss trials. **b)** Participants were excluded if they rated their happiness as 0, 0.5 or 1 more than 50% of the time. Two participants were excluded under this criterion. Separate participants were excluded based on each criteria.

4.4.2 Questionnaire Measures

Nine Participants did not complete the MASQ or RRSQ. These were unique participants to those removed in the proceeding section. All questionnaires were highly correlated with a mean Pearson's *r* of 0.670 and a minimum *r* of 0.352. All questionnaires scores were not significantly different between the two versions ran of the task (all $p > 0.11$, MASQ $p = 0.309$; Figure 4-1)

4.4.3 Participants have Stronger Affective Responses to Losses Compared to Gains: Evidence for an Affective Negativity Bias

We first confirmed that choice outcomes impacted happiness ratings. We found that participants were happier after gaining points than after missing out on gaining (gaining: 0.565 ± 0.113; missing out on gaining: 0.422 ± 0.120; $z = 7.673$, $p < 0.0001$) and that they were happier after avoiding losing, than after losing (losing: 0.390 ± 0.120; avoiding losing: 0.519 ± 0.120; $z = 7.414$, $p < 0.0001$). Interestingly, participants reported to be happier after avoiding losing, than after missing out on gaining, while the outcome is zero points in both trial types ($z = 6.492$, $p < 0.0001$). This suggests an effect of expectations, as well as the involvement of outcomes in happiness fluctuations. We next turned to comparing gain and loss trials overall. Participants were happier in gain trials (0.517 ± 0.107) compared to loss trials (0.449 ± 0.109; $z = 6.811$, $p < 0.0001$). Furthermore, those who were happier in gain blocks were also happier in loss blocks ($\rho(76) = 0.757$, $p < 0.0001$) suggesting happiness ratings capture important variance about overall mood state, independent of task events.

We next designed computational models to capture both the aforementioned expectation and outcome effects in fluctuations in happiness ratings. All models we compared are shown in Section 4.6.1 (*Models of Affective Dynamics*). For the succeeding analyses, we used a model including the recent history of expectations as well as the recent history of the difference between the choice outcome and the expectation, namely the prediction error to predict each participant's happiness ratings. The full model is described by below (Equation 4-1 *Expectations Model*, which is a formation of Equation 1-3 *Affective Dynamics Model*). In the model comparison in Section 4.6.4, this model is described as *Expectations Model (3b)*.

$$Happiness_t = C + w_{EVloss}\sum_{j=1}^{t}\tau^{t-j}\ EVLoss_j + w_{EVgain}\sum_{j=1}^{t}\tau^{t-j}\ EVGain_j + w_{PEgain}\sum_{j=1}^{t}\tau^{t-j}\ PEGain_j + w_{PEloss}\sum_{j=1}^{t}\tau^{t-j}\ PELoss_j$$

Equation 4-1 Expectations Model

In the Expectations Model C , w_i and τ are free parameters. The constant term C describes baseline mood. This was on average around the centre of the scale (0.535 ± 0.103). Weights were included that parameterized the influence of expected value and prediction errors on happiness ratings, termed *affective dynamics* (Figure 4-4b). Affective Dynamics for gain and loss trials were modelled separately as expected value (wEV^{loss}: 0.178 ± 0.210; wEV^{gain}: 0.0686 ± 0.0882) and prediction error weights (wPE^{loss}: 0.537 ± 0.403; wPE^{gain}: 0.414 ± 0.287). The decay parameter τ captured the influence of previous trial events on the happiness rating (0.643 ± 0.244), suggested that approximately 4-5 trials in the past influenced the current rating (Figure 1-3). Interestingly, prediction error weights were significantly higher in the loss trials than the gain trials ($z = 4.45$, $p < 0.0001$; Figure 4-4c) as were expected value weights ($z = 4.24$, $p < 0.0001$). This suggests that participants had stronger affective reactions to equivalent outcomes in loss trials compared to gain trials. For example, using the average prediction error parameter estimates, a prediction error of winning 60 points with a chance of 20% would boost happiness by 20 points in a gain trial, but a prediction error of losing 60 points would decrease happiness by 26 points in the loss trial. We consider this to be evidence of an *affective negativity bias*. In the continuation of the results we will refer to the difference in prediction error weights as the *negativity bias* (work concerning the difference in expected value weights is in Section 4.6.3).

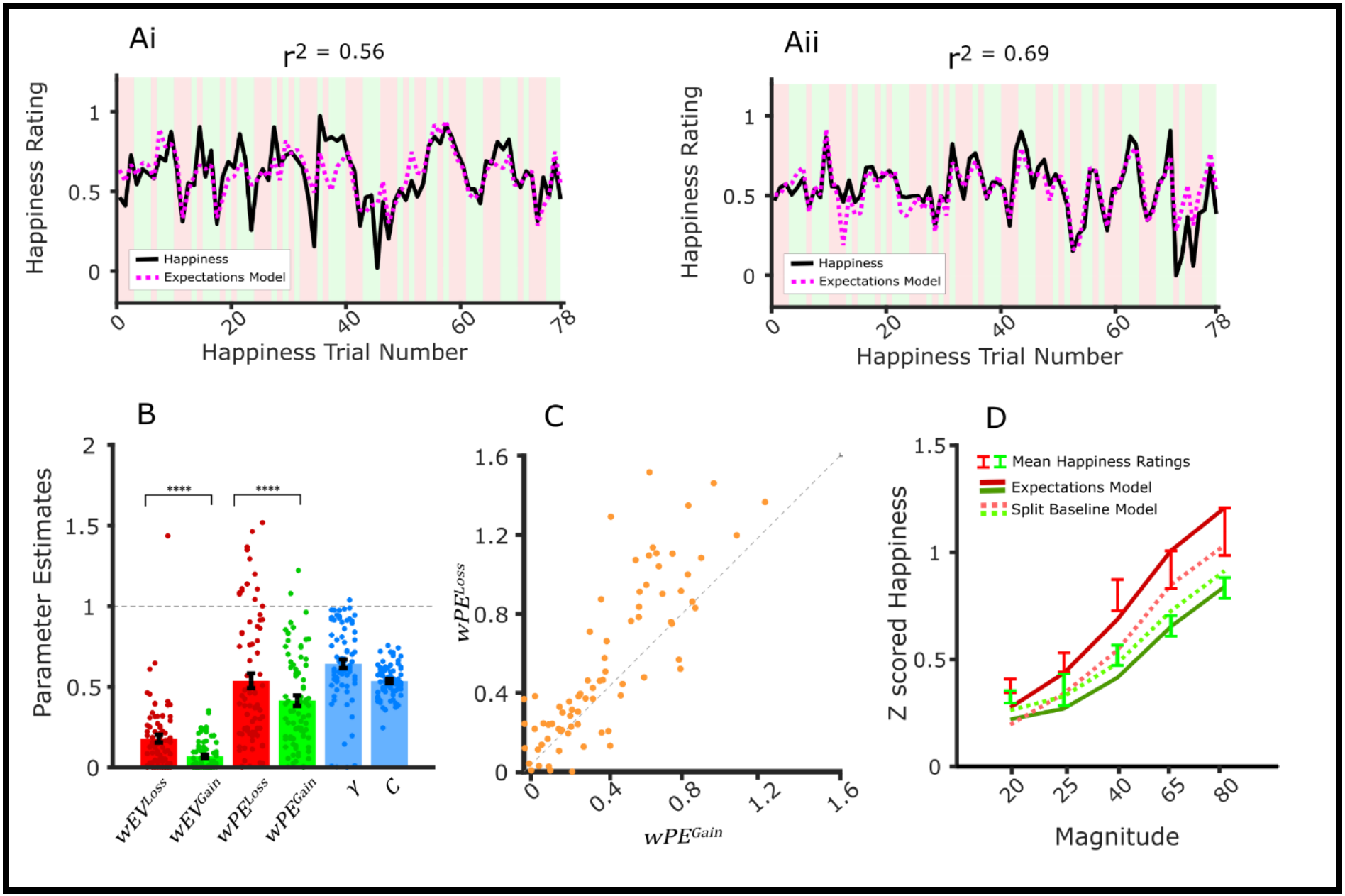

Ai
r2 = 0.56
Happiness Rating
Happiness
Expectations Model
Happiness Trial Number
Aii
r2 = 0.69
Happiness Rating
Happiness
Expectations Model
Happiness Trial Number
B
Parameter Estimates
****
****
wEVLoss
wEVGain
wPELoss
wPEGain
γ
C
C
wPELoss
wPEGain
D
Mean Happiness Ratings
Expectations Model
Split Baseline Model
Z scored Happiness
Magnitude

Figure 4-4 Affective Responses to Losses are Greater than those for Gains

A) Fluctuations in momentary happiness (black filled line) are well explained by a model (magenta dotted line) which includes parameters that estimate the influence of expected value, prediction errors on happiness (i.e. Expectations model). Happiness ratings made after gain trials are shown with a green background, ratings made after a loss trial are shown with a red background. **Ai)** Shows the happiness ratings for the participant with the median r^2 value. **Aii)** Shows the participant with the median *negativity bias* (0.091). **B)** The Expectations model included parameter estimates which weighted events in gain trials (green bars) and loss trials (red bars) separately (termed *affective dynamics*). Affective dynamics parameter estimates convert trial-by-trial points to a change in the happiness rating. If the expected value was +80 points, and the $\mathrm{wEV}^{\mathrm{gain}}$ is 0.25, happiness ratings would increase by 20% of the scale. Baseline parameter C and decay parameter τ (blue bars) were fit to all trials. Bars shown the mean parameter estimate and the error bars show the standard error of the mean. Each participant's individual estimate is shown as a data point in the same width of the bar. **C)** Parameter estimates for prediction errors were greater in loss trials than in gain trials ($z = 4.45$, $p < 0.0001$). **D)** Error bars show the mean average z-scored happiness rating for each magnitude outcome. Filled lines show the *Expected Model* estimates. Dotted lines show the estimates of an alternative model where the affective dynamic parameters are fit to loss and gain trials together, and a separate baseline parameters (C) are fit to loss and gain trials.

[*p < 0.05, ** p < 0.01, *** p < 0.001, , **** p < 0.0001]

4.4.4 Negativity Bias Associated Increased Sensitivity to Losses

We next tested whether the parameter estimates in the affective dynamics model related to decision-making. Overall, participants made the risky choices in a similar proportion between gain trials and in loss trials (gain trials: 0.488 ± 0.149; loss trials: 0.524 ± 0.187; $z = 1.152$, $p = 0.128$). Participants number of risky choices in gain trials was not related to their choices in loss trials ($\rho(76) = -0.192$, $p = 0.0926$). The proportion of risky choices did not correlate with any of the self-report questionnaires (all $\rho < 0.188$, $p > 0.122$). Mean happiness was not related to the overall propensity to take risks; the proportion of risky choices made did not correlate with mean happiness in loss blocks ($\rho(76) = 0.096$, $p = 0.403$) nor gain blocks ($\rho(76) = -0.066$ $p = 0.565$).

Focusing on the negativity bias for happiness ratings, we looked for whether this was related to any differences in choices made by the participants. An asymmetry between how one feels about positive and negative events may lead us to change our behaviour accordingly to maximize feeling good and avoid feeling bad. In other words, if losing points makes us feel worse than winning the equivalent gains, we may actively seek to avoid large losses more than we look to approach large gains. In order to investigate this we tested whether the prediction error parameters were correlated with risk taking in loss and gain trials separately. The prediction error weight was negatively correlated with proportion of risky choices in the gain trials ($\rho(76) = -0.308$ $p = 0.0060$), suggesting that those that were happier in response to outcomes in gain trials also chose the risky option less frequently. We did not see a relationship in loss trials ($\rho(76) = -0.055$ $p = 0.632$), nor a relationship between the negativity bias and the difference in proportional risky decisions when loss compared to gain trials ($\rho(76) = -0.112$ $p = 0.331$).

In order to test whether any of the affective dynamics were related to differences in the subjective perception of magnitudes or probabilities used to make choices, we used a series of prospect theory-like model to fit the choices which included a probability weighting function (Section 4.6.4 *Comparing Models of Choice*). For the succeeding analyses, we used a model which fits free parameters for gain and loss trials separately (the equation below does not specify trial frame for clarity; Figure 4-5).

$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}}$$

$$U_{safe} = w(p)(Magnitude)^{\alpha}$$

$$U_{risky} = w(p)(Magnitude)^{\alpha}$$

Equation 4-2 Prospect Theory Model with Tversky Probability Weighting Function

Choice probabilities were then determined by a softmax rule which converts the difference in subjective utilities (U_{risky} and U_{safe}) to the probability of choosing the riskier option. The softmax includes an inverse temperature parameter (μ) which quantifies choice stochasticity (Equation 4-3 & Figure 1-2 Softmax Rule Simulation) As above, we fit μ separately for both loss and gain trials but omit this from the equation below.

$$P_{gamble} = \frac{1}{1 + e^{-\mu(U_{risky} - U_{safe})}}$$

Equation 4-3 Softmax Rule

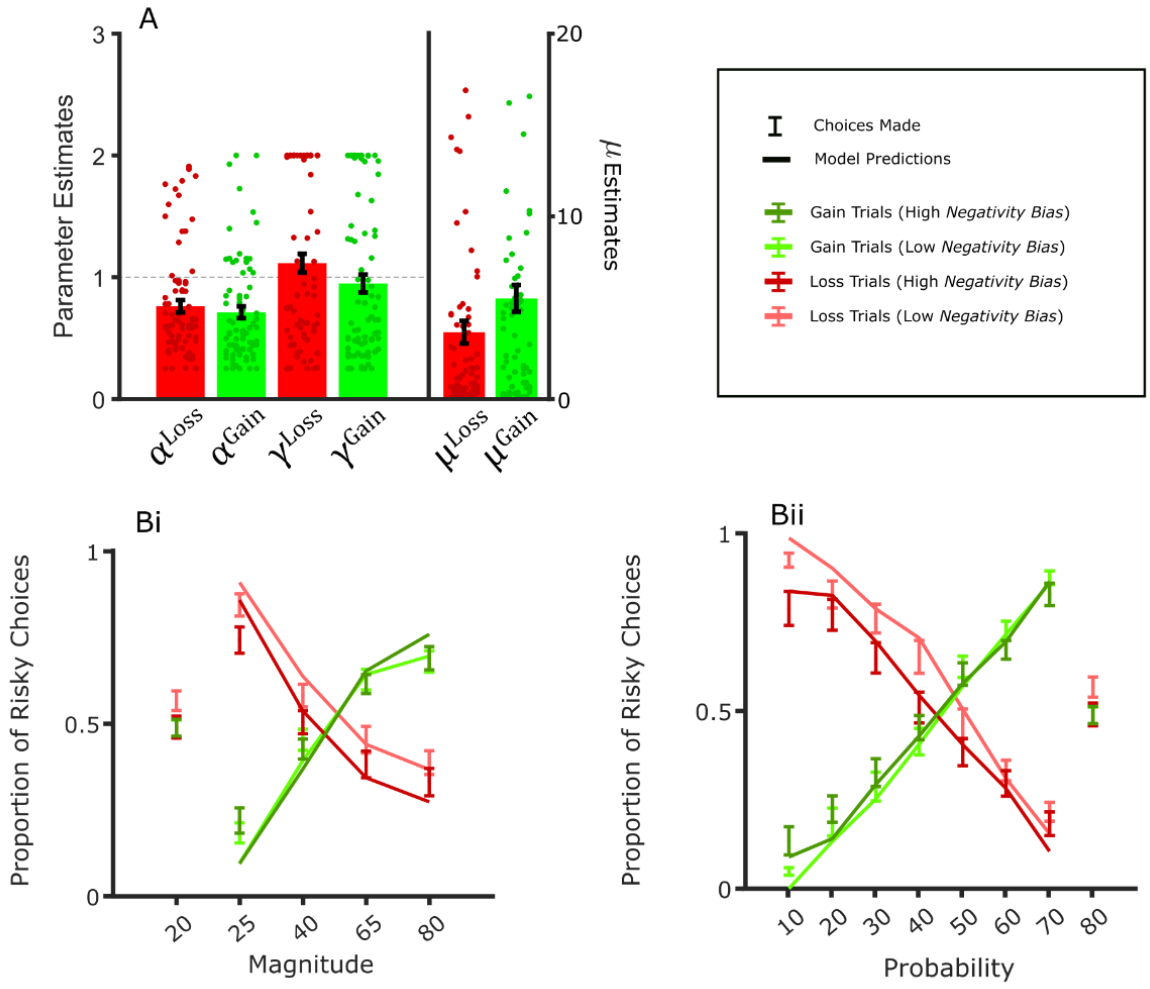


Figure 4-5 Prospect Theory Style Model of Decision-making

(A) The Prospect theory style model included parameters which estimate the influence of the subjective perception of magnitudes (α) and probability (γ) on choice. It also includes a parameter which indexes how deterministic participant's choices were (μ). Bars shown the mean parameter estimate and the error bars show the standard error of the mean. Each participant's individual estimate is shown as a data point in the same width of the bar. Parameter estimates for loss trials are in red, estimates for gain trials are in green. **(B)** The mean proportion of risky choices is shown in error bars with the standard deviation (red for loss trials, green for gain trials) for each magnitude **(Bi)** and probability **(Bii)**. Joined lines indicate the Prospect Theory model's mean predictions. We median split the data based on whether participants had a high negativity bias (mean 0.287, N = 39) or low negativity bias (mean -0.041, N = 39) for presentation purposes.

We next tested whether there were correlations between the prediction error weights with the subjective perception of magnitudes (α), and probabilities (γ). We found correlations between these in gain and loss trials, but they did not exceed the multiple-comparisons adjusted threshold (α^{loss} : ρ(76) = 0.217, $p^{HB} < 0.0167$**,** p = 0.057; γ^{gain}: ρ(76) = 0.217, $p^{HB} < 0.0167$, p = 0.057) [18]. We next tested whether the negativity bias showed a relationship with any parameters. Here we found that α^{loss} correlated with the negativity bias (ρ(76) = 0.2996, $p^{HB} < 0.0083$, p = 0.0079, Figure 4-6c), but not any other estimates of subjective magnitude or probability, nor the difference between them. This suggests that those who were comparatively more upset about losses than they are about gains, were also more likely to choose the smaller magnitude more likely loss option over the larger magnitude less likely loss option. The relationship between α^{loss} and the negativity bias was in part largely driven by the participants who showed a more positive bias; where they showed greater prediction error weighted responses to gains than losses. When they were removed from the analysis (N = 19) there was no significant correlation (ρ(57) = 0.0420, p = 0.751). We also tested the *Prospect Theory Model* with only one set of parameters (i.e. all shared between gain and loss) which demonstrated the same effect for the α fit to both gain and loss (α: 0.798 ± 0.493, ρ(78) = 0.297, $p^{BH} < 0.0167$, p = 0.0084). This may suggest the negativity bias is associated with increased sensitivity to magnitudes overall (higher values of α indicating relatively greater risk taking for gains, and safe choices for losses).

4.4.5 Negativity Bias Associated with Reduced Anhedonia Symptoms

We next looked to see whether the negativity bias was associated with any of the questionnaire measures. We found that greater self-reported scores on the MASQ were negatively correlated with the average happiness (ρ(67) = -0.402, $p^{HB} < 0.01$, p = 0.0006) as were scores on the STAI (ρ(76) = -0.273, $p^{HB} < 0.02$, p = 0.0157) and the BDI (ρ(76) = -0.265, $p^{HB} < 0.03$, p = 0.0189). We next used the parameter estimates from the *Expectations Model* to test whether this may explained by a lower mood baseline or an increased negativity bias. The MASQ questionnaire negatively correlated with both the baseline parameter C (ρ(67) = -0.415, $p^{HB} < 0.0167$, p =

[18] The test statistics in these paragraphs were identical when rounded, this is not a transcription error.

0.0004, Figure 4-6a) and the negativity bias (ρ(67) = -0.290, $p^{HB} < 0.0333$, p = 0.0155, Figure 4-6b). It did not correlate with the difference in expected value parameter estimates between loss and gain trials (ρ(67) = -0.108, p = 0.501). An exploratory analysis showed that the relationship with MASQ was present for $\mathrm{wPE}^{\mathrm{loss}}$ (ρ(67) = -0.251, p = 0.0372) but not $\mathrm{wPE}^{\mathrm{gain}}$ (ρ(67) = 0.131, p = 0.284). These findings suggest that those with greater self-reported anhedonia have lower mood baselines, and also reduced negativity bias (Further analysis to support the robustness of these results can be found in Section 4.6.3.1 (*The Asymmetry of Prediction Errors is not caused by Variance in Mood Baselines*.).

Having seen relationships between the negativity bias with MASQ, and with risk sensitivity for losses (α^{loss}), we looked to see whether MASQ scores were also associated with α^{loss}. We did not find a relationship between α^{loss} nor any other parameters in the model (all ρ(57) < -0.215, p > 0.0764). Interestingly we did find a relationship between MASQ scores and response times for both gain (MASQ: ρ(57) = 0.436, $p^{BH} < 0.01$, p < 0.000182; RRSQ: ρ(78) = 0.306, $p^{BH} < 0.02$, p = 0.0105; BDI, ρ(78) = 0.238, $p^{BH} < 0.03$, p = 0.0249) and loss trials (MASQ, ρ(57) = 0.433, $p^{BH} < 0.01$, p = 0.000205). Higher MASQ scores (mean score 71.177, N = 34) had an average response time of 1.900 minutes for Loss trials and 1.70 for Gain trials, while lower MASQ scores (mean score 49.971, N = 35) had an average of 1.48 minutes for Loss trials and 1.35 minutes for Gain trials. MASQ scores were not associated with response times for happiness ratings for gain trials (ρ(57) = 0.209, $p^{BH} < 0.02$, p = 0.0856) nor loss trials (ρ(57) = 0.195, $p^{BH} < 0.02$, p = 0.109), suggesting that anhedonia may be associated with prolonged deliberation time for choices.

Interestingly, we found that the RRSQ positively correlated with reporting time for happiness ratings, however the p values exceeded the adjusted threshold (Gain trials: ρ(78) = 0.284, $p^{BH} < 0.01$, p = 0.0182; Loss trials: ρ(78) = 0.288, $p^{BH} < 0.01$, p = 0.0166). This may suggest that participants reporting increased rumination in their day to day life take more time to consider their happiness ratings.

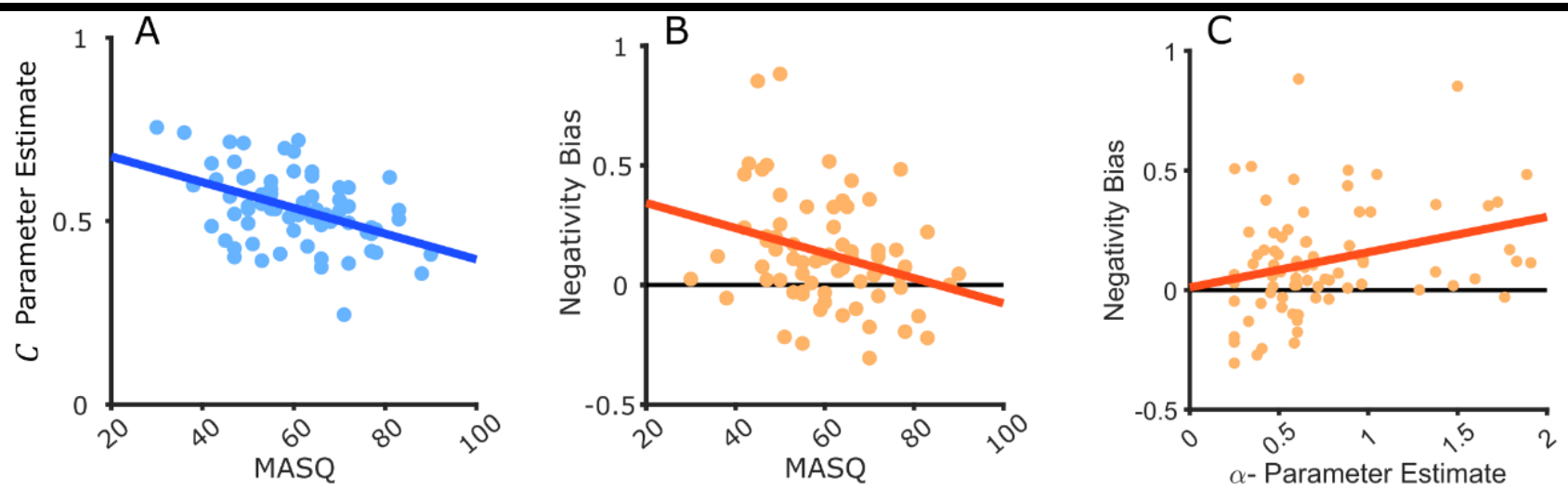


Figure 4-6 Anhedonia Symptoms Associated with Decreased Mood Baseline and Reduced Affective Asymmetry

(A) The baseline parameter C negatively correlates with scores on the MASQ questionnaire, demonstrating that participants who report greater anhedonia symptoms have a lower mood baseline. **(B)** The difference in negativity bias was negatively correlated with MASQ scores suggesting those that report less anhedonia symptoms show increased affect reactivity to losses compared to gains. In order to test whether the difference in prediction error weights included variance driven by changes in the baseline parameter we ran a partial correlation test and found that the negativity bias correlated with MASQ scores when the baseline was partialled out ($\rho(67) = -0.31$, $p = 0.009$). **(C)** The negativity bias was positively correlated with α^{Loss}. Higher values of α^{Loss} indicate the higher magnitudes being overweighed and choosing the riskier option appears less attractive. The analysis was run again removing those with negative values of negativity bias (where the prediction weights were greater for gains than losses (N = 19) and the correlation was no longer significant ($\rho(78) = 0.0420$, $p = 0.751$).

4.4.6 Affective Dynamics were influenced by Positive and Negative Futures

In the previous sections, we have shown evidence for a negativity bias, where participants show larger changes in their happiness ratings for prediction errors in loss trials compared to gain trials. We found the negativity bias is related to risk sensitivity in choices, particularly sensitivity to prospective losses. Next, we looked to see whether future prospects had an effect on happiness ratings. We used the *Expectations Model* to regress out the influence of all trial-by-trial events on happiness to determine if the future blocks had any influence on the ratings (a method previously used in Vinckier et al., 2018a). Positive residual happiness means that the model has under predicted the participant's happiness ratings (i.e. they are happier than would be estimated given the current trial expectations and outcomes). Negative residual error means that the model has over predicted the participant's happiness ratings (i.e. they are less happy than would be estimated given the current trial expectations and outcomes). We first tested whether the future had an effect on happiness ratings overall, by comparing the happiness residuals between positive and negative futures. These were significantly different across all trials (0.0229 ± 0.053, $z = 3.44$, $p < 0.00058$) in both loss trials (0.0220 ± 0.0613, $z = 2.96$, $p < 0.0031$) and gain trials (0.0251 ± 0.0654, $z = 2.98$, $p > 0.0029$; Figure 4-7a). As a robustness check, we tested whether these differences were present in both the first and second half of the experiment, where it was found to be the case in both loss trials (first half: $z = -2.08$, $p = 0.038$; second half: $z = 2.015$, $p = 0.044$) and gain trials (first half: $z = 2.52$, $p = 0.012$; second half: $z = 2.63$, $p = 0.009$). Across all trials, the mean difference in happiness residuals for positive vs negative futures was positively correlated between gain and loss trials ($\rho(76) = 0.367$, $p < 0.001$; Figure 4-7b).

We next looked to see whether we could see effects in negative and positive futures separately. For looking at positive and negative futures separately, we compared residual happiness for the mixed prospects future in order to account for whether an individual's residuals were on average positive or negative. Residual happiness was significantly positive for positive futures (0.0096 ± 0.037, $z = 2.25$, $p = 0.024$), and significantly negative for negative futures (-0.013 ± 0.040, $z = 2.90$, $p = 0.0037$). Positive futures were significantly positive in gain trials (0.0124 ± 0.0444, $z = 3.0158$ $p = 0.0026$) but not significant in loss trials (0.0080 ± 0.0517, $z = 1.163$, $p =$

0.2448). Negative futures were significantly negative in loss trials (-0.014 ± 0.0406, z = -2.93, p = 0.0034) and showed a negative trend for gain trials (-0.0127 ± 0.0576, z = -1.7009, p = 0.0890).

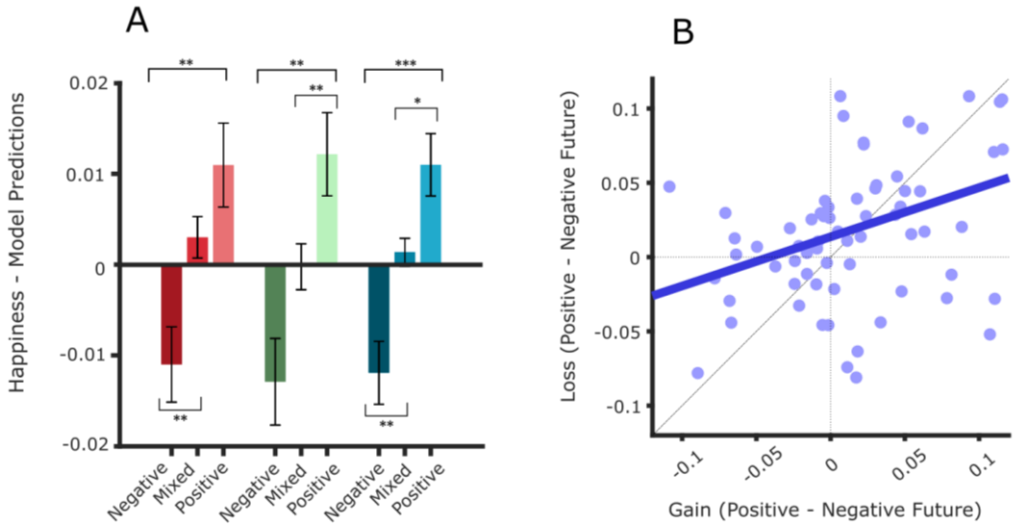


Figure 4-7 Influence of the Future on Happiness Ratings

(A) The difference in Happiness ratings and the generated model predictions (residual happiness) was significantly different between positive and negative futures for all trials (blue bars), and when split by loss trials (red bars) and gain trials (green bars). Residual happiness was significantly positive when the future was positive for all trials and in gain trials. The residual happiness was significantly negative for negative futures for all trials, and loss trials. An average residuals value of 0.02 would mean that happiness ratings are 2% greater on average under a given future prospect. **(B)** The difference in residual happiness for positive and negative futures was significantly positive in both loss and gain trials, and the differences also correlated between them.

[*p < 0.05, ** p < 0.01, *** p < 0.001]

4.4.7 Influence of the Positive and Negative Futures on Decision-making Increases with Proximity

Having shown that the valence of the future prospects influenced happiness ratings, we next we looked at whether positive or negative futures effected decision-making. We used the

Prospect Theory model to regress out the influence of magnitude and probability on choice to explore if the future blocks had any influence on the decision-making. Positive choice residuals means that the model under predicted the risky option would be chosen. Negative choice residuals meant that the model under predicted the safe option would be chosen. The greater the residuals, the more the model under predicted the choice made. We first tested whether future prospects had an effect on decision-making overall, by comparing the choice residuals between positive and negative futures. These were not significantly different overall (0.015 ± 0.068, $z = 1.69$, $p = 0.091$), nor when split between loss trials (0.012 ± 0.097, $z = 1.58$, $p = 0.115$), or gain trials (0.0174 ± 0.102, $z = 0.809$, $p = 0.418$). The mean difference in choice residuals for loss trials was positively correlated with the mean difference in happiness residuals ($\rho(76) = 0.246$, $p = 0.029$), but not in gain trials ($\rho(76) = -0.030$, $p = 0.796$) nor overall $\rho(76) = 0.115$, $p = 0.316$). This may suggest that in loss trial's those whose mood was more affected by future prospects also had changed their decision-making based upon future prospects. However as we do not see this effect in gain trials we do not consider it robust enough for further analysis.

Unlike happiness ratings, decision-making is sampled multiple times within one block of trials. We leveraged this to explore whether we could see changes in decision-making at the beginning or end of a block of trials. We found that the choice residuals were significantly different between positive and negative futures on the 3rd trial of each block across all trials, (0.0454 ± 0.157, $z = 2.607$, $p = 0.0091$; Figure 4-8ai), loss trials (0.0594 ± 0.222, $z = 2.56$, $p = 0.0105$; Figure 4-8aii), but not gain trials (0.0228 ± 0.27, $z = 0.516$, $p = 0.606$; Figure 4-8aiii). All effects seen were positive, suggesting that positive futures increased risky choices where as negative futures increased safe choices. We did not see this effect on the 1st trial of each block (all trials: 0.011 ± 0.097, $z = 1.25$, $p = 0.210$; loss trials: 0.014 ± 0.146, $z = 1.08$, $p = 0.281$; gain trials: 0.0079 ± 0.133, $z = 0.167$, $p = 0.868$).

We next tested whether any influence of the future on decision-making increased throughout the block. We fit general linear models to the choice residuals as a function of the trial number separately for all combinations of future valence and trial valence (Figure 4-8b). A positive slope for the influence of trial number on future prospects would indicate an increase in gambling as the future was approached. There was a significant difference between the influence of future on trial number slopes across all trials when comparing positive and negative futures (0.0288 ± 0.084, $z = 2.95$, $p = 0.0032$). We then tested this separately for trial types and found a significant

contrast in loss trials (0.0393 ± 0.115, z = -2.87, p = 0.0042), but not in gain trials (0.0206 ± 0.129, z= 1.75, p = 0.080). As a robustness test we split the data in half and tested whether we could observe the difference in positive and negative futures in both halves (first half: 0.0344 ± 0.104, z = -2.77, p = 0.0056; second half: 0.0239 ± 0.123 z = -1.99, p = 0.047).

Finally, we explored whether we could see effects in negative and positive futures separately. To test the significance of these separately, we fit another set of general linear models which included a slope that captured the influence of decision-making changing across the block under all future conditions, to remove any variance associated with this. Choice residuals showed a positive slope for positive futures across all trials (0.010 ± 0.032, z = 2.68, p = 0.0074), but the slope for negative futures was not significant (0.0049 ± 0.028 , z = -1.41, p = 0.159). For loss trials only, we found the positive and negative futures were not significant (positive future: 0.0074 ± 0.041, z = 1.606, p = 0.108; negative future: -0.0082 ± 0.041, z = -1.736, p = 0.083), and only positive futures in gain trials (positive future: 0.0125 ± 0.050, z = 2.28,p = 0.022; negative future: -0.0027 ± 0.0383, z = -0.535, p = 0.592).

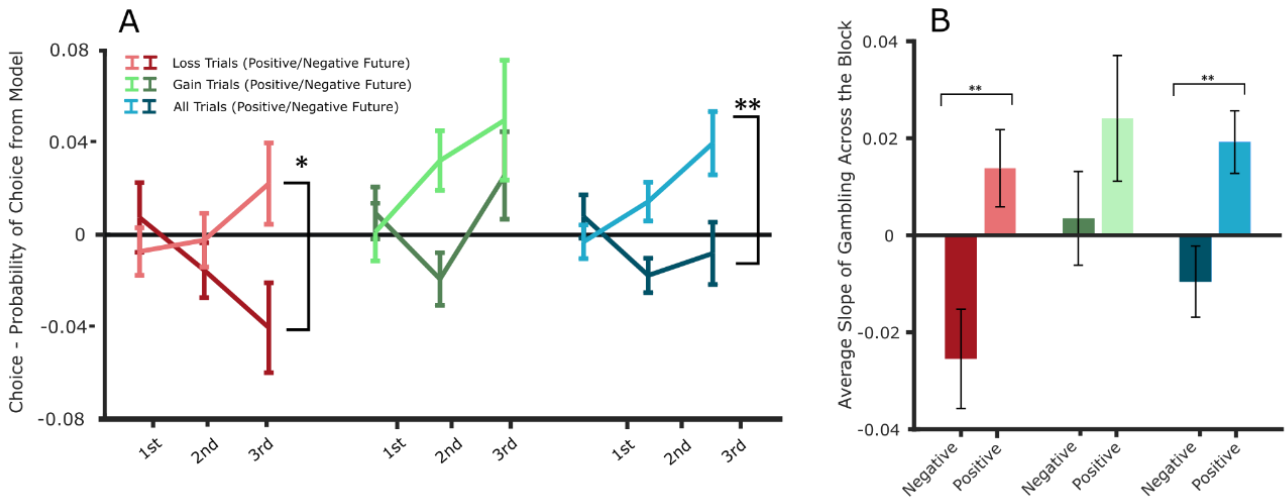


Figure 4-8 Influence of the Future on Decision-Making

(A) Choice Residuals (y-axis) are the difference in the choice made by the participant (1 for riskier choice, zero for safe choice) and the probability of making the risky choice from the Prospect Theory style model. Positive choice residuals means that the model under predicted the risky option would be chosen. Negative choice residuals meant that the model under predicted the safe option would be chosen. The greater the residuals, the more the model under predicted the choice made. The choice residuals are shown for loss trials (red error bars), gain trials (green error bars), and all trials (blue error bars). Darker colours indicate negative futures, and lighter colours indicate positive futures. Choice residuals were significantly different in the 3rd trial of each block in loss trials, and across all trials. **(B)** Slopes were fitted to the choice residuals and trial position which were significantly different between positive and negative futures for loss trials, and across all trials. We did not appear to see any effects of trial number in gain trials, this appears to be driven by the 3rd trial when the future is negative being positive. We can speculate that is may be due to an additional gain trial being treated as a positive surprise.

*p < 0.05, ** p < 0.01, *** p < 0.001

4.5 Discussion

In a novel paradigm where participants were aware of the valence of near future trials while making decisions in current trials, we demonstrate that moment-to-moment changes in densely sampled happiness ratings relate to both short and long term prospects. In the short term, using a computational model of affective dynamics we find evidence for an affective negativity bias, where happiness ratings (modelled as prediction errors) were more extreme for loss trials then gain trials. We also show that the valence of longer term prospects affects affective ratings and tendency to make risky decisions: positive futures are associated with higher affective ratings and increased risky taking; negative futures are associated with lower affective ratings and decreased risk taking. Where possible, for affective biases we looked to interpret results which contrast between gain and loss trials, and positive and negative futures. This is due to these statistics being more robust to individual participant's differential usage of the scale, but limits the interpretations we can make about the independent influence of individual trial type of valence of future.

Negativity biases have been observed across multiple domains (learning from reinforcement: Gershman, (2015); increased attention: Yechiam & Hochman (2013); mixed prospect choices: Charpentier et al (2016)). Here we show robust evidence for its presence in affective responses, as observed using densely sampled ratings. Prior work has suggested a utility of such a bias in learning, where it allows for the avoidance of future negative surprises (Michely et al., 2020; Pulcu & Browning, 2017). Consistent with this view, we see a relationship between a greater negativity bias and increased risk sensitivity for losses in a prospect theory model where risk sensitivities for gains and losses are parametrised separately. We also see this effect in a model where risk sensitivity is captured as one parameter, which suggests the negativity bias may be associated with increased sensitivity to magnitudes overall (higher values indicating relatively greater risk taking for gains, and safe choices for losses). However as we found the parameters in the prospect theory style model were highly correlated, we believe these results should be replicated before strong inferences can be made.

We also found that future prospects were associated with affective responses in the direction of their valence; participants were happier when the future was positive, compared to when the future was negative. This effect was present in both gain and loss trials. We examined the influence of positive and negative futures separately on affective ratings, and found significant

effects in both cases. When further split by trial type, we only found significant effects in loss trials and negative futures, and gain trials and positive futures. However this may be due to statistical power being sufficiently reduced; one future for each trial type corresponds to less than 10 ratings per participant. This observation could suggest that dwelling on the nature of the future has an affective property, which may contribute to the observation that optimism is associated with greater life outcomes.

The size of the effects for the influence of future prospects on affective ratings and decision-making were robust but relatively small. For example, the difference between happiness ratings between positive and negative futures was on average approximately 2% of the entire scale, compared to the average difference in prediction error weights which was 12% of the scale. This is perhaps not-surprising due to there being no task associated with the presence of the future blocks, which limits the attention required to them. Studies exploring anticipation typically focus on the anticipated event entirely, with no interweaving task related events (e.g. Iigaya et al., 2016). In contrast, here we looked to design a novel task which capitalised on a more natural tendency to dwell on the future, and explore how it interacts with current events (i.e. decisions).

Regarding the influence of future prospects on decisions, we were surprised to not find an effect of the future on decision-making when including all trials, particularly as many participants informally reported that they did change their decisions based on the future blocks. Intriguingly, we found that when we looked across the whole block, we could see the risk taking increased as participants approached the end of the block with a positive future, compared to as they approached a negative future where risk taking appeared to reduce. A change in decision-making as distance to the future decreases is consistent with neuroimaging work on anticipation, where a ramping signal in the value encoding areas of the brain is observed, following surprising news about a future reward to the receipt of the reward itself (Iigaya et al., 2019). The anticipation of positive and negative futures may drive a similar neural mechanism which leads to risky gain prospects being perceived as more attractive. Studies exploring dread have suggested a similar increase in affective intensity and neural activation as a painful shock stimulus approaches in time (however in pain related areas, and not value encoding: Berns et al., 2006).

A lower baseline mood being associated with increased depressive symptoms has been observed in studies using dense affective sampling (Blain & Rutledge, 2020; Rutledge et al., 2017). We replicated this result here, where the baseline mood parameter is negatively correlated with the Mood and Anxiety Symptom Questionnaire. This questionnaire measures self-reported anhedonia symptoms, however we prescribe caution at interpreting this result to be specifically associated with anhedonia, as we also saw strong correlations between average happiness and other measures, including anxiety (State Trait Anxiety inventory) and depressive symptoms (Beck's Depression Inventory II). In addition to lower baselines being associated with greater MASQ scores, we also saw that the affective negativity bias was lower in greater MASQ scores. We confirmed the independence of these results using a partial correlation test. A reduced negativity bias, if linked to reduced avoidance behaviour, could suggest a mechanism by which mood related symptoms are worsened by being unaided by negative affect to avoid more future negative affect. Other studies have shown evidence for attenuated neural prediction error signals in patients with major depressive disorder (e.g. Kumar et al., 2008). Major depressive disorder, and anhedonia more specifically, have been related to reward sensitivity, as opposed to attenuated learning rates (as indexed by the MASQ questionnaire: Huys et al., 2013). It may be that differences in mood related prediction errors are more appropriately described as differences in reward sensitivity, and understood distinctly from learning related prediction errors. It is notable in this study; we did not find strong evidence for a prediction error model better describing the happiness ratings than a model using the value of the outcome alone.

Prior studies using dense affective sampling have not observed any differences between prediction error weights in affective dynamics models and depressive symptoms (Blain & Rutledge, 2020; Rutledge et al., 2017). These experimental tasks use either all trials of the same type (gain trials; Blain & Rutledge, 2020) or pseudo-randomise the order of different trial types (gain, loss and mixed; Rutledge et al., 2017). It may be that we were able to observe prediction error relevant variance with symptoms for two experimental reasons. Firstly, we leveraged the difference in prediction errors between two trial types, which partially normalises for each participants individual scale usage. Secondly, it may be that by structuring trials into blocks of gain and loss trials we increase the relevant emotional salience of each trial type. This may

correspond to the difference in comparing a surprising good or bad event, with the difference in a good or bad day.

Altogether, we have shown that expectations in the short and longer term can influence affective state, estimated from happiness ratings, and decision-making. This approach of estimating both affective and decision biases within the same study extends our understanding of how both these biases interact, perhaps leading to a better understanding of how low mood and mood disorders affect our choices and perception of the world. Future work should focus on optimising computational models for both happiness ratings and decision-making, perhaps developing model based solutions for estimating latent variables that affect both mood and choice.

4.6 Extended Analyses & Results

4.6.1 Models of Affective Dynamics

4.6.1.1 The Model Space

To explore the influence of task events on fluctuations in happiness ratings, we considered 2 families of affective dynamics models. The first family, referred to as *Expectation Models,* used *expected values* and *prediction errors*. The second family, referred to as *Rewards Models,* used *outcomes* only. Within each family we tested two model versions: one with a shared baseline between gains and losses, and one which modelled signed outcomes and zero outcomes separately (Table 4-1 *Affective Dynamics Candidate Models*). All models had separate affective dynamic weights for gains and losses, and shared a forgetting factor (τ) between gains and losses. The reason for this was in previous studies, is has been shown that the number of trials which influence the current rating exceeds the lengths of the blocks in this study (e.g. Blain & Rutledge, 2020; Rutledge et al., 2014).

4.6.1.2 Model Selected for Analysis in the Main Text

The following sections examine the models for complexity and generalizability trade-off (BIC Model comparison), split-half reliability of parameter estimates, and model identifiability. We conclude from these tests that the *Expectations Model (3b)* is the best for analysing differences in happiness ratings. Importantly, we do not conclude that our model comparison is able to demonstrate the affective responses are *driven by* prediction errors, and expectations, compared to other candidate models.

4.6.1.3 *BIC Model Comparison*

First, we performed a formal model comparison using Bayesian Information Criterion, as has been standard for affective dynamics models (e.g. Blain & Rutledge, 2020; Rutledge et al., 2014). The best fitting model according to this criterion was *Rewards Wins & Losses (2b)* (Table 4-1; Figure 4-9a) with a mean ΔBIC of 6.585 compared to the best fitting model of the Expectations family (Figure 4-9b&c).

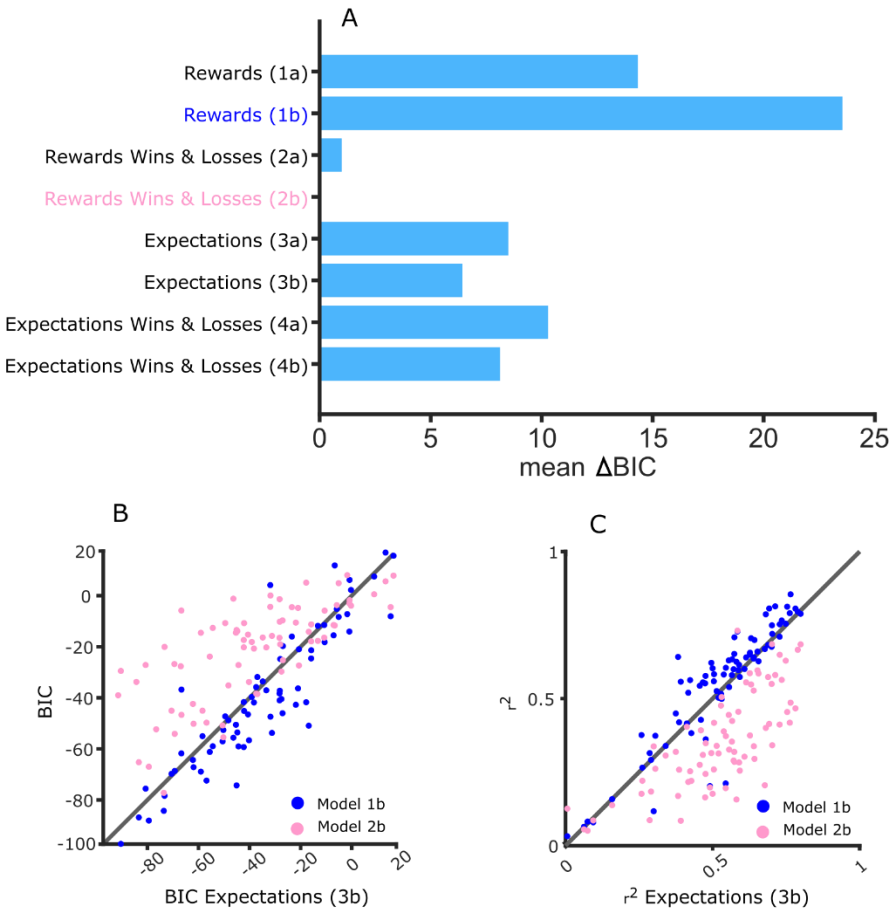


Figure 4-9 Bayesian Model Comparison for Affective Dynamics Models

(A) Bayesian Information Criteria for comparing the candidate models. The x-axis shows the average difference between the BIC for the model with the lowest BIC (Rewards Win & Losses (2b)) and each other candidate model. The models are fit to z-scored happiness ratings, in order to not bias the model comparison towards participants who use a lesser degree of the happiness rating scale. **(B)** The difference in each participants BIC scores between Expectations model (3b) and Rewards Wins & Losses (2b) (pink data points), and Rewards (1b) (blue data points). The difference in r^2 for the same models is shown in **C)**. All models were fit using positive bounds.

Table 4-1 Affective Dynamics Candidate Models

Model Name	Affective Dynamics	Constant	Baseline
Rewards (1a)	Outcome (2) [-1 0.5 2]		(2) [-0 0.5 1]
Rewards, One Baseline (1b)			(1)
Rewards Wins & Losses (2a)	Signed Outcomes (2) [-1 0.5 2]	Unsigned Outcome (2) [-1 0.5 2]	(2) [-0 0.5 1]
Rewards Wins & Losses, One Baseline (2b)			(1)
Expectations (3a)	Expected Value (2) [-1 0.5 2] Prediction Errors (2) [-1 0.5 2]		(2) [-0 0.5 1]
Expectations, One Baseline (3b)			(1)
Expectations, Split Baseline (3c)	Expected Value (1) Prediction Error (1)		(2)
Expectations Wins & Losses (4a)	Expected Value (2) [-1 0.5 2] Prediction Errors for Wins (2) [-1 0.5 2] Prediction Errors for Losses (2) [-1 0.5 2]		(2) [-0 0.5 1]
Expectations Wins & Losses, One Baseline (4b)			(1)

Here we show two families of candidate models Rewards Win & Losses Models (2) and Expectation Models (3). The round bracketed number indicates the amount of parameters of that kind in the model, where (2) indicates a separate parameter for gain and loss trials and (1) indicates they are shared. The square brackets indicate the [lower bound, starting point, upper bound] for each parameter when it is fit with fmincon (MATLAB). Where models are described as having 'positive bounds' in the main text, the affective dynamics parameters have a lower bound of 0. Each model is presented in its full version, indicated by 'a' and any subsequent versions below only provide numbers of parameter if they change from the 'a' model. Every model also includes a forgetting factor (τ) which is constrained at [0 0.1 1]. Expectations (3c) is used for simulations and robustness checks.

4.6.1.4 Assessing Model Reliability

Reliability is a paramount issue for models used in computational psychiatry. In order for parameters in computational models to be compared to self-reported questionnaire measures (or any other measures taken from real life e.g. a diagnosis) they need to be estimated with a high degree of internal consistency (e.g. test retest reliability, or split-half reliability) (Brown et al., 2020; Browning et al., 2020; Paulus et al., 2016). This poses a difficult question for models of mood, where emotional engagement may vary throughout the task.

With this in mind, we tested the split-half reliability of all candidate models. This procedure involved fitting each participant's first and second half of their data separately (39 blocks per sample, ~98 trials) and correlating the estimated parameters fit to the data from the first half, to those fit to the second half. We also tested two different boundary constraints to see if it improved reliability. In the first set of constraints (unconstrained) the affective reactivity parameters were allowed to fit under zero, in the second set (positive bounds) the affective reactivity parameters were constrained to be above zero. The reason for this was because affective dynamic parameters below zero are difficult to interpret, for example a negative prediction error weight in gain trials would suggest that an individual was happier when they lost points. Despite this, previous publications using affective dynamics models have not constrained the parameters to be above zero (e.g. Blain & Rutledge, 2020; Rutledge et al., 2014). We test here whether formalising our expectation that constraining these weights to be positive improves split-half reliability and should be considered an axiom for future instantiations of affective dynamics models. In both sets of boundary constraints the baseline mood parameter (C) and decay parameter (τ) were constrained to exceed zero.

We tested whether constraining all parameters to be positive improved reliability of parameter estimates between the first and second half of the experiment. Broadly, all the models showed good split-half reliability, where most parameter estimates (with the exception of τ) correlated between the first and second half (all the statistics for these analyses are presented in Table 4-2 Split-Half *Reliability*). In almost all models and parameters, the split-half reliability improved with the inclusion of positive bounds.

Table 4-2 Split-Half Reliability

Model Names	**Parameter Reliability (R)**						**r^2**
Rewards Wins and Losses (2)	UnSign		wSign		τ	C	
(2a) Unconstrained	.467***	.527***	.331**	.409***	.28*	.320* .108^{ns}	.581
(2a) Positive bounds	.543***	.692***	.467***	.645***	.131^{ns}	.461*** .387***	.577
(2b) One Baseline	.367***	.621***	.478***	.540***	.140^{ns}	.341***	.554
Expectations (3)	wEV		wPE		τ	C	
(3a) Unconstrained	.172^{ns}	.189	.669***	.693***	.139^{ns}	.343*** .268**	.548
(3a) Positive bounds	.303**	.223*	.673***	.722***	**.308****	.440** .375***	.539
(3b) One Baseline	.41***	.218^{ns}	.701***	.698***	.190^{ns}	.450***	.527

Each column shows first the loss trials estimate, and second the gain trial estimate. If there is only one value, the parameter is fit to both trial types.

[ns non-significant, * p < 0.05, ** p < 0.01, *** p < 0.001]

4.6.1.5 Parameter Correlations

In each of the four candidate models there was a high degree of correlation between baseline parameters for each trial frame and their affective dynamics parameters (Figure 4-10). In the *Rewards Model (2a)* the baseline for gains negatively correlates with positively signed outcomes ($r(76) = -0.504$, $p < 0.0001$), and positively correlates with unsigned outcomes ($r(76) = 0.306$, $p = 0.0065$; Figure 4-10a). The inverse can be seen in the loss trials, where the baseline positively correlates with negatively signed outcomes ($r(76) = 0.336$, $p = 0.0027$), and negatively correlates with unsigned outcomes ($r(76) = -0.530$, $p < 0.0001$). The same direction of relationships is observed in the *Expectations Model (3a)*, where the baseline for gains negatively correlates with positively signed affective dynamics (expected value, $r(76) = -0.591$, $p < 0.0001$) and the baseline for loss trials positively correlates with negatively signed affective dynamics (expected value, $r(76) = 0.680$, $p < 0.0001$; ; Figure 4-10b). Shared variance between baselines and dynamics presents an issue for comparing between gain and loss trials, as variance may be either masked by this trade-off, or artefacts occur causing erroneous significant differences between trial types discussed in Section 6.3.2 *Can we Identify both Mood Baselines and Affective* Responses?).

We next tested whether we could reduce the covariance between the baseline being fit to both gain and loss trials together. In the *Rewards Model (2b)* the covariance reduced only marginally; the baseline parameter negatively correlated with affective dynamics for loss trials (unsigned outcome: $r(76)$ -0.374, $p = 0.0007$; signed outcome: $r(76) = 0.278$, $p = 0.0139$), and gain trials (unsigned outcome: $r(76) = 0.240$, $p = 0.0347$; signed outcome, $r(76) = -0.469$, $p < 0.0001$; Figure 4-10c). However fitting a shared baseline reduced covariance a considerable degree in the *Expectations Model (3b)*; the expected value weights for loss trials $r(76) = 0.376$, $p = 0.0008$), and gain trials $r(76) = -0.427$, $p = 0.0001$), and also the prediction error weights reduced (loss trials, $r(76) = -0.0257$, $p = 0.823$; gain trials, $r(76) = 0.0035$, $p = 0.98$; Figure 4-10d).

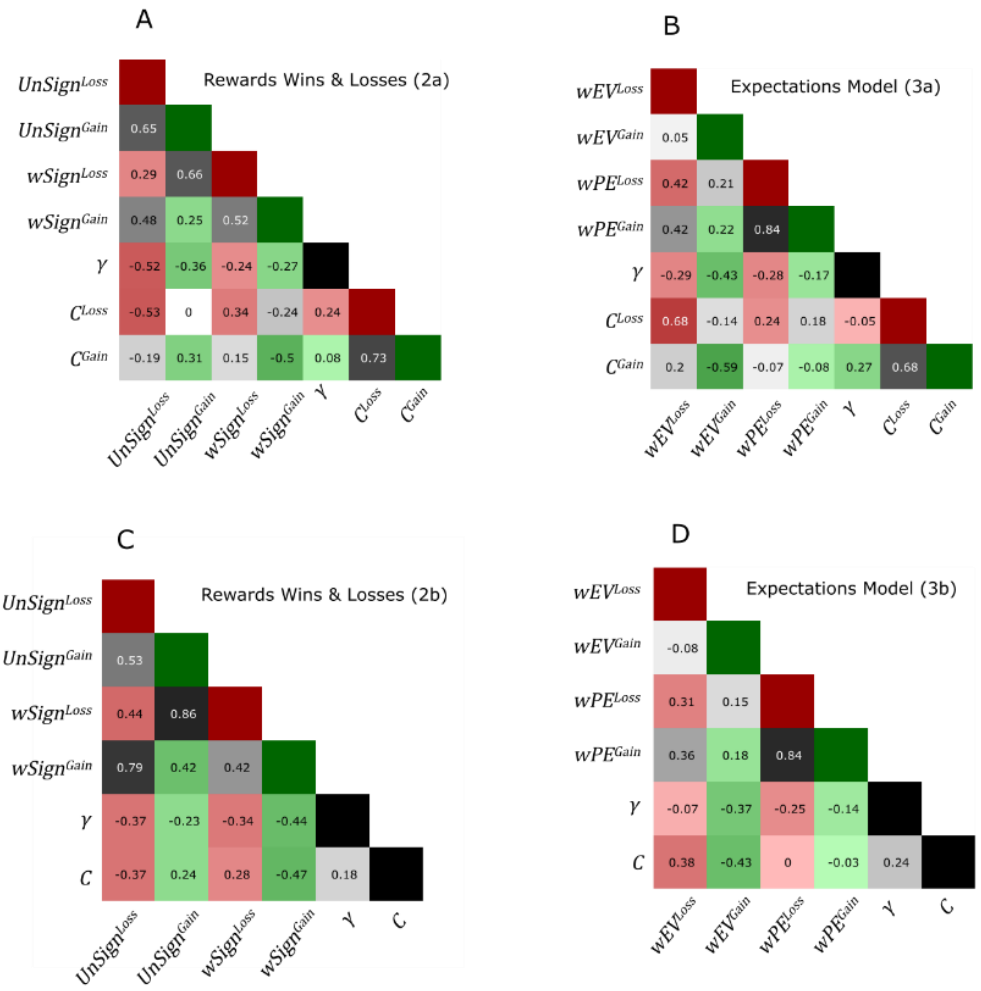


Figure 4-10 Correlation Matrices for Candidate Models

The Pearson's r values for each parameter compared to each other parameter in the same model are shown in for **(A)** *Rewards Wins and Losses (2a),* **(B)** *Expectations Model (2b)*, **(C)** *Rewards Wins and Losses (2b)*, and **(D)** *Expectations Model (3b)*. These models were all fit with positive bounds. The r values are written at the intersection of each parameter pair, with the saturation of the colour also indicating the value (saturated colours indicate lower absolute values of r). The maximum unsaturated tone is shown in the diagonal squares where the r value is 1. An r value of zero is indicated by a white tone. Red Squares indicate any correlations where at least one parameter is fit to only loss trials (e.g. C and wPE^{loss}), or (wEV^{loss} and wPE^{loss}). Green Squares indicate any correlations where at least one parameter is fit to only gain trials (e.g. C and wPE^{gain}), or (wEV^{gain} and wPE^{gain}). Grey Squares indicate correlations where parameters are fit to both gain and loss trials equally (e.g. C and τ), or are fit to one set of trials and not another (e.g. wEV^{gain} and wEV^{loss}).

4.6.2 A Model Recovery Technique for Evaluating Generative Performance in Continuous Data

Generative performance is how well the model predictions capture data features of interest, and their ability to make unique predictions compared to other candidate models (Palminteri et al., 2017). One method of testing generative performance is by performing model recovery; fitting a series of candidate models to the generated data of one model of interest (test model), and confirming that a new model comparison prefers the test model over the other candidate models. If the test model makes unique predictions, then it should be preferred by BIC to other candidate models. If one of other candidate models is preferred to the test model, this suggests that it makes similar predictions to the test model and thus the theories instantiated by these models cannot be described as falsifiable (in this case).

To the authors knowledge there have not been attempts in the literature to explore model recovery for affective ratings. Model recovery is a popular technique when comparing models that make binary choices, as the *generative data* is probability estimates that are converted into choices – thus is clear to see how two different candidate decision models may make different probability estimates (e.g. the likelihood of choosing a particular option as 0.6 or 0.95) but have the same behaviour (i.e. choosing the option). When predicting point estimates of mood, the *generated data* requires no conversion to simulate the behaviour. This means that it is very easy to recover any model due to the high level of precision of the predicted data. Here a novel model recovery method was used which involved adding noise to the generative data of the test model and refitting the candidate models. The noise term was the standard deviation of the residuals of the observed data and the generated data. This allows for amount of noise added to be both proportional to the participants use of the scale and appropriate for the noisiness of the model (akin to an inverse temperature parameter in a decision model). If a model makes unique predictions to other candidate models it tends to be recoverable. If a model is not recoverable compared to another model it suggests it is not possible to confidently distinguish between the two models.

We looked to test the identifiability of the *Expectations Model (3b)*. We first confirmed the noise term correlated with the r^2 of the model (when normalized by the standard deviation of each participants ratings: $\rho(76) = -0.923$, $p < 0.0001$), demonstrating that where the model explained more of the variance in the happiness ratings, the simulated noise was lower. The model

recovery did prefer the *Expectations Model (3b)* in terms of BIC Figure 4-11a) However the ΔBIC to the other candidate models was similar in degree to BIC ratings shown in Figure 4-9. Next we looked at the r^2 measures of fit for gain and loss trials separately for the *Expectations Model (3b)* and *Rewards Model (2b)*. If models are identifiable, r^2 should be greater for the test model than the alternative candidate model. For loss trials both the *Expectations model (3b)* and the *Rewards Model (2b)* had greater r2 for their own recovered model compared to the candidate (*Expectations:* r^2 for *Expectations* 0.498 ± 0.222; r^2 for *Rewards* 0.448 ± 0.222, z = 3.086 ,p = 0.0020; *Rewards*: r^2 for *Rewards* 0.531 ± 0.232; r^2 for *Expectations* 0.468 ± 0.2104, z = 3.99 ,p < 0.0001; Figure 4-11bi). However, in the gain trials the only the Rewards model had greater r^2 (*Expectations:* r^2 for *Expectations* 0.490 ± 0.209; r^2 for *Rewards* 0.471 ± 0.209, z = 1.629 ,p = 0.104; *Rewards*: r^2 for *Rewards* 0.547 ± 0.223; r^2 for *Expectations* 0.506 ± 0.223, z = 2.58 ,p = 0.0098; Figure 4-11bii).

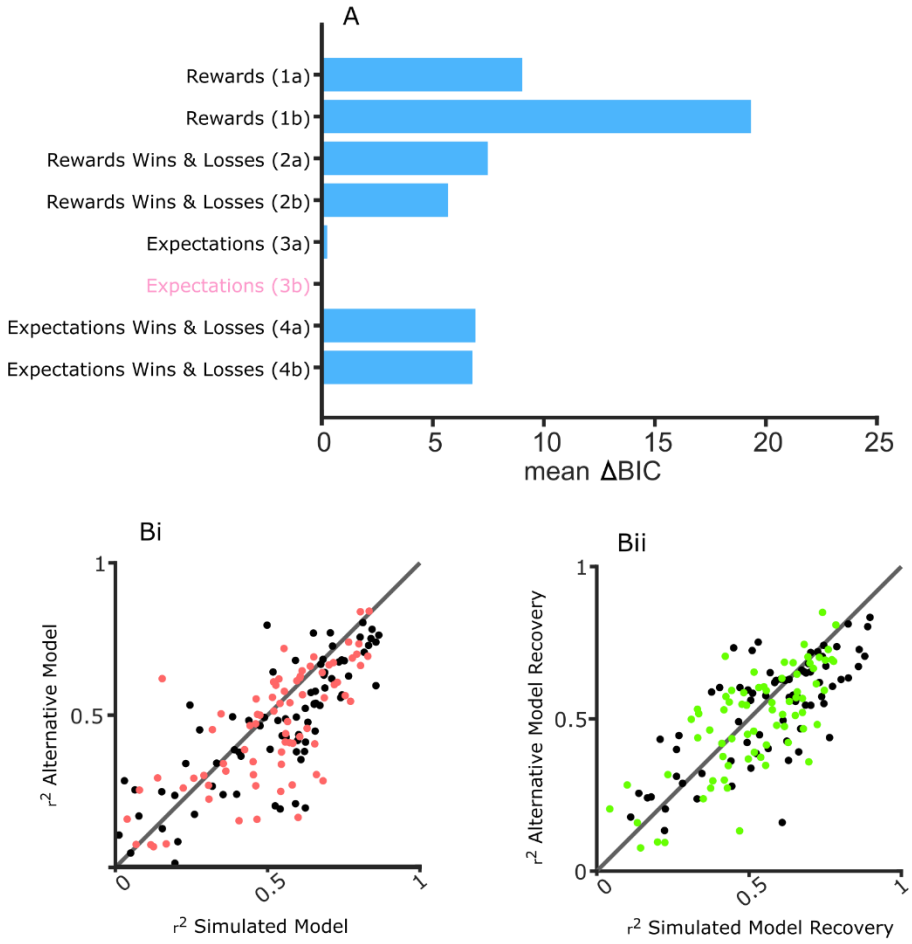


Figure 4-11 Affective Dynamics Model Recovery

(A) Bayesian Information Criteria for comparing the recovered *Expectations Model (3b)* with all other candidate models. The x-axis shows the average difference between the BIC for the model with the lowest BIC (*Expectations Model (3b)*) and each other candidate model. All models were fit using positive bounds. **(B)** We simulated the *Expectations Model (3b)* and the *Rewards Wins & Losses Model (2b)* with noise and recovered each model with each of the models. The r^2 of the simulated model is on the x-axis, and the r^2 of the alternative model is on the y axis. The coloured data points indicate simulations from the *Expectations Model (3b)*. The r^2 for loss trials is shown in **(Bi)** and **(Bii)** shows the gain trials. Data points below the origin indicate that the trials were best fit in the simulated model compared to the alternative model. Loss trials both had greater r^2 in the simulated model, however gain trials only had greater r^2 when the *Rewards Wins and Losses Model (2b)* was simulated.

4.6.3 Testing the Robustness of the Negativity Bias to Artefacts

4.6.3.1 *The Asymmetry of Prediction Errors is not caused by Variance in Mood Baselines.*

A mood baseline deviating from the centre of the scale may cause artefacts in the estimates of affective reactivity weights, due to asymmetries in the amount of space on the scale available above and below the baseline. When $C > 0.5$ positive responses may be truncated compared to negative responses. Here, we test whether variance inferred as a negativity bias could be caused by this.

In the *Expectations model (3b)* the difference in prediction error weights did not correlate with C ($\rho(76) = 0.0219$, $p = 0.849$), however the difference in expected value weights positively correlated ($\rho(76) = 0.475$, $p < 0.0001$). This correlation suggests that when C is higher, wEV^{loss} is estimated as greater than wEV^{gain}; this may be due to there being more room for negative expectations to be reported to be than positive expectations.

4.6.3.2 *Could the Appearance of the Negativity Bias be caused by differences in Baseline Mood for Gains and Losses?*

It is not implausible to suspect that the *negativity bias* could be an artefact of different mood baselines for gain and loss trials due to parameter trade off. To assess this concern, we fit a version of the *Expectations Model (3b)* with separate baseline estimates (C) for gain and loss trials, but shared affective dynamics betas (*Expectations Model 3c;* r^2 : mean 0.493; median, 0.536; standard deviation, 0.186). In this new model, the baselines were not significantly different ($z = 1.27$, $p = 0.203$), suggesting that separable baselines were not adequate to describe the difference in prediction error parameter estimates.

To substantiate further whether this concern was plausible, we explored whether a difference in baselines between gains and losses *could* be fit as a negativity bias. In order to perform this analysis we generated a new set of happiness ratings where the loss baseline was lower than the gain baseline. We did this by taking the generated ratings from a model where all the parameters were shared between gain and loss trials, and subtracting one standard deviation of the baseline estimate (0.121) from all of the happiness ratings which were sampled at the end of loss trials. We confirmed we had successfully generated separable baselines by fitting the same model, but a separate baseline for gain and loss trials to these new ratings. We

confirmed the baseline estimates were significantly different ($z = 7.12$, $p < 0.0001$). Next we fit the *Expectations Model (3b)* to the new happiness ratings, and compared the difference in affective dynamics weights. We found that the model estimated greater expected value weights for loss trials than gains (mean difference: 0.111 ± 0.241, $z = 3.98$, $p < 0.0001$), and a minor increase in the prediction error for gains compared to losses (mean difference: -0.023 ± 0.081, $z = -2.253$, $p = 0.024$). This is consistent with our analysis of covariance, where we saw high correlations between expected value and baseline parameters.

4.6.3.3 Removing Happiness Ratings at the Boundaries does not affect the Negativity Bias

We next looked to explore if we could still observe the *negativity bias* if we removed ratings that may be vulnerable to being artificially bounded by the scale. We fit the *Expectations Model (3b)* to only the happiness ratings that where within 0.1 and 0.9 of the scale. Marginally more happiness ratings at the end of loss blocks were removed than in gain blocks (loss trials: 4.795 + 6.205; gain trials 4.167 + 5.540,$z = 1.96$, $p = 0.050$). In the model fit to the subset of ratings, the difference in affective dynamics weights were still significantly greater in loss trials than in gains (expected value weights: wEV^{loss} , 0.183 + 0.197, wEV^{gain}, 0.058+ 0.078, $z = 4.46$, $p < 0.0001$; prediction error weights: wPE^{loss} , 0.507+ 0.414, wPE^{gain}, 0.352+ 0.229, $z = 4.78$, $p < 0.0001$).

4.6.4 Comparing Models of Choice

When developing the design matrix for this study we looked to balance emotional salience of choices and outcomes, and attention paid to the future prospects. To this end we had both choices associated with a probability, to maximize the number of surprising outcomes, and used a constant decision option (i.e. the safe option) so decision-making was not cognitively overtaxing, allowing for attention to be paid to the future blocks. When fitting prospect theory style models of choice to the data, we found that the design matrix was not optimized for arbitrating between models and identifying parameters. Here, we briefly present some basic model comparison, but we do not believe these models should be relied upon for strong inferences about connections with affective dynamics.

We fit four prospect theory style models with different probability weighting functions. Probability weighting functions distort the effect of objective probabilities on subjective utilities

to describe participants choices 'overweighing' (treating a probability as greater than it is) and 'underweighting choices' (treating a probability as smaller than it is). The four probability weighting functions we tested were: Tversky function (Tversky & Kahneman, 1992; Figure 4-12), Lattimore function (Lattimore et al., 1992) and two versions of the Prelec function (Prelec, 1998).

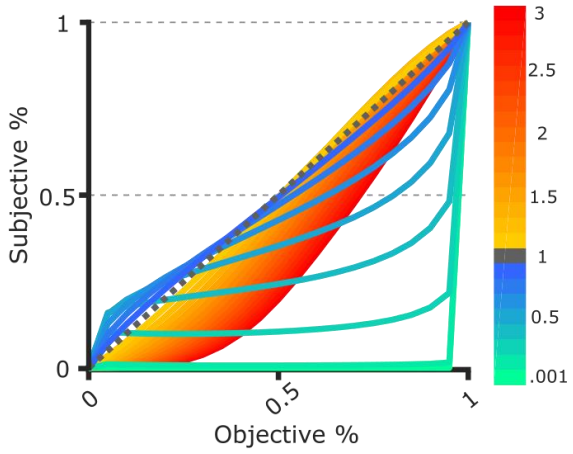


Figure 4-12 Tversky & Kahneman Probability Weighting function

The probability weighting function as included in Tversky & Kahneman (1992) includes one free parameter (γ) which converts objective to subjective probability. The colour bar on the right indicates values of γ. When γ <1 low probabilities are over weighted and moderate to high probabilities are underweighted. When γ > 1 lower probabilities are underweighted and higher probabilities are marginally overweighed. Average study values of γ tend to be < 1.

Each of these models fit the data well, where the average pseudo r^2 values all exceeded 0.50 (Tversky: 0.545 ± 0.213; Lattimore, 0.555 ± 0.214; Prelec Delta, 0.555 ± 0.214; Prelec, 0.542 ± 0.215). A BIC comparison preferred the model with the Tversky Weighting Function (Figure 4-13a). However when looked to recover these models and we found they all preferred the Tversky model (mean difference in BIC between Tversky model: Lattimore ΔBIC = 3.93; Prelec Delta, ΔBIC = 0.260; Prelec, ΔBIC = 5.85). The two best fitting models according to BIC both showed considerable parameter covariance (Figure 4-13b & c).

It is unlikely that these issues seen here were due to stochasticity of decision-making, as average risk taking was highly correlated between the first and second half of the experiment (loss trials, $\rho(76) = 0.784$, $p < 0.0001$; gain trials, $\rho(76) = 0.589$, $p < 0.0001$). Stochasticity parameters for the model were relatively high, suggesting behaviour was relatively deterministic based on other parameter estimates (μ^{Loss}: 3.601 ± 5.462; μ^{Gain} :5.46 ± 6.483; Figure 1-2 Softmax Rule Simulation).

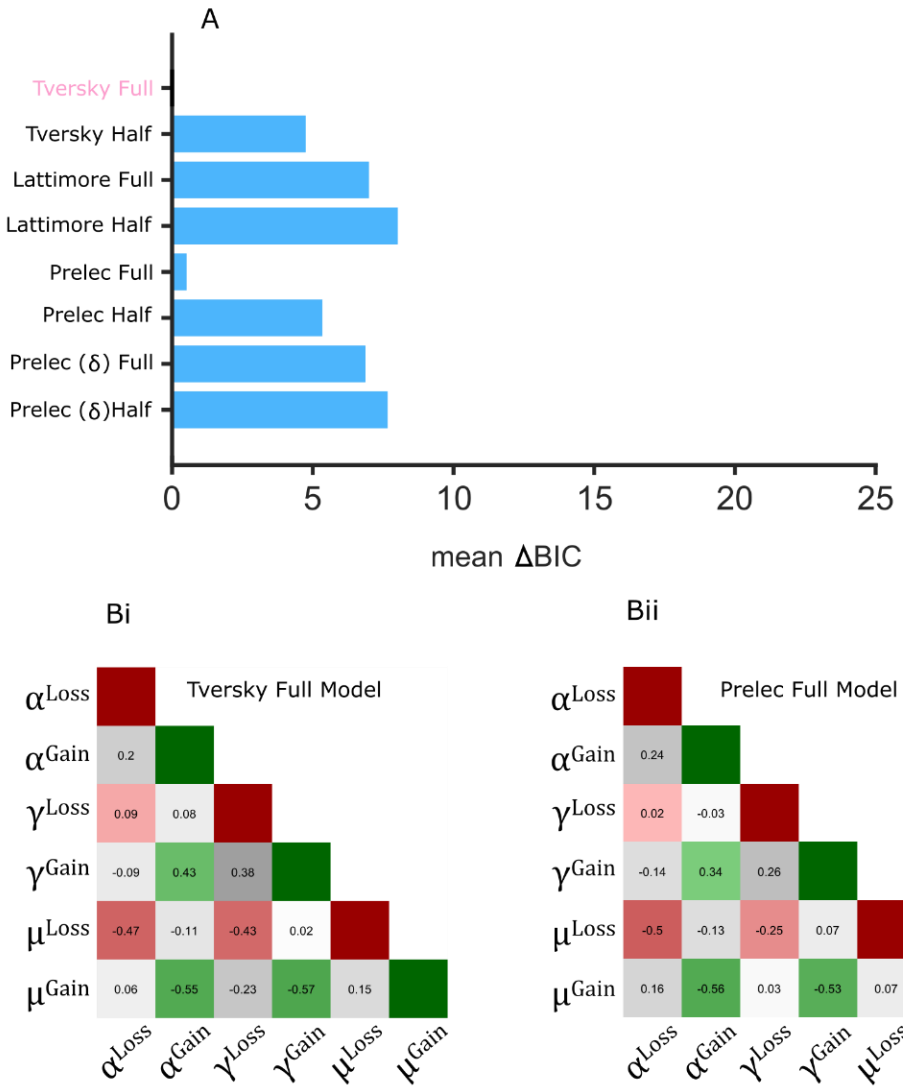


Figure 4-13 Models of Choice

(A) Bayesian Information Criteria for comparing the candidate models. The x axis shows the average difference between the BIC for the model with the lowest BIC (Tversky & Kahneman Full Model) and each other candidate model. Each model is compared in its 'full' version: where there is a separate parameter fit for gain and loss trials, and a 'half' version where the parameters are all shared between gain and loss trials. **(B)** Shows correlation matrices for the best fitting model according to BIC: Tversky Full Model and the 2nd best: The Prelec Full Model. Full details of the colour map can be found in the caption for Figure 4-10.

4.6.5 Alternative Parameter Estimates with MASQ and Negativity Bias

Here, we looked to replicate the result in the main results section in the *Prelec Model*. We tested the prediction error weights with the subjective perception of magnitudes (α), and probabilities (γ). We found correlations between these in gain and loss trials, but they did not exceed the adjusted threshold (α^{Loss}: ρ(76) = 0.219, $p^{HB} < 0.0167$, p = 0.054; γ^{Gain}: ρ(76) = 0.232, $p^{HB} < 0.0167$, p = 0.042) [19]. We next tested whether the negativity bias showed a relationship with any parameters. Here we found that α^{Loss} correlated with the negativity bias (ρ(76) = 0.351, $p^{HB} < 0.0083$, p = 0.0017), but not any other estimates of subjective magnitude or probability, or the difference between them.

4.6.6 Dissociating the Negativity Bias and Changes in Risk Taking

Differences in decision-making causes different values of expected value and prediction errors to be sampled between participants. As we saw a relationship between the negativity bias and increased sensitivity for losses (α^{Loss}), thus it is prudent to test whether this relationship may be an artefact. In order to test whether particular patterns in decision-making bias the *Affective Dynamics Model* by constraining particular estimates (and differences between them), we fit a version of the *Expectations Model* with each parameter shared between gain and loss trials (*Expectations 3c*) and fit the generated data (plus noise) with *Expectations Model (3b)*. If an artefact was caused by changes in decision-making, we would expect to see differences in the prediction errors weights in the new parameter estimates that correlated with subjective estimates of magnitude and probability. None of the prospect theory free parameters correlated with prediction error estimates in this new model, nor the negativity bias (all ρ(76) < 0.19, p > 0.107).

[19] The test statistics in these paragraphs were identical when rounded, this is not an error.

5. Future Prospects of a New Smartphone Application

5.1 Abstract

Large-scale data sets collected through smartphone-based tools have been heralded as one of the most exciting new resources for answering questions in computational psychiatry. So far, the first plays for each participant have been used for the majority of data analysis. We propose that collecting repeated plays in the same individual, sampled in a similar style to experience sampling methods (ESM), may provide valuable new data sets. The *Future Prospects Task* was adapted from a laboratory behavioural study to be included in a new smartphone app *The Happiness project*. Previously, the *Future Prospects Task* found evidence for an affective negativity bias (increased affective responses to losses compare to gains), and influences of positive and negative futures on affective state and risky decision-making. Here, we present some early pilot data (N = 2,768) from *The Happiness Project* where we show evidence for positive futures increasing risky decisions, and negative futures decreasing risky decisions. We consider the appropriateness of this gamified version of *The Future Prospects Task*, and the app generally for providing insights for mood disorders with longitudinal testing.

5.2 Introduction

If we wish to bridge the gap between insights from cognitive psychology and clinical treatment, focus needs to move from not just merely describing behaviour but predicting outcomes (Browning et al 2020). The tools of computational psychiatry are primed to capitalize on the collection of large-scale multivariate datasets generated through smartphone-based tools (Gillan & Rutledge, 2021; Insel, 2017). Our understanding of the mechanisms of mental illness is complicated by the inclusive roles of individual environment, for example urban living (Paykel et al., 2000), socioeconomic status (Lorant et al., 2003), and social isolation (Richard et al., 2017). In order to advance neurocognitive models of mental illness, we need to invest in methods that can generate multivariate data sets with considerably large samples. Smartphone-base tools can contribute by collecting passive data and user generated input.

Passive data is data about how smartphone users interact with their phones (e.g. how many outgoing texts they send a day), and how movement in the world can be sensed through their phones using functions like accelerometers (e.g. how fast they moved). User generated input can include demographic data (which has been shown to be broader, in terms of demographics like age and education; Brown et al., 2014), experience sampling measures of current activities and events, questionnaire measures, and data from gamified cognitive tasks. Previously in this thesis we explored data from a risky decision-making task in *The Great Brain Experiment* (GBE) – a smartphone app containing several gamified cognitive tasks which each took less than 5 minutes to play.

In the past two decades, research based apps including gamified cognitive tasks have recruited large scale samples under the motivation being a *citizen scientist*. Large scale samples give the opportunity to uncover strong but small effect sizes in individual differences and cross-sectional temporal effects. The stop-signal task in the GBE uncovered a decline in reactive control over the lifespan, with a greater decline in women than men (Smittenaar et al., 2015). A decline in pavlovian approach bias was seen over life time, but not pavlovian avoidance bias (Rutledge et al., 2016). Between 2014-2020 , four of the nine peer-reviewed papers utilizing data from the GBE made significant claims about how cognitive traits and abilities change across the lifespan (McNab et al., 2015; Rutledge et al., 2016; Smittenaar et al., 2015; Teki et al., 2016). Other apps include *Sea Hero Quest*, where participants take part in a sea adventure navigating mazes, and chasing fantastical sea creatures (Coutrot et al., 2018). Using the wealth of data generated (over 27,000 individuals), a measure of spatial navigation was developed which was found to be sensitive to a genetic risk factor for Alzheimer's disease in an subsequently recruited in-lab sample (Coughlan et al., 2019). In the past few years, this approach has been adopted by several labs leading to new apps. These include *Brain Explorer* developed by Tobias Hauser and colleagues; a space quest themed app looking to understand why so many mental health problems emerge during adolescence. *Neureka* is an app developed by Claire Gillan and colleagues, which aims to collect data to identify cognitive markers for individual risk of developing dementia.

One way in which smartphone-based tools can contribute to prediction of mental health outcomes, is to capture repeated tests over time for individual participants. This has been the standard of *experience sampling methods* (ESM), which has allowed causal inferences to be

made by examining the time series of data. For example, over 28,000 users downloaded the *58 Seconds app* and provided data over a 1 month period on their mood and activities. This allows researchers to observe that mood increasing activities were selected when in a bad mood (Taquet et al., 2016), an effect they found to be reduced in individuals with a history of depressive symptoms (Taquet et al., 2020). Furthermore, the app *Track Your Happiness*, demonstrated a causal link between mind wandering proceeding reduced happiness (N = 2,250; Killingsworth & Gilbert, 2010). A repeated sampling framework could be utilised with gamified cognitive tasks. Published work from gamified cognitive tasks has largely been focused on the first play of each individual, however many users played games multiple times, with one player contributing over 150 unique plays in the risky decision-making task in the GBE.

Smartphone-based tools may be particularly useful for conditions which are characterized by a high prevalence in the population, and recurrent episodes such as Major Depressive Disorders (MDD) (Solomon, 2000). Spontaneous recovery is also common in MDD; the resolution of symptoms without intervention (Frank et al., 1991). Clinically, this can mean that often individuals no longer show symptoms when they are able to be observed by a clinician, or recover during treatment, independently of the invention itself. Being able to track mood and cognitive biases over time may be extremely useful for discovering risk markers for depressive episodes.

MDD is characterized by negative cognitive biases, particularly particular in terms of attention, memory and interpretation (Mathews & MacLeod, 2005) and have been suggested as a target for behavioural intervention (Browning et al., 2012). For example, following a sad mood induction (sad music and reading negative statements), individuals with previous episodes of depression rated autobiographical positive memories with less vividness than healthy controls (with no history of depression) (Werner-Seidler & Moulds, 2011). Often it is inferred that these negative information processing biases may be accompanied by negative affective biases (i.e. the conscious experience of negative vs positive affect), however this has not been explicitly tested. In previous work in a non-clinical sample, we found evidence for an *affective negativity bias* when comparing affective ratings for positive vs negative outcomes (Chapter 4), consistent with evidence that we see a negativity bias at population, albeit less extreme than in MDD (e.g. Gershman, 2015; Pulcu & Browning, 2017). However we saw evidence that participants reporting greater depressive symptoms had lower affective negative biases. Collecting large

scale data sets about affective bias gradients, and their evolution over time, may provide a useful insight into how valence biases cause, or are caused by depressive symptoms. Furthermore, patients with MDD have also shown to have distinct negative biases in terms of their perception of the past and the future. For example patients with MDD were found to describe prospective positive events in terms of less vivid imagery (e.g. *imagine you will have lots of energy and enthusiasm*) compared to healthy controls (Morina et al., 2011). No difference was found in vividness of prospective negative events (e.g. *imagine someone close will reject you*) and healthy controls. Similarly, in a non-clinical population low BDI-II scores were associated reduced vividness of imagination about future positive events, but not negative future events (Holmes et al., 2008).

Importantly, smartphone-based tools can be used alongside clinical studies to enrich data collected during the study, and provide longer term measures without considerable additional cost and patient burden of laboratory testing. Indeed, smartphone tools do not necessarily need to be used independently of other more traditional methods, they can act as an inexpensive add-on to a behavioural or neuroimaging study looking to explore changes over time. For example, positive affect recorded using ESM has been associated with increased variability in physical locations; an association which showed a stronger link for participants exhibiting greater functional connectivity between the hippocampus and ventral striatum, areas of the brain associated with spatial navigation and reward (Heller et al., 2020). While techniques like fMRI are unlikely to become portable, wearable EEG has been utilised alongside smartphone-based tools to show that the degree to which prediction errors from a reward learning task could be decoded in the EEG signal was predictive of changes in mood later in the day (N = 10, Eldar et al., 2018). An example of a behavioural study that could have utilised smartphone-based cognitive tasks was that of Michely et al (2020) who linked SSRI usage to asymmetric learning rates after a 7 day course. The researchers compared learning rates before and after SSRI course, this study may have been enhanced by short repeated testing on a smartphone based learning task for 5 minutes each day providing a time course of change in the effect of interest.

The data presented in this short chapter is from the first release of the new smartphone application *The Happiness Project* developed by the Rutledge lab. The app was launched on January 4th 2021, and as of February 7th 2021 has over 8,000 downloads from 123 countries.

Co-development of the smartphone application has been a major project of my PhD over 3 years. We have worked to develop an infrastructure using an incentive badge system and push notification schedule that encourages longitudinal participation. Longitudinal data sets will be available in the future, and here I present some early pilot analysis from the first players.

The analysis presented is concerned with establishing that the main effects of interest from the laboratory based version of the *Future Prospects Task* (Chapter 4) are present in the data, in this relatively small sample. Specifically, we predict that participants will report greater happiness in gain blocks than in loss block, and will also report greater happiness when the future contains all gain trials (i.e. positive futures) compared to when it contains all loss trials (i.e. negative futures). We will explore whether the valence of the future blocks also affects risk taking as seen in the lab-based study, where positive futures increase risk taking and negative futures decrease risk taking. We predict that these effects will be observable without utilising the model residuals for analyses. This is due to the shorter nature of the smartphone task (compared to the much longer laboratory task), where the future blocks describe a large proportion of the entire game play. Further work is required to explore model-based analyses, which will focus on reliable parameter estimates in longitudinal data. At this stage we will test whether using a simple affective dynamics model, a baseline mood parameter negatively correlates with depressive symptoms, as would be expected in this design.

In Chapter 3 (*Risk Taking For Potential Losses But Not Gains Increases With Time Of Day*) we observed that risk taking was greater for loss trials as compared to gain trials in a gamified cognitive task, counter to what is seen in the broader literature (Ruggeri et al., 2020). We speculated this was due to the novelty of the smartphone design where a spinner animation took several seconds before revealing the outcome increasing positive anticipation for potential gains (increasing the attractiveness), and an effect of negative anticipation for potential losses (decreasing the attractiveness). We predict we will see the same effect in this task, which includes an animation before the outcome is revealed.

5.3 Experimental Procedures

5.3.1 The Happiness Project

Download from: https://rutledgelab.org/

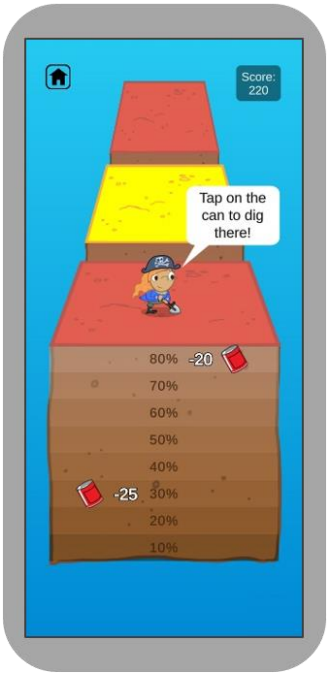
The Happiness Project features gamified cognitive tasks (hence forth referred to as 'games') of under 4 minutes. The original release (4th January 2021) featured four games based on laboratory experiments. *Spinner* is based on the *Reward and Decision-making* task from *The Great Brain Experiment* (the task used in Chapter 3), but also includes forced trials where participants were presented with only one choice. *Wheel of Fortune* is based on the task from (Eldar & Niv, 2015) where participants learn the values of probabilistic rewards with a large positive or negatively valenced mood induction in the middle of the experiment. *Fishing* is based on Niv et al., (2007) which explores opportunity costs and how much vigour participants employ under different rewarding statistics. Finally *Digging for Treasure* is based on the *Future Prospects Task* where the user plays as a pirate digging for treasure on red islands (loss trials) and yellow islands (gain trials). All games featured several happiness ratings throughout. All games apart from *Fishing* are mostly comprised of two alternative forced choice trials.

Ai

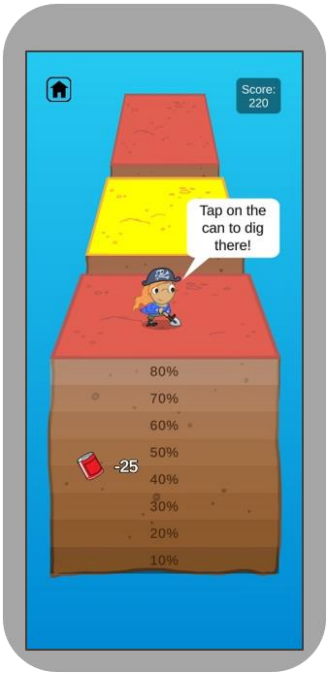
Aii

Aiii

Bi



Bii



Biii

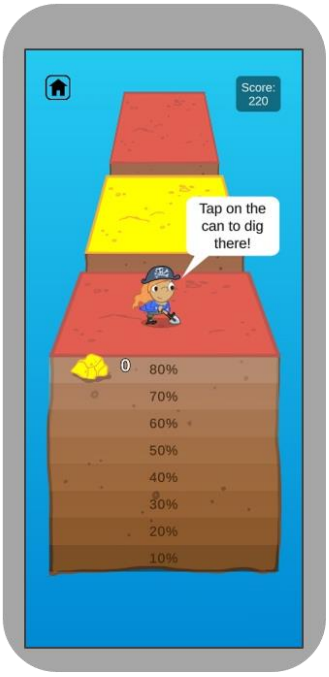


Figure 5-1 The Happiness Project

The Future Prospects task was adapted for the smartphone app *The Happiness Project* **(C)** into a task called *Digging for Treasure* where the principles of the game are the same as those described in Figure 3-1). In the player makes a choice between two gold coins by tapping on the left or right side of the screen **(Ai)**. Gold coins are always worth positive points (i.e. gain trials). Immediately after the choice is made, the unchosen option disappears. After (800ms) an animation is shown where the coin moves up the screen at a slowly, and the outcome is revealed at a random position **(Aii)** between where the coin started and the top of the screen **(Aiii).** In **(A)** the coin reaches the top of the screen and grows larger before disappearing, and the score is updated by 80 points. In **(B)** where the player chooses between rusty cans (i.e. loss trials). In **(Biii)** the can turns into a yellow rock (0 points) before it reaches the top. Once it reaches the top it grows larger before disappearing. When the player transitions between blocks, the pirate sprite jumps in the air and the islands transition forwards so they land on the next island. The animation plays after the happiness trial **(E),** or immediately after the trial if there is no happiness rating at the end of that block. This version of the task has minimal instructions prior to the task, and

the pirate sprite provides some useful reminders throughout the game. **(C)** shows the landing screen of *The Happiness Project* where icons lead to gamified cognitive tasks and questionnaires. (*top left to bottom right)*: *Fishing*, designed by Akshay Nair; *Spinner* (Viking hat), designed by *Bastien Blain*; *Wheel of Fortune* (scarab beetle) designed by Liam Mason & Ilinca Ureche-Angelescu; *Digging for Treasure* (gold coin) designed by Rachel Bedder; and a questionnaire icon. Questionnaire development was led by Matilde Vaghi.

5.3.2 Task Design

The experimental design (other than length) was largely the same as described in Chapter Gain and Loss Asymmetry for Short and Long Term Prospects4 (Figure 4-1) with a few adaptions. New animations were included for the block transition, and outcome reveal (detailed in Figure 5-1a & b). In *Digging for Treasure* players experienced 8 blocks of trials (10 were viewable, but the game terminated before the final two blocks were played). The order of blocks was generated by creating 100,000 different possible orders of 5 gain and 5 loss blocks and filtering them based on set criteria (Table 5-1). Each block contains 3 or 4 trials. The 28 trials for each game are generated from two sets of all available probabilities (14 trials). Each individual probability is randomly allocated a low magnitude risky option [25, 40] and a high magnitude risky option [65 80]. These trials are then duplicated in a randomized order (28 trials). The safe option remains the same as in the laboratory task, 80% chance of receiving a magnitude of 20. Each version had 6 happiness ratings which were distributed randomly throughout the blocks.

Different playable versions of *Digging for Treasure* are exported from MATLAB as json files (JavaScript Object Notation) and added to the app by the outsourced development team (*Crysberry*). The probabilities and magnitudes of each trial are set in each individual json file, with the outcome determined in real time when the player makes a choice. The version of the task experienced by each player were randomly sampled without replacement. Therefore it is possible for the participant to play the same block set and trial set more than once, but the probabilistic outcomes would be determined in each instance.

Table 5-1 Digging for Treasure Block Order Criteria

Criterion	Justification
Ends with [x-GL] or [x-LG]	Does not end with the block order that may lead to the greatest emotional response (i.e. positive or negative futures).
Begins with [l-GL] or [g-LG]	Switching back and forth allows first time players to get used to how the future blocks mean.
No instances for [g-GGG] or [l-LLL]	Four Loss blocks in a row may lead the participant to terminate the game early.
At least 1 of each of the following [l-LL],[l-GL],[l-LG],[l-GG],[g-GG], [g-LG],[g-GL][g-LL]	To allow a parametric function for mood or decision-making to be fit to the future.blocks.

Mixed Futures describe when there is both a gain and loss block in the future. Positive futures are two gain blocks [GG]. Negative futures are two loss blocks [LL]. Here, the order of blocks is indicated by [current block – future block 1, future block 2]. 'x' indicates the current block is unspecified. For example, [l-GG] indicates a loss block with a positive future, and [x-GG] indicates a positive future and any current block.

5.3.3 Questionnaires

The app currently includes 4 questionnaires: Patient Health Questionnaire (PHQ-8; Kroenke et al., 2001); Generalised Anxiety Disorder Scale (GAD-7; Spitzer et al., 2006); the behavioural subscale from the Apathy and Motivation Index (bAMI; Ang et al., 2017); and the Hypermanic Personality Scale (HPS; Eckblad & Chapman, 1986). Questionnaires were introduced on the home screen as items to be completed in a staged manner after completion of some of the games.

5.3.4 Participants

Players downloaded the app from either Google Play or the Apple App Store. The majority of players who downloaded the app were people signed up to an online course *The Science of Wellbeing* (hosted on website coursera.org and designed by Laurie Santos), who received an

email announcement about the new app (amongst other items). Players are not paid for their participation.

5.4 Results

All analysis were performed with each participant's first play, unless otherwise stated. All analyses that refer to other plays only count by the number of full-completed plays.

5.4.1 Statistical Procedures

Two-sided Wilcoxon signed rank tests were used for all comparisons of all task metrics. Spearman rank correlations across participants were used to test for relationships between parameter estimations and questionnaire scores. All averages were reported as (mean ± standard deviation), unless otherwise stated. All analyses were performed using MATLAB. The minimum p value threshold reported is $p < 0.0001$.

5.4.2 Risk Taking

Players gambled more for gains (0.545 ± 0.235) then for losses (0.489 ± 0.159, $z = 8.676$, $p = 0.0001$). Importantly, players chose the option with the greater expected value in loss trials (0.557 ± 0.311; significantly better than chance $z = 8.593$, $p < 0.0001$; Figure 5-2a) and in gain trials (0.665 ± 0.237; significantly better than chance $z = 28.912$, $p < 0.0001$; Figure 5-2a). During in-lab piloting sessions, we received feedback from participants that it was not immediately intuitive how to interpret the onscreen probabilities, especially in loss trials. To explore whether participants understand the game better in the second play, we looked at whether players increased the proportion of times they chose the choice with the best expected value. In players with at least two full games ($N = 955$), the proportion of higher expected value choices increased for loss trials (second play 0.615 ± 0.338; $z = 3.758$, $p < 0.001$) but not gain trials (second play 0.662 ± 0.257; $z = 0.662$, $p < 0.501$).

Next we looked to see if we could see any impact of the future on decision-making. Participants made risky choices more when the future was positive than negative in loss trials (positive

future: 0.506 ± 0.337; negative future: 0.474 ± 0.334; $z = 4.533$, $p < 0.0001$; Figure 5-2cii) and made more risky choices when the future was positive than negative in gain trials (positive futures: 0.587 ± 0.302; negative futures: 0.544 ± 0.313; $z = 5.958$, $p < 0.0001$; Figure 5-2cii). In the laboratory based study we did not see model free changes in risk taking, so this represents a promising increase in effect for the gamified short version of the task.

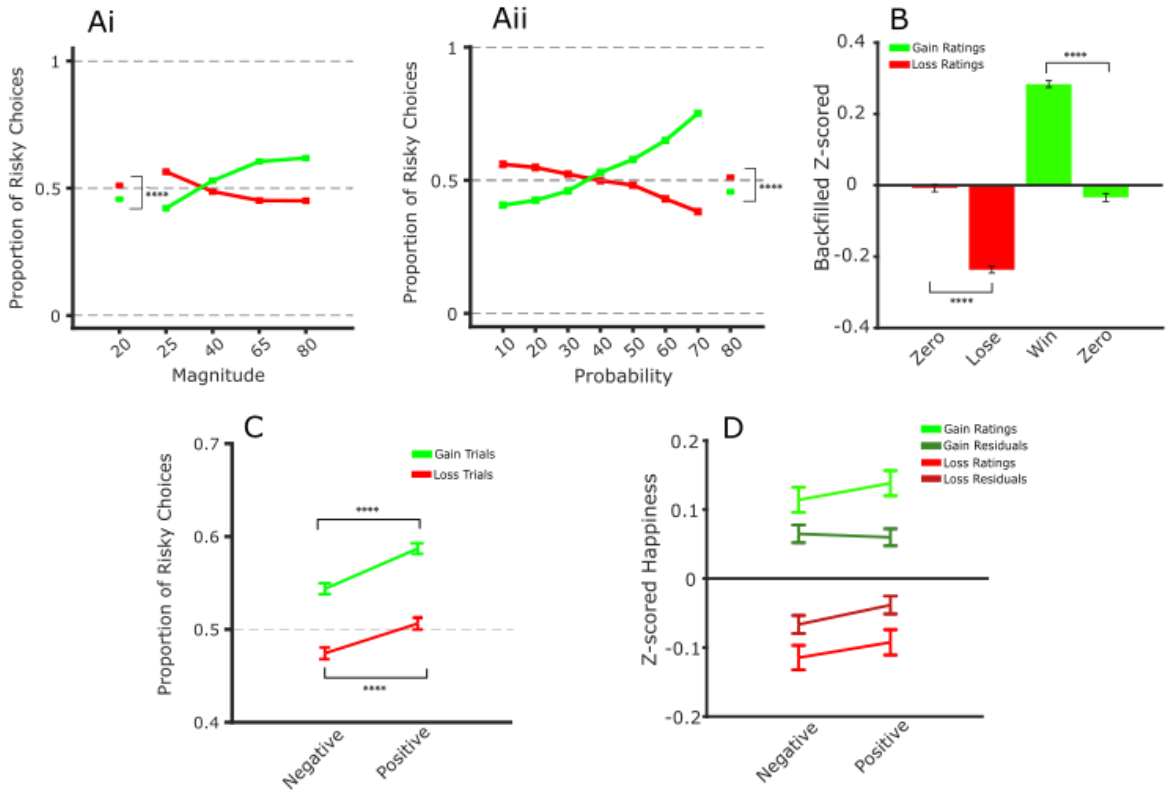


Figure 5-2 Decision-making and Average Happiness Ratings

(A) The mean proportion of risky choices is shown in error bars with the standard deviation (red for loss trials, green for gain trials) for each magnitude **(Ai)** and probability **(Aii)**. **(B)** We report the average happiness for these events as raw scored date in the main text, but z-score the data for clarity here. **(C)** The proportion of risky choices chosen by participants increases when the future is positive (two gain blocks) compared to negative (two loss blocks) when making decisions in both gain and loss trials. **(D)** We did not observe any differences in happiness ratings for positive or negative futures in the ratings, or in the residual ratings after the fitting an affective dynamics model.

5.4.3 Average Happiness Ratings

We first confirmed that choice outcomes impacted happiness rating. Players were happier after gaining points than missing out on gaining (gaining: 0.565 ± 0.113; missing out: 0.422 ± 0.120; $z = 21.055$, $p < 0.0001$) and were happier after avoiding losing than losing (losing: 0.511 ± 0.161; not losing: 0.540 ± 0.160; $z = 7.414$, $p < 0.0001$). Participants were happier overall in gain blocks (0.543 ± 0.135) compared to loss blocks (0.521 ± 0.154; $z = 19.79$, $p < 0.0001$). Using the residuals from an affective dynamics model (see Section 5.4.4) we compared residual happiness (z-scored prior to model fitting) for positive and negative futures for gain and loss trials. We did not find significant differences between happiness for positive and negative futures in loss trials (positive futures: -0.0384 ± 0.637; negative futures: -0.0666 ± 0.6344; $z = 1.12$, $p = 0.263$; Figure 5-2d) nor in gain trials (positive futures: 0.0598 ± 0.602; negative futures: 0.0647 ± 0.629; $z = 0.231$, $p = 0.818$; Figure 5-2d). We also did not find significant differences in the model free data (both $z < 1.00$, $p > 0.39$).

5.4.4 Modelling

The fluctuations in the mood ratings were well captured by affective dynamics models (N = 2,223). We used a version of the affective dynamics model (Equation 1-3) with a baseline ($\mathcal{C}$) , decay parameter (τ), and weights shared between gain and loss trials for expected value and prediction errors (wEV and wPE) (Figure 5-3a). We first fit a version without positive bounds on the affective dynamic weights which had very high average r^2 (0.714 ± 0.238). A considerable proportion of players had r^2 values that exceeded 0.7 (61.7%), this could mean that the happiness ratings are being over fit. The inclusion of positive bounds reduced the average r^2 (0.640 ± 0.286), with 52.1% of participants with r^2 values of over 0.7. Each game only has 6 happiness ratings (the model has 4 parameters), thus it needs to be considered how this overfitting will affect our ability to look at trajectories with precision without suitable power. We examined concatenated plays of at least three (N = 360), and found the r^2 reduced (0.409 ± 0.216; 9.31% over 70%). It may be that to fit longitudinal data the full amount of games needs to be leveraged to estimate change in parameters that are not over fit to the data.

5.4.5 Questionnaire Data

Average happiness correlated with all 4 questionnaires (PHQ-8, ρ(840) = -0.198, p < 0.0001; GAD7, ρ(788) = -0.183, p < 0.0001; HPS ρ(45) = -0.283, p = 0.054; BAMI ρ(347) = -0.212, p < 0.0001). The baseline parameter estimate for the model also correlated with each questionnaire except the HPS (PHQ-8, ρ(793) = -0.198,p < 0.0001; GAD7, ρ(753) = -0.171, p < 0.0001; HPS ρ(37) = -0.299, p = 0.161; BAMI ρ(312) = -0.171, p < 0.0024). The sample size varied between tests as different players completed different questionnaires, and the model did not converge for all participants.

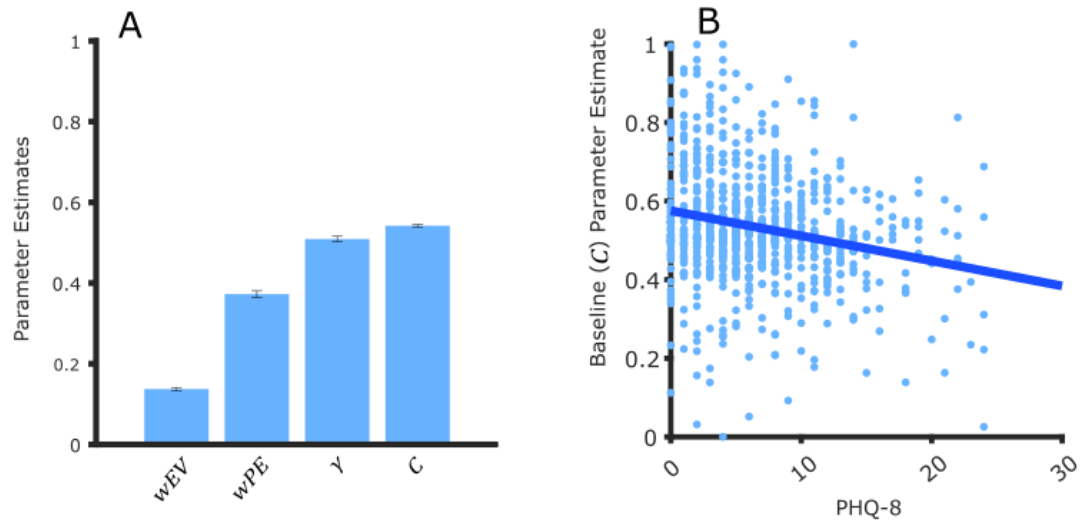


Figure 5-3 Affective Dynamics Model fit to Pilot Data in *The Happiness Project*

(A) Parameter estimates for an *affective dynamics model*, where the parameter are all shared between gain and loss trials. Mean parameters shown here were estimated with positive bounds, error bars indicate the standard error of the mean. **(B)** The parameter estimate for baseline mood (C) was negatively correlated with depressive symptoms (PHQ-8).

5.4.6 Repeated Plays

On average each participant played 1.836 times (±1.943). 65% of participants only had 1 complete play. 6.0% of participants played 5 times or more. One participant played 39 times. One of the aims of the app was to encourage people play on multiple days and examine how

parameter estimates may change as a function of their incidental mood. Without a working notification system, 74.1% of participants played on one day only and 19.3% of participants played at least 2 days apart. While this is lower than needed for our longitudinal tests, it is promising that almost 20% of people played on average 2 days apart.

Data collected by smartphone may be more likely to include some happiness ratings of reduced quality. We tested this by looking at the number of times a participant rated their happiness as either end of the scale or the centre. We included a participant in the analysis if they had at least 2 ratings which were not all at the centre or either end of the scale. We found in the first play 13.22% of participants were removed (12.10% had 5 or more ratings of 0.5). The amount of participants required to be removed appeared consistent across multiple plays: the second play (N = 955) required 12.57% to be removed, and the 5th play (N = 165) 16.36% required removal.

To examine whether participants behaviour was stable and thus likely to be reliably capturing individual differences we compared data between participants first and second play. Proportion of risky choices for gains and losses correlated between first and second play (Gains $r(955) = 0.393$; Losses: $r(955) = 0.471$; both $p < 0.0001$). Gains, Losses, and receiving zero in both frames all correlated between plays (all $r > 0.45$, $p < 0.0001$).

5.5 Discussion

The results show that at a basic level, happiness ratings and decision-making behaviour are consistent with what we would expect for rational behaviour and happiness ratings for our most basic model assumptions. For example, participants reported greater happiness ratings when they won points, compared to when they missed out on winning points. We saw an effect of future prospects on decision-making without regressing out trial-by-trial events, where positive futures are associated with increased risk taking compared to negative futures. This may be due to increased power from sample size, or the effects may be greater in the adapted version than the in lab behavioural version. We postulate that the new design may increase the effect sizes, due to the design novelty of the future blocks and the shorter task meaning they represent a larger proportion of all possible winnings. If this is the case, it may be possible to examine the effect of positive or negative futures along an *affective gradient*. Here, we could examine

affective gradients regarding current events (i.e. current trial) and future beliefs, to determine whether biases held here can be captured under one gradient mechanism. Surprisingly, we did not see an effect of future prospects on happiness ratings. This may be due to their being significant changes in decision-making based on future prospects, and therefore different outcomes distributions may mask any effects on happiness ratings. Future simulation work can be performed to test the viability of this as an explanation.

We did not perform analyses exploring affective gradients in the pilot data. The affective dynamics model used to estimate this effect in Chapter 4 required prediction error parameters to be split between gain and loss trials, using a minimum of five parameters. Here, we show that in a model with only four parameters (where all parameters are shared between gains and losses), the happiness ratings are likely to be over fit. Thus, making inferences about affective ratings may need to be leveraged by combining happiness ratings from multiple game plays, a data set we plan to collect in the future. We were able to demonstrate that the baseline parameter was negatively correlated with depressive symptoms (consistent with previous work in Chapter 4, and in Blain & Rutledge (2020)).

We hypothesised that we may see greater risk taking for losses compared to gains, due to the animation for the risky option taking a longer amount of time than for the safer option (as the coin or can has to travel to the top of the screen which takes longer in the risky options). We did see that participants made the risky choice more for gains than losses, however participants informally also reported some confusion on understanding the loss trials in the first plays. Thus, this hypothesis can be revisited once a significant number of participants have played multiple times and we can be assured they fully understand the loss trials.

While here I present the pilot data from only one of the gamified tasks, many of the games feature happiness ratings and two alternative forced choices. It may be that the data from multiple tasks can be leveraged to constrain parameter estimation (e.g. hierarchical estimation of affective dynamics or baseline estimates). Furthermore, affective characteristics from multiple tasks can be combined to form different emotional profiles, that at best can combined may increase predictive power, or alternatively allow for an analysis of shared variance to work out the specificity of the construct they claim to measure (e.g. as shown in the meta analysis ESM work by Dejonckheere et al., 2019). For example, the *Wheel of Fortune* game based on the work by Eran Eldar and colleagues (Mason et al., 2017) allows for the estimation of a *mood bias*

which indexes the degree to which a mid-task mood manipulation (i.e. a windfall or large loss of points on a spinning wheel) affects the perception of subsequent outcomes (i.e. a mood bias on learning). Combining estimates of affective gradients and mood bias could provide a deeper characterisation of emotional profiles. For example, a positivity affective bias may be beneficial, but if combined with a high mood bias, this can lead to instability in mood (Eldar et al., 2016; Mason et al., 2017).

A further discussion about the potential of gamified cognitive tasks for computational psychiatry can be read in Section 6.3.4 (*The Future of Smartphones for Cognitive Science & Computational Psychiatry*).

6. General Discussion

6.1 Overall Summary

In this thesis I have used computational modelling of affective dynamics and decision-making to investigate how exploring the difference between individual responses to gains and losses may provide new insights. Chapter 3 introduced the approach of using smartphones to capture large scale data sets to illuminate small, but important, effect sizes in diurnal changes in cognitive biases. Here we showed that the time of day constrains attitudes to risk regarding prospective losses, but not gains. Using computational modelling we demonstrated this effect could be described as reduced risk sensitivity for losses increasing throughout the day. By being able to show diurnal changes to prospective losses but not gains, we were able to speculate on the role of serotonin in this behaviour, which has been implicated in loss related behaviours, and shows a distinct circadian rhythm itself.

After the groundwork had been set for comparing gains and losses, in Chapter 4 I expanded this to the domain of mood and affective responses. I designed a novel paradigm (*the Future Prospects Task*) where gain and loss trials were organised into short blocks of trials, and participants had knowledge of both short and long term expectations. We fit an affective dynamics model to dense sampling of affective state throughout the experiment. This model revealed a *negativity bias*: where affective responses to losses exceeded those of equivalent gains. We found this negativity bias was linked to risk sensitivity to losses, and the reflection effect. We also demonstrated that the valence of longer term prospects, affected ratings of affective state (or happiness ratings), and found intriguing results showing that the influence of the future's valence on decision-making is most observable in proximity to the future. The neural mechanisms behind this may be similar to those of anticipation and dread, which I further explore below (Section 6.2.2).

Both these studies show how comparing between losses and gains, whether it be as main effects (e.g. time of day being associated with differences in risk taking for losses, but not gains), or within individuals (e.g. comparing between affective responses for gains and losses) can provide valuable insights into the mechanisms that drive our choices and feelings. In *The Great*

Brain Experiment we were able to provide mechanistic hypotheses regarding the change in risk taking seen throughout the day due to its influence on loss, but not gain sensitivity. In the *Future Prospects Task* comparing between affective responses to gains and losses as a within subjects measure allowed us to be more confident that affects were not reflections of individual differences in use of the continuous scale.

The *affective negativity bias*, and the influence of the future on affective state and decision-making may both be understood in terms of their position an affective gradient. I suggest that understanding how an individual's position on the affective gradient may fluctuate over time may provide useful insights into mood disorders, when explored alongside measures of sub-clinical and clinical symptoms. In Chapter 5 I present pilot data from a gamified version of the the *Future Prospects Task* from the new smartphone app *The Happiness Project* which was designed over 3 years of my PhD. The long-term goal of this app is to collect data on affective and cognitive biases with repeated longitudinal testing, expanding on the work done with *The Great Brain Experiment*.

In the smartphone-based version of the *Future Prospects Task* we were able to replicate some, but not all of the main effects from the laboratory based version. For example, we saw that the valence of the future prospects had effects on decision-making in both versions (and to a larger degree in the smartphone version), but the valence did not appear to effect happiness ratings in the smartphone version. Additionally, at this early stage, we were not able to see the presence of a *negativity bias* in the smartphone version; however both of these results may be due to the simplistic modelling done with the limited pilot data. In Chapter 3, we also noted divergences between the main effects (i.e. how risk taking was less prevent in loss trials), and what is typically observed in the literature. We suggested this may be due to the increased effects of anticipation and dread when using a animation of a spinner for risky (but not safe) choice options. This effect was also seen in the smartphone version of the *Future Prospects Task,* however this may be due to participants not immediately understanding the loss trials. Where behavioural and smartphone results are seen to diverge, it may be prudent to run both laboratory and smartphone versions of the task in the same individuals to examine whether they are systematic changes due to the format of the task, or if it may be due to differences in sampling pools. This work can be done for *Future Prospects Task,* with participants visiting the laboratory before completing the smartphone version for follow-ups. *This* is a crucial question

for computational psychiatry, should the field continue to utilise smartphone applications for prediction and longitudinal testing.

In this final chapter, I will discuss some future directions for behavioural and neuroimaging work based on the results from Chapter 1. I will also explore what I believe to be some of the theoretical limits of using descriptive models for mood, and some methodological limitations. I will also briefly discuss where I see use the use of gamified cognitive tasks (as featured Chapters 4 & 5) contributing to the computational psychiatry framework in the near future.

6.2 Future Directions

6.2.1 How can we Further Explore the Role of Expectations?

The majority of computational studies of mood have focused on how outcomes may drive affective responses, particularly in terms of how surprising they are. While affective dynamics models include a role for expectations, this is implicitly part of the response at the time point of outcome, and does not necessarily correspond to affect experienced at the point of using expectations to make a choice. Indeed, a 2nd behavioural study of my PhD (not featured here) focused on trying to capture affective responses to being presented with a (mixed prospects) choice, but found inconclusive results. One important issue with exploring expectations and mood is that ideally it requires participants to attend to the prospects of something that has not happened yet, compared to something that is happening now (i.e. the outcome of a choice recently made). Thus, there may be considerable individual differences in attention to expectations which may be conflated with the strength of affective responses to expectations. This becomes even more difficult the further in the future expectations will be experienced, as attention is likely to be drawn to other things. It is well known that our beliefs and expectations about the world guide our decisions, it may be that beliefs play a much stronger role in mood than we currently expect, but are undervalued due to the difficulty in experimentally testing such hypotheses.

Finally, another hindrance in the direct study of expectations is *when* we are able to sample affective states. In the Future Prospects task we are able to show decision-making changing throughout a block, but only collect happiness ratings at the end of each block. Why not include happiness ratings throughout the block? When exploring the role of emotions, designing a task

which does not annoy the participant (unless annoyance is the effect of interest) is of paramount importance. Happiness ratings can feel like an interruption when a participant has been instructed to try and maximize reward, and does not optimize for this by reporting their happiness. Thus, we felt it was best to sample happiness at the end of each block before the block transition animation, making it predictable when they would be *interrupted*. However it is worth noting, the assumption that within block, (or within trial) will worsen data quality is not data driven itself. Future work focusing on methodology of sampling happiness ratings could explore ways to enable this.

6.2.2 An fMRI Study Exploring Future Prospects

I began data collection for an fMRI study exploring how future prospects affect mood and decision-making in December 2019 which was disrupted by COVID-19 (Perlman, 2020). The aim of this study was to expand upon the results found in Chapter 4, where we demonstrated increased affective responses for losing points compared to gains points (*affective negativity bias*), and changes in decision-making as future prospects increased in proximity. With some notable design changes (Figure 6-1 & Figure 6-2), we ran the same task paradigm in MRI with a view to collect 50 participants for model based analyses. We had collected 15 participants, including 5 pilot participants, in total when scanning was terminated. Below, I will briefly describe some of the hypotheses and analyses we had planned for this study.

Regarding the *affective negativity bias* we planned to explore whether those with greater prediction error parameter weights (from an affective dynamics model) for losses compared to gains also showed greater activity in the ventral striatum at the point of outcome presentation for loss trials. In previous fMRI studies that explored affective dynamics models, model parameter estimates were not correlated with trial-by-trial activation strength in the ventral median prefrontal cortex (vmPFC), or the ventral striatum (VS), but the objective prediction errors of the trials themselves were (Rutledge et al., 2014; Rutledge et al., 2017). This suggests that the prediction error weights from affective dynamics models may not reflect differences in neural prediction errors. However individual parameter estimates for prediction errors have been shown to correlate with neural activity when they drive subsequent choices (e.g. in reinforcement learning tasks, (Niv et al., 2012) and when mood predicts choice (Vinckier et al.,

2018a)), thus we predicted we may find a relationship in this study, due to the relationship observed between prediction errors and risk aversion for losses.

We noted in the behavioural work that parameter weights for expected value were highly correlated with the baseline mood parameter and thus the observed increase in expected value weights for losses compared to gains must be interpreted with caution. We planned to leverage fMRI to infer whether these differences are substantive by comparing activation in the vmPFC at the point of choice onset between gains and loss trials. However, the caveat above again applies, that individual participant's parameter estimates in an affective dynamics model have not been shown to consistently correlate with neural activation in value encoding areas.

Our most exciting neural hypothesis pertained to the changes in decision-making we see throughout each block as a function of the valence of the future block. Here, we would be guided by the results shown by Kiyohito Iigaya. Briefly, Iigaya and colleagues present a paradigm where participants choose to receive prior information about the rewarding nature of a stimulus to be presented after less than a minute. They observed a prediction error signal in the dopaminergic midbrain when information about a future reward is revealed, and increasing activity in the vmPFC as they approach the delayed rewarding outcome, which they describe as a signal of anticipation (Iigaya et al., 2019). This study was done with primary rewards (attractive pictures), and we hypothesise we will see similar neural activation in the dopaminergic midbrain and vmPFC for secondary rewards. Notably our study uses categorical future prospects (positive or negative), which are not deterministically relate to future outcomes (i.e. a positive future does not *guarantee* gaining points, it only guarantees that it will be possible to and that no points will be lost), thus we will explore whether those with greater neural prediction error signals at the point the new future is revealed (Figure 6-2) also show greater anticipatory signals in the vMPFC. We predict that the strength of participant's anticipatory signal in the vmPFC and prediction error in the VTA/SN will positively correlate with changes in decision-making throughout the length of the block. Thus, this will extend our behavioural findings by describing the neural mechanism by which positive and negative futures may change behaviour. In a further analysis, Iigaya and colleagues found that those with greater functional connectivity between the hippocampus and the vmPFC and midbrain showed greater anticipatory and reward prediction error signals. They suggest this to be due to the hippocampus' role in *imagining* future reward, which could also be explored within our study.

Iigaya and colleagues focused on the anticipation of rewards in the brain. Other studies have presented similar paradigms that explore the experience of dread (Berns et al., 2006), but these have utilised electric shocks. To my knowledge neural models including both the anticipation of gains and dread of losses have not been compared within the same study. We hypothesised that we would observe the same mechanism at play for future losses. Specifically we would observe negative prediction errors when the future was revealed, and a decrease in activation in the vmPFC as the future approaches in proximity, corresponding to reduced risky decisions being chosen. We were also interested in comparing whether the strength of the prediction error signal for positive futures is present to the same degree for negative futures.

In this adapted version of the *Future prospects task* we simplified the future by only showing two difference valences – one positive block or one negative block. One reason for this was because the neutral futures included in the laboratory version of the task where difficult to interpret, as they where temporally related to the positive and negative futures (i.e. a negative future can only be succeeded by a neutral future with a loss block first then a gain block, but not a neutral future with a gain block first). We included a third condition where the future was not shown to participants on 50% of the blocks (uncertain future). The inclusion of this condition allows for information about positive and negative futures to be less frequent and thus more surprising. We also have the opportunity to compare whether receiving information about the future constitutes a neural information prediction error in irrespective of its valence. However, we have no strong hypotheses regarding this at this stage.

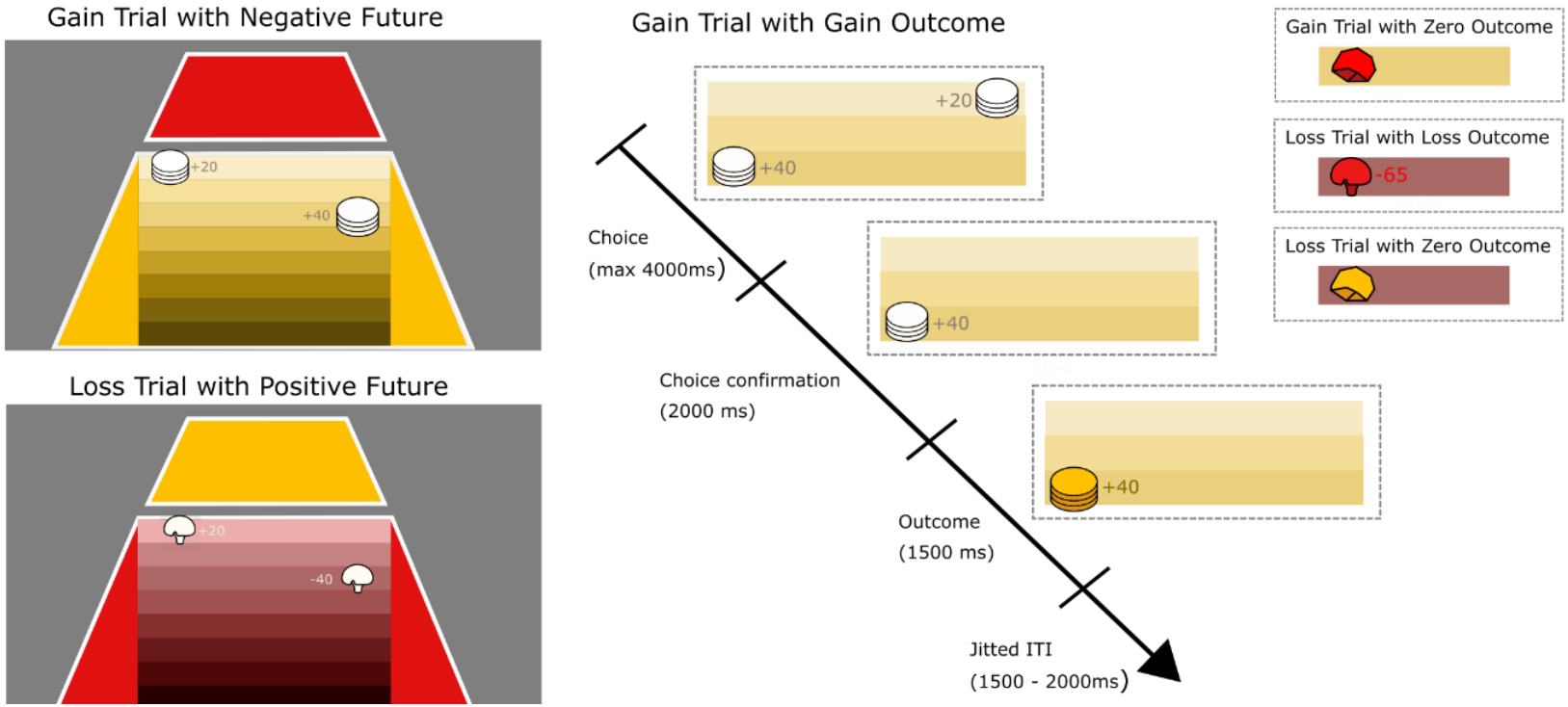


Figure 6-1 Future Prospects Task Adapted for fMRI

The adapted task was identical to the behavioural version (Figure 3-1) with the following changes. We increased the average number of trials per block to 3.5 (from 2.5) in order to increase our ability to observe any signal change in neural activation throughout each block. A total of 210 choice trials were performed for each participant in MRI scanner (excluding practice sessions in and out of the scanner). We adapted the colour scheme for the monitor brightness and contrast in the MRI scanner. 60 blocks performed in the scanner.

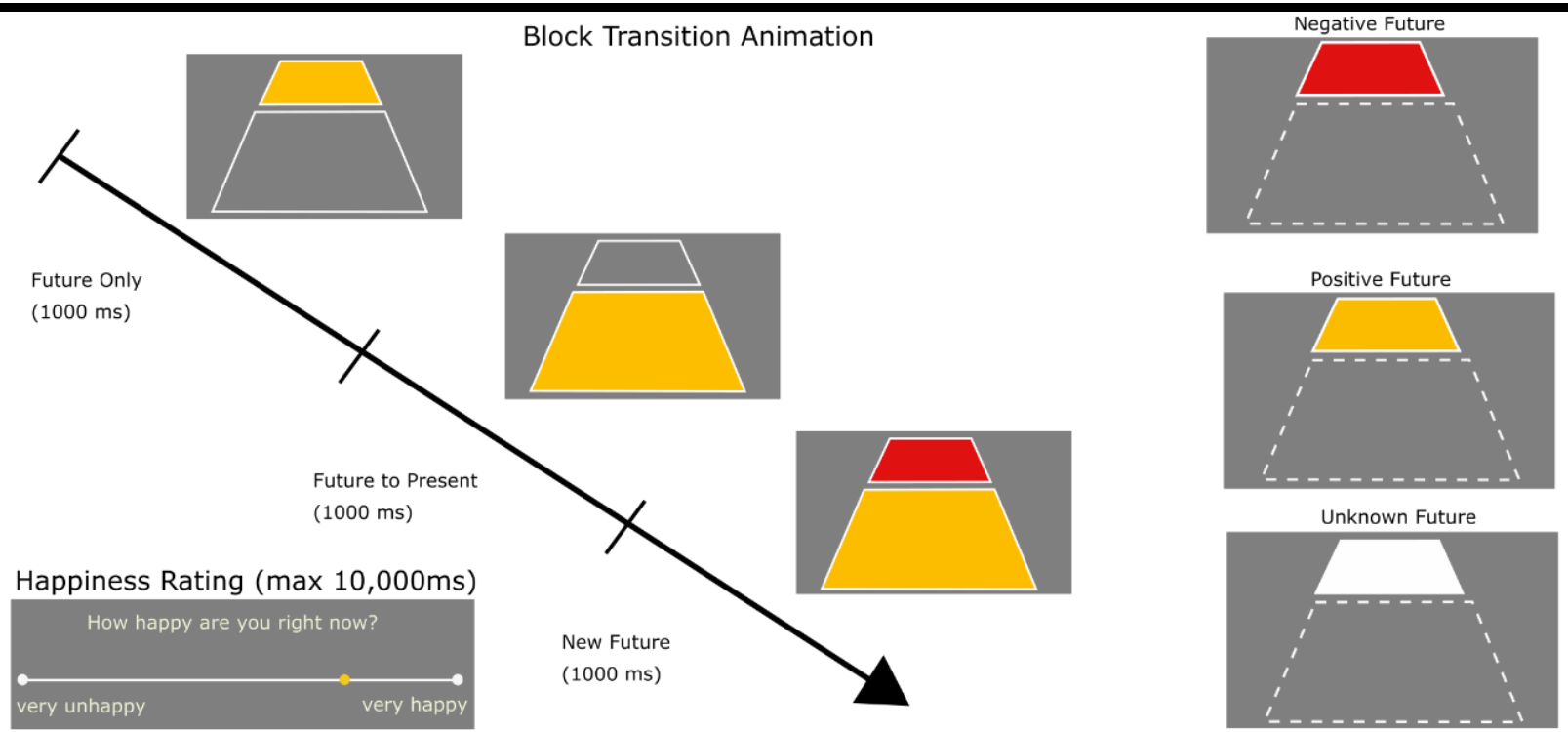


Figure 6-2 Future Block Transition for fMRI

We made changes to the block transition to make sure we could record neural activation when the new future was introduced without confounds of other task events. One significant adaptation from the behavioural version of the task was removing mixed futures, and including unknown futures (30 in total, equally divided between gain and loss).

6.3 Methodological Limitations

6.3.1 How do we Know we are Sampling Affective States?

This thesis began with a description of how mood, emotions and feelings have been defined within the literature. While the field may have broadly converged on definitions, there is no guarantee our participants also subscribe to these. In Chapters 4 and 5, participants are asked to make a rating of how happy they currently feel, ranging from the labels *very unhappy* to *very happy* in the laboratory studies, or emoticons that illustrate these labels in all the smartphone-based studies. Individual subjective interpretations can feed into multiple parts of the process of reporting ones current affective state. For example, participants may interpret the word *happy* to mean reflecting their current affective state, or an evaluative judgement about how well things are going, reflecting their current luck. I believe it is likely they use a combination of the two (i.e. introspection of state, and retrospection of events). It is not clear whether these two interpretations would require distinct models (i.e. reflecting on how well have done may refer to the amount of recent positive prediction errors).

When looking to make inferences about affective dynamics, we statistically analyse the degree to which participants move the cursor away from their model-inferred baseline. To be able to compare individual participant's responses to the same events, we assume they consider the range of the scale to be the same possible emotional range. This assumption may be problematic if incorrect, especially if assumptions about scale range co-vary with symptom dimensions. In view of this, where possible I have only made major claims, which compare affective responses within individuals for different events and use these as individual differences to compare between participants. For example the *affective bias* analysed in Chapter 4, relies upon comparing affective prediction errors between gain trials and loss trials within an individual.

Participants are financially incentivised to make rational choices in the decision-making elements of the task, as the money they take away from the experiment is a function of the total points accrued. There is no such equivalent for rating their current affective state, beyond doing what they are asked to do by the experimenter. Thus, it is also possible that participants will report their happiness based on what they predict the experimenter is expecting to see. Affective ratings are sampled frequently throughout the task, so the participants may be aware

the experimenters expect to see variance in ratings on such a time scale. For example, a participant may predict that after winning (or losing) a large amount, the experimenter may be looking to see a considerable degree of change in their rating, or that they should rate their happiness as greater after a gain trial than after a loss trial. However, there is evidence and arguments which suggests that participants do not do this to a considerable degree. Firstly, reporting ones current affective state in response to being asked about *how happy you are* is easy and intuitive to do; particularly as we aim to make the tasks as engaging as possible when considering design and length. Secondly, a prediction error model in many studies performs better at fitting the data than an outcomes based model. Unless a participant is well versed in cognitive psychology, it is unlikely they would use a model with additional complexity to report their affective state. Furthermore, consistency of the magnitude of parameter estimates between participants, for example the size of the prediction errors (larger) compared to the degree of change in affective state from the future prospects (similar), also suggests a similar model of affective responses across individuals rather than being the result of demand characteristics.

Researchers using continuous or visual analogue scales to model affective ratings have been hesitant to validate the use of the scale with other, particularly physiological measures (e.g. pupillometry, heart rate, galvanic skin response). One reason for this may be that it is not clear how subjective affective reports should map onto objective physiological responses. Future studies exploring these kinds of connections may reveal vital individual differences in how people use information about their bodily state to infer their own feelings, and their degrees of affective responses.

6.3.2 Can we Identify both Mood Baselines and Affective Responses?

Affective dynamics models of mood have two main elements: a baseline which affective responses fluctuate around, and those affective responses themselves. These elements may be ideal candidates for emotional phenotyping, e.g. it may be useful to distinguish between someone who has a very low mood baseline but does not have a strong affective response to a negative stimulus, and someone who operates from a higher mood baseline but has considerable mood swings when encountering negative stimuli. As descriptive models (by

definition) do not explain behaviour, their value is in their ability to estimate latent variables to a high degree of accuracy. Adequately distinguishing between baseline mood and patterns of mood dynamics requires them both to be able to be estimated independently[20].Experimental results presented in Chapter 4 found that anhedonia symptoms were associated with a lower mood baseline and a reduced negativity bias. We used simulation tests to explore whether they can be independently estimated when using a bounded scale. Here, I present a more thorough account of this concern, and consider whether our current methodological approach is equipped to estimate unique variance in mood fluctuations which can be attributed to baselines and responses. I will also suggest some new ways to explore estimating latent variables which could improve the reliability of these parameter estimates.

The identifiability of baseline and affective dynamics parameters governing mood dynamics is at risk when using a bounded scale for reporting affective state. On average, most studies have found that baseline mood is estimated around the centre of the scale, however greater depressive symptoms have been associated with lower baselines (in healthy controls (Blain & Rutledge, 2020) and in patients with major depressive disorder (MDD) (Rutledge et al., 2017). Rutledge and colleagues observe baseline estimates as low as zero in MDD patients. In situations where the mood baseline is low, it may not be possible to capture the full range of negative prediction errors, leading these parameters to compete for variance. Figure 6-3 illustrates the potential problems that can arise from directly comparing affective dynamics between gains and losses when the baseline is not the centre of the scale.

6.3.3 What Variance is most Important?

When fitting a computational model of any kind the experimenter chooses a range of task events which will allow them to fit their function of interest. For example, when fitting a prospect style theory model it is prudent to include choices where one option is extremely attractive (or unattractive) in order to sample the participant's choice at the extremities.

[20] Estimating baseline mood and affective reactions independently does not assume they are independent processes. For example, baseline mood may determine the range and asymmetry in affective responses.

Furthermore, repeatedly sampling choices which involve two options that only marginally differ in their desirability (or undesirability) allows for an estimation of an indifference point and how deterministic behavior is. Here poses an issue for studies which explore affective responses, as we only sample affective ratings regarding choices that have been made (and rarely the counterfactual). This creates sampling difference *between* participants; each participant's experience of expected values and reward prediction errors depends upon their individual choice biases. For example, if a participant never chooses high win low probability options they will not experience large unexpected windfalls. Furthermore, there are sampling differences *within* participants driven by stochasticity of choices and stimulus probability; for example extreme outcomes may be sampled only once, compared to small outcomes being sampled multiple times.

To the author's knowledge, all currently published work in mood modelling involves fitting the observed data to generated data using a total variance explained approach. However, it may be that some parts of the function have greater relevant variance to a symptom dimension of interest and thus should be prioritized when comparing between models (Palminteri et al., 2017). For example; if we wished to understand whether people at risk of depressive relapse showed differential affective responses to emotionally salient stimuli compared to those not at risk, would it be more relevant to compare reactions to common events which are likely to have a smaller response, or compare reactions to extreme events which are likely to have a greater response? Using the prevailing model fitting approach common events are likely to have a greater impact on parameter estimates and, which may leave valuable variance about responses to rarer events unexplained by latent variable estimates in the computational model This creates an interesting question for comparing participants along an important dimension (i.e. self-reported symptom questionnaire) where it is possible that behaviour also varies along the same dimension, which may lead to differences in outcomes sampled, and thus bias parameter estimates in an affective dynamics model in unexpected ways.

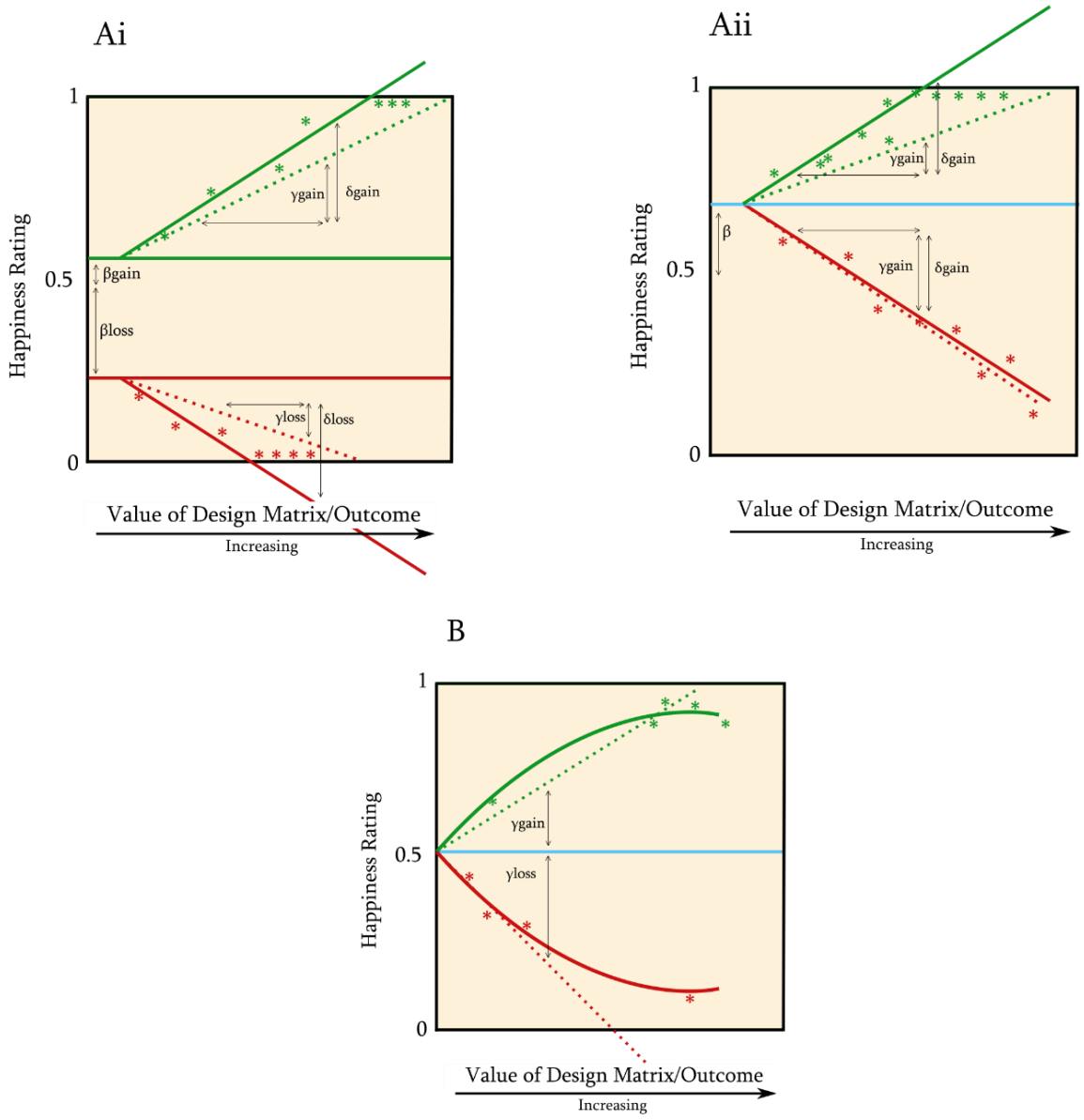


Figure 6-3 Baselines and Boundaries

Here is illustrated a scenario where we wish to estimate baselines and affective responses (e.g. weights for expected value) for gain trials (green) and loss trials (red). The y-axis shows the scale which happiness ratings are sampled on. The x-axis shows increasing values of a task event (e.g. expected value, or signed outcomes). Filled lines describe the 'real function', i.e. what would be fit to the mood ratings if the scale had no boundaries and no noise (linear functions of happiness to affective responses: δ^{gain} and δ^{loss}). Asterisks describe sampled ratings based upon that function. Dotted lines describe how the sampled ratings may be fitted using the estimation strategies included in this thesis (γ^{gain} and γ^{loss}, i.e. estimating

parameters using an optimization procedure). For the sake of this illustration, we assume the fitted baseline is identical to the baseline from the 'real function', and we only describe a scenario where the estimated baseline affects the estimations of affective dynamics, rather than when they may compete for variance. **(Ai)** Shows a scenario where the baseline for the gain trials is closer to the centre of the scale (0.5) by a distance of β^{gain}. In loss trials, the baseline is a further absolute distance from the centre (β^{loss}). Importantly, the sign of the distance from the centre is the same as the sign of most of the affective responses in the frame, i.e. In gain trials, the baseline may be more likely to be greater than centre of the scale ($\beta^{gain} > 0$), and most of the affective responses will be positive ($\gamma^{gain} > 0$), and in loss trials the baseline is more likely to be less than the scale ($\beta^{loss} < 0$), and most affective responses will be negative ($\gamma^{loss} < 0$). In this scenario, the absolute distance from the centre is not equal between gain and loss trials ($\beta^{loss} \neq \beta^{gain}$). The consequence of this asymmetrical baseline (and affective responses being in the direction as the distance from the baseline), is that the parameters for affective responses (γ^{gain} and γ^{loss}) are likely to be predictably mis-estimated depending on how distant baseline parameters (β^{loss} and β^{loss}) are from 0.5. In the illustration we see the two real reactivity functions are equal in magnitude but not sign (δ^{gain} and δ^{loss}), but γ^{gain} is closer to true δ^{gain} given β^{gain} is closer to 0.5. **(Aii)** In an alternative model, the gain and loss trials share a baseline. In this case, the mis-estimation in the parameter for affective responding to gains is inversely proportional to that for losses, which is determined by how far and in what direction the baseline parameter is from 0.5. Here, the baseline is much higher than the centre of the scale ($\beta > 0$), and thus affective responses to gains are (γ^{gain}) truncated compared to those to losses (γ^{losses}). **(B)** Illustrates another scenario, this time where the real functions are non-linear (in this case concave). Non-linear functions fit as linear slopes may be bias by choice behaviour, due to certain parts of the function being sampled more frequently than others. For example, if smaller expected values are sampled more frequently in loss trials, and larger expected values are sampled more frequently in gain trials, this could lead to artifactual asymmetries in linear functions fitted to the sampled ratings. Alternatively, it may be that non-linearity is not part of the real function, but observed due to participants being resistant to endorse the extreme poles of the scale.

Concerns about misestimating parameters as a function of populations of interest endorsing extreme responses on self-report is ubiquitous. Indeed, this has been a particular concern in clinical literature. When patient samples score more highly on some measure, ceiling effects may mask importance variance related to the study. Model based solutions have been proposed to mitigate ceiling (or floor) effects in self-report questionnaires, such as the *Tobit Model* (Tobin, 1958). Briefly, a model is fit to the data by splitting the likelihood function into two parts. The first part estimates the probability that the observed data is censored (at the boundary) or uncensored (not impeded by the boundary) given the predictor variances (e.g. the objective trial-by-trial prediction errors). The second part of the function applies only to the uncensored data. This approach was shown to more accurately fit censored data compared to a standard least squares approach in simulation work (McBee, 2010). While I am not prescribing this model in particular, such statistical approaches should be leveraged to determine if it improves indicators of better model fit. This may be particularly useful for scales with reduced granularity, such as those used in smartphone studies where the visual analogue scale is limited in width (size of the screen) and granularity of response (size of finger tap).

6.3.4 The Future of Smartphones for Cognitive Science & Computational Psychiatry

Much has been written on the prospective benefits of smartphone-based tools (for a recent review; Gillan & Rutledge, 2021). Recruiting from a more diverse pool of participants is likely to increase generalizability of insights to clinical samples, as samples collected on smartphone-based tools and online samples tend to be more diverse (Buhrmester et al., 2011). Western university in-lab testing has frequently been critiqued for sampling uniquely WEIRD (White Educated Industrial Rich Democratic) participants (Henrich et al., 2010). 85% of participants from *the Great Brain Experiment* were non-university students (Brown et al., 2014). Thus, there may be limitations in how easily results from smartphone samples may be compared to studies that have been carried out within a physical university setting. Further studies could explore whether discrepancies in results are attributable to differential samples, or laboratory vs smartphone methods.

Furthermore, the full potential of multivariate data sets has not yet explored. Phenotyping based on machine learning driven clusters of demographics, experience sampling methods and

gamified task data could be used for predicting time course of illness or treatment response (Gillan & Whelan, 2017). Classifier performance has been shown to be improved when trained on parameters estimated from computational models rather than on the observed data used to estimate the parameters (on simulated agents, Huys et al., 2016). Future app development could also consider the use of generating trial types in real-time, to maximise estimating an effect of interest in individuals (e.g. using Bayesian adaptive algorithms; Pooseh et al (2018)).

Large-scale data sets may also be useful for hypothesis generation for future studies. For example, our finding that time of day affects loss sensitivity, was not a priori predicted when the *Great Brain Experiment* was developed. Further studies based on these insights could include questionnaire measures on individual diurnal chorotypes (i.e. morning larks and night owls; Horne & Östberg, 1976). Participants could be prompted to play games featuring loss prospects at different times in the day, and even times of the year to explore whether this behaviour varies in individuals with seasonal affective depression, which has shown to be linked to circadian rhythm disruption (Leibenluft et al., 1995; Magnusson & Partonen, 2005).

While there are often no participant costs, the development of smartphones can be highly costly, and labour intensive depending on whether the development is kept within the lab or outsourced to an app developer or company (Teki et al., 2016). With the increased popularity in online testing via platforms such as Amazon Mechanical Turn (MTurk), these online platforms may in future show greater cost-benefit trade-offs, especially if they can incentivise repeated plays, and provide more flexible updating of tasks based on initial insights. While smartphone-based cognitive tasks may be highly beneficial for hypothesis testing, it is worth noting that this may not be possible within the same platform, if a significantly large sample is required. Large samples are often driven by online publicity when the app is realised and can garner the most media attention, thus it may not be possible for the same app to collect another sample of the same size to test follow up hypotheses. For example, *The Great Brain Experiment* (GBE) was first released in March 2015 and was downloaded 16,000 times in the first month of public availability, a significant percentage of its downloads in total. Reliance on a large sample to be well powered for a hypotheses means can mean you only get *one shot* to collect your data.

While smartphone-based data collection shows much potential for questions of interest in computational psychiatry, this method does include some necessary limitations. Previously. high quality data collection has often relied upon the sample size outweighing the error from

multiple sources inherent in collecting data in on a smartphone. These sources include participants being in an uncontrolled environment, using relatively small device and tasks being relatively short. While some gamified cognitive tasks have found effect sizes comparable to laboratory versions, typically a significantly smaller effect size is observed (Brown et al., 2014). Furthermore, to answer questions regarding individual differences, which may be relevance to computational psychiatry, it is not sustainable to rely upon large samples to balance the signal to noise ratio. Below, I briefly reflect on some sources of noise and make suggestions for how these could be mitigated in applications designed in the future.

Participants are able to complete activities in the smartphone app at any time of their choosing. Whilst this means a greater degree of convenience for taking part in scientific research for the participant, the experimenter does not typically receive data on aspects of the environment that may affect a participant's performance, such as noise levels and degree of multi-tasking. It may be that changes in effects of interest (e.g. mood, response times) may correlate with changes in environment. For example, a change in baseline mood may co-occur with a change in living situation and depressive symptoms. Researchers should look to collect data reflecting this if they believe it may affect their main hypotheses, which can be analysed as covariates. Further to this, the length of each task is often made relatively short so they may be completed easily throughout the day. This enforces limitations on the complexity of the task, as participants must be able to quickly understand how they are required to perform and many will likely skip longwinded instruction screens. Sources of error may also relate to the smartphone device itself, particular in terms of the size of the screen, which may affect the granularity of response (particularly with continuous scales). Researchers could consider developing scales that make use of the height and width of the screen, or provide additional visual feedback on the position of the cursor.

6.4 Final Comments

6.4.1 The Limitations of Descriptive Models of Mood

One question that has plagued me during my PhD, is what can we learn from descriptive models of mood?[21] I have looked to draw connections between mood and decision-making by means of comparing parameter estimates from models fit to data collected simultaneously, but analyzed orthogonally (i.e. choices and happiness ratings). Here, I have tried to draw inferences about how affective reactivity and decisions may be linked. How fluctuations in mood depend upon choices is defined within the affective dynamics model (i.e. as the objective expected values), however I have not formalized how mood itself may bias decisions in a model-based way. One reason for this is the functional role of mood is still a matter of academic debate. Understanding the goal of a system, may be the best way to understand how the mechanism works (Marr & Poggio, 1977). Not having a clearly defined functional role makes it difficult to make normative statements about mood or affective responses, as we cannot quantify the degree to which a particular response or state is close to achieving such a goal. Instead, we are perhaps bias to making assumptions based on the explicit pleasant or unpleasantness of such a state; I do not enjoy being in a bad mood, thus a low mood baseline must be sub-optimal. Furthermore, a low mood baseline being associated with depressive symptoms risks being *guilty by association.* Thus, observing connections between symptoms and descriptions of properties of mood is limited without a functional explanation. It may be that low baseline mood serves as an adaptation to being in an environment to many negative events, by reducing the dynamic range for negative responses.

6.4.2 Okay, so what makes us happy?

The word *happiness* serves as a tool in mood research in two ways. Firstly (and most importantly), its ubiquitous colloquial usage makes it an easy to grasp concept by participants being asked to make swift intuitive judgements about their affective state. Secondly, it serves as a strong brand, which can be particularly useful for the dissemination of work regarding the computational modelling of mood to a broader public audience. A by-product of this, is

[21] In addition, to an actual plague.

researchers are often asked, or asked to comment on – *what makes people happy*? *What should we do to be happy?* It is my firm belief that people should terminate asking computational modellers of mood this question; we do not know better than anyone else does. The field is in its infancy, and we have only begun to start understanding the ways in which human emotions evolve and interact with our behaviour.

6.4.3 No but seriously, what makes us happy?

In this thesis I have explored the role of future prospects on mood. While these effects are small in an experimental setting, it may be that they are considerable larger in real life when we attend to thoughts on our own expectations about the future, in terms of our personal goals. I hope the results here are built upon to develop a greater understanding of how expectations and beliefs about the future affect our affective state and mental wellbeing. If anything, based on my work I advocate for the role of hoping for the best, and believing your future to be bright.

I wish to take this opportunity to thank my internal examiner, Professor Sarah Garfinkel and my external examiner, Professor Michael Browning for examining my PhD thesis.

x. References

Abitbol, R., Lebreton, M., Hollard, G., Richmond, B. J., Bouret, S., & Pessiglione, M. (2015). Neural Mechanisms Underlying Contextual Dependency of Subjective Values: Converging Evidence from Monkeys and Humans. *Journal of Neuroscience*, *35*(5), 2308–2320. https://doi.org/10.1523/JNEUROSCI.1878-14.2015

Ackermann, R., & DeRubeis, R. J. (1991). Is depressive realism real? *Clinical Psychology Review*, *11*(5), 565–584. https://doi.org/10.1016/0272-7358(91)90004-E

Ågren, H., Koulu, M., Saavedra, J. M., Potter, W. Z., & Linnoila, M. (1986). Circadian covariation of norepinephrine and serotonin in the locus coeruleus and dorsal raphe nucleus in the rat. *Brain Research*, *397*(2), 353–358. https://doi.org/10.1016/0006-8993(86)90638-4

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, *108*(4), 441–485. https://doi.org/10.1037/0096-3445.108.4.441

Amsterdam, J. D., Settle, R. G., Doty, R. L., Abelman, E., & Winokur, A. (1987). Taste and smell perception in depression. *Biological Psychiatry*, *22*(12), 1481–1485. https://doi.org/10.1016/0006-3223(87)90108-9

Ang, Y.-S., Lockwood, P., Apps, M. A. J., Muhammed, K., & Husain, M. (2017). Distinct Subtypes of Apathy Revealed by the Apathy Motivation Index. *PLOS ONE*, *12*(1), e0169938. https://doi.org/10.1371/journal.pone.0169938

Arnett, J. (1992). Reckless behavior in adolescence: A developmental perspective. *Developmental Review*, *12*(4), 339–373. https://doi.org/10.1016/0273-2297(92)90013-R

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Auerbach, R. P., Pagliaccio, D., & Pizzagalli, D. A. (2019). Toward an Improved Understanding of Anhedonia. *JAMA Psychiatry*, *76*(6), 571–573. https://doi.org/10.1001/jamapsychiatry.2018.4600

Barrett, L. F. (2016). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, nsw154. https://doi.org/10.1093/scan/nsw154

Bassi, A., Colacito, R., & Fulghieri, P. (2013). 'O Sole Mio: An Experimental Analysis of Weather and Risk Attitudes in Financial Decisions. *The Review of Financial Studies*, *26*(7), 1824–1852. https://doi.org/10.1093/rfs/hht004

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is Stronger than Good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Bechara, A., & Damasio, A. R. (2005). The Somatic Marker Hypothesis: A Neural Theory of Economic Decision. *Games and Economic Behavior*, *52*, 336–372. https://doi.org/10.1016/j.geb.2004.06.010

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Manual for the beck depression inventory-II. *San Antonio, TX: Psychological Corporation*, *1*, 82.

Beedie, C., Terry, P., & Lane, A. (2005). Distinctions between emotion and mood. *Cognition & Emotion*, *19*(6), 847–878. https://doi.org/10.1080/02699930541000057

Beevers, C. G., Worthy, D. A., Gorlick, M. A., Nix, B., Chotibut, T., & Todd Maddox, W. (2013). Influence of depression symptoms on history-independent reward and punishment processing. *Psychiatry Research*, *207*(1), 53–60. https://doi.org/10.1016/j.psychres.2012.09.054

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bennett, D., Davidson, G., & Niv, Y. (2020). *A model of mood as integrated advantage*. PsyArXiv. https://doi.org/10.31234/osf.io/dzsme

Berns, G. S., Chappelow, J., Cekic, M., Zink, C. F., Pagnoni, G., & Martin-Skurski, M. E. (2006). Neurobiological Substrates of Dread. *Science*, *312*(5774), 754–758. https://doi.org/10.1126/science.1123721

Blain, B., & Rutledge, R. B. (2020). Momentary subjective well-being depends on learning and not reward. *ELife*, *9*, e57977. https://doi.org/10.7554/eLife.57977

Blakemore, S.-J., & Robbins, T. W. (2012). Decision-making in the adolescent brain. *Nature Neuroscience*, *15*(9), 1184–1191. https://doi.org/10.1038/nn.3177

Bossaerts, P., & Murawski, C. (2015). From behavioural economics to neuroeconomics to decision neuroscience: The ascent of biology in research on human decision making. *Current Opinion in Behavioral Sciences*, *5*, 37–42. https://doi.org/10.1016/j.cobeha.2015.07.001

Bouchard, L. C., Carver, C. S., Mens, M. C., & Scheier, M. F. (2018). Optimism, health, and well-being. In *Positive psychology: Established and emerging issues* (pp. 112–130). Routledge/Taylor & Francis Group.

Brown, H. R., Zeidman, P., Smittenaar, P., Adams, R. A., McNab, F., Rutledge, R. B., & Dolan, R. J. (2014). Crowdsourcing for cognitive science – the utility of smartphones. *PLOS ONE*, *9*(7), e100662. https://doi.org/10.1371/journal.pone.0100662

Brown, V. M., Chen, J., Gillan, C. M., & Price, R. B. (2020). Improving the Reliability of Computational Analyses: Model-Based Planning and Its Relationship With Compulsivity. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(6), 601–609. https://doi.org/10.1016/j.bpsc.2019.12.019

Browning, M., Carter, C. S., Chatham, C., Ouden, H. D., Gillan, M., Baker, J. T., Chekroud, A. M., Cools, R., Dayan, P., Gold, J., Goldstein, R. Z., Hartley, C. A., Kepecs, A., Lawson, R. P., Phillips, M. L., Pizzagalli, D. A., Powers, A., Rindskopf, D., Roiser, J. P., … Paulus, M. (2020). Realizing the Clinical Potential of Computational Psychiatry: Report from the Banbury Center Meeting, February 2019. *Biological Psychiatry*, *88*(2), e5-1e10.

Browning, M., Holmes, E. A., Charles, M., Cowen, P. J., & Harmer, C. J. (2012). Using Attentional Bias Modification as a Cognitive Vaccine Against Depression. *Biological Psychiatry*, *72*(7), 572–579. https://doi.org/10.1016/j.biopsych.2012.04.014

Browning, M., Holmes, E. A., & Harmer, C. J. (2010). The modification of attentional bias to emotional information: A review of the techniques, mechanisms, and relevance to emotional disorders. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(1), 8–20. https://doi.org/10.3758/CABN.10.1.8

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, *125*(3), 367–383. https://doi.org/10.1037/0033-2909.125.3.367

Caraco, T. (1981). Energy budgets, risk and foraging preferences in dark-eyed juncos (Junco hyemalis). *Behavioral Ecology and Sociobiology*, *8*(3), 213–217. https://doi.org/10.1007/BF00299833

Charpentier, C., De Neve, J., Roiser, J. P., & Sharot, T. (2016). Models of affective decision-making: How do feelings predict choice? *Psychological Science*, *27*(6), 763–775. https://doi.org/10.2139/ssrn.2734088

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. *The Lancet Psychiatry*, *3*(3), 243–250. https://doi.org/10.1016/S2215-0366(15)00471-X

Chew, B., Hauser, T. U., Papoutsi, M., Magerkurth, J., Dolan, R. J., & Rutledge, R. B. (2019). Endogenous fluctuations in the dopaminergic midbrain drive behavioral choice variability. *Proceedings of the National Academy of Sciences*, *116*(37), 18732–18737. https://doi.org/10.1073/pnas.1900872116

Chiavacci, S. J., Ward, M. P., & Benson, T. J. (2015). Why fledge early in the day? Examining the role of predation risk in explaining fledging behavior. *Behavioral Ecology*, *26*(2), 593–600. https://doi.org/10.1093/beheco/aru236

Christakou, A., Gershman, S. J., Niv, Y., Simmons, A., Brammer, M., & Rubia, K. (2013). Neural and Psychological Maturation of Decision-making in Adolescence and Young Adulthood.

Journal of Cognitive Neuroscience, *25*(11), 1807–1823. https://doi.org/10.1162/jocn_a_00447

Ciarleglio, C. M., Resuehr, H. E. S., & McMahon, D. G. (2011). Interactions of the serotonin and circadian systems: Nature and nurture in rhythms and blues. *Neuroscience*, *197*, 8–16. https://doi.org/10.1016/j.neuroscience.2011.09.036

Clark, J. E., Watson, S., & Friston, K. J. (2018). What is mood? A computational perspective. *Psychological Medicine*, *48*(14), 2277–2284. https://doi.org/10.1017/S0033291718000430

Clementz, B. A., Sweeney, J. A., Hamm, J. P., Ivleva, E. I., Ethridge, L. E., Pearlson, G. D., Keshavan, M. S., & Tamminga, C. A. (2015). Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers. *American Journal of Psychiatry*, *173*(4), 373–384. https://doi.org/10.1176/appi.ajp.2015.14091200

Cléry-Melin, M.-L., Schmidt, L., Lafargue, G., Baup, N., Fossati, P., & Pessiglione, M. (2011). Why Don't You Try Harder? An Investigation of Effort Production in Major Depression. *PLOS ONE*, *6*(8), e23178. https://doi.org/10.1371/journal.pone.0023178

Conversano, C., Rotondo, A., Lensi, E., Della Vista, O., Arpone, F., & Reda, M. A. (2010). Optimism and Its Impact on Mental and Physical Well-Being. *Clinical Practice and Epidemiology in Mental Health : CP & EMH*, *6*, 25–29. https://doi.org/10.2174/1745017901006010025

Cools, R., Nakamura, K., & Daw, N. D. (2011). Serotonin and dopamine: Unifying affective, activational, and decision functions. *Neuropsychopharmacology*, *36*(1), 98–113. https://doi.org/10.1038/npp.2010.121

Correa, Á., Alguacil, S., Ciria, L. F., Jiménez, A., & Ruz, M. (2020). Circadian rhythms and decision-making: A review and new evidence from electroencephalography. *Chronobiology International*, *37*(4), 520–541. https://doi.org/10.1080/07420528.2020.1715421

Coughlan, G., Coutrot, A., Khondoker, M., Minihane, A.-M., Spiers, H., & Hornberger, M. (2019). Toward personalized cognitive diagnostics of at-genetic-risk Alzheimer's disease. *Proceedings of the National Academy of Sciences*, *116*(19), 9285–9292. https://doi.org/10.1073/pnas.1901600116

Coutrot, A., Silva, R., Manley, E., de Cothi, W., Sami, S., Bohbot, V. D., Wiener, J. M., Hölscher, C., Dalton, R. C., Hornberger, M., & Spiers, H. J. (2018). Global Determinants of Navigation Ability. *Current Biology*, *28*(17), 2861-2866.e4. https://doi.org/10.1016/j.cub.2018.06.009

Cowen, P. J., & Browning, M. (2015). What has serotonin to do with depression? *World Psychiatry*, *14*(2), 158–160. https://doi.org/10.1002/wps.20229

Crockett, M. J., Clark, L., Apergis-Schoute, A. M., Morein-Zamir, S., & Robbins, T. W. (2012). Serotonin modulates the effects of pavlovian aversive predictions on response vigor. *Neuropsychopharmacology*, *37*(10), 2244–2252. https://doi.org/10.1038/npp.2012.75

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*(1), 126. https://doi.org/10.1186/1741-7015-11-126

Dawson, C. J., de Meza, D., Henley, A., & Arabsheibani, G. R. (2015). *The Power of (Non) Positive Thinking: Self-Employed Pessimists Earn More than Optimists* (SSRN Scholarly Paper ID 2655283). Social Science Research Network. https://papers.ssrn.com/abstract=2655283

Dayan, P., & Huys, Q. J. M. (2008). Serotonin, inhibition, and negative mood. *PLoS Computational Biology*, *4*(2). https://doi.org/10.1371/journal.pcbi.0040004

Dayan, P., & Huys, Q. J. M. (2009). Serotonin in affective control. *Annual Review of Neuroscience*, *32*(1), 95–126. https://doi.org/10.1146/annurev.neuro.051508.135607

de Meza, D., & Dawson, C. (2020). Neither an Optimist Nor a Pessimist Be: Mistaken Expectations Lower Well-Being. *Personality and Social Psychology Bulletin*, 0146167220934577. https://doi.org/10.1177/0146167220934577

Dejonckheere, E., Mestdagh, M., Houben, M., Erbas, Y., Pe, M., Koval, P., Brose, A., Bastian, B., & Kuppens, P. (2018). The bipolarity of affect and depressive symptoms. *Journal of Personality and Social Psychology*, *114*(2), 323–341. https://doi.org/10.1037/pspp0000186

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of

psychological well-being. *Nature Human Behaviour*, *3*(5), 478–491. https://doi.org/10.1038/s41562-019-0555-0

Dillard, A. J., Midboe, A. M., & Klein, W. M. P. (2009). The Dark Side of Optimism: Unrealistic Optimism About Problems With Alcohol Predicts Subsequent Negative Event Experiences. *Personality and Social Psychology Bulletin*, *35*(11), 1540–1550. https://doi.org/10.1177/0146167209343124

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R. N., Zebley, B., Oathes, D. J., Etkin, A., Schatzberg, A. F., Sudheimer, K., Keller, J., Mayberg, H. S., Gunning, F. M., Alexopoulos, G. S., Fox, M. D., Pascual-Leone, A., Voss, H. U., … Liston, C. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, *23*(1), 28–38. https://doi.org/10.1038/nm.4246

Dunn, E. W., Wilson, T. D., & Gilbert, D. T. (2016). Location, Location, Location: The Misprediction of Satisfaction in Housing Lotteries: *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/0146167203256867

Eckblad, M., & Chapman, L. J. (1986). Development and validation of a scale for hypomanic personality. *Journal of Abnormal Psychology*, *95*(3), 214–222. https://doi.org/10.1037/0021-843X.95.3.214

Eisenberger, N. I., Berkman, E. T., Inagaki, T. K., Rameson, L. T., Mashal, N. M., & Irwin, M. R. (2010). Inflammation-Induced Anhedonia: Endotoxin Reduces Ventral Striatum Responses to Reward. *Biological Psychiatry*, *68*(8), 748–754. https://doi.org/10.1016/j.biopsych.2010.06.010

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Eldar, E., & Niv, Y. (2015). Interaction between emotional state and learning underlies mood instability. *Nature Communications*, *6*(1). https://doi.org/10.1038/ncomms7149

Eldar, E., Roth, C., Dayan, P., & Dolan, R. J. (2018). Decodability of Reward Learning Signals Predicts Mood Fluctuations. *Current Biology*, *28*(9), 1433-1439.e7. https://doi.org/10.1016/j.cub.2018.03.038

Eldar, E., Rutledge, R. B., Dolan, R. J., & Niv, Y. (2016). Mood as Representation of Momentum. *Trends in Cognitive Sciences*, *20*(1), 15–24. https://doi.org/10.1016/j.tics.2015.07.010

Fareri, D. S., Martin, L. N., & Delgado, M. R. (2008). Reward-related processing in the human brain: Developmental considerations. *Development and Psychopathology*, *20*(4), 1191–1211. https://doi.org/10.1017/S0954579408000576

Frank, E., Prien, R. F., Jarrett, R. B., Keller, M. B., Kupfer, D. J., Lavori, P. W., Rush, A. J., & Weissman, M. M. (1991). Conceptualization and Rationale for Consensus Definitions of Terms in Major Depressive Disorder: Remission, Recovery, Relapse, and Recurrence. *Archives of General Psychiatry*, *48*(9), 851–855. https://doi.org/10.1001/archpsyc.1991.01810330075011

Fuller, P. M., Gooley, J. J., & Saper, C. B. (2006). Neurobiology of the sleep-wake cycle: Sleep architecture, circadian regulation, and regulatory feedback. *Journal of Biological Rhythms*, *21*(6), 482–493. https://doi.org/10.1177/0748730406294627

Gard, D. E., Kring, A. M., Gard, M. G., Horan, W. P., & Green, M. F. (2007). Anhedonia in schizophrenia: Distinctions between anticipatory and consummatory pleasure. *Schizophrenia Research*, *93*(1), 253–260. https://doi.org/10.1016/j.schres.2007.03.008

Gershman, S. J., Pesaran, B., & Daw, N. D. (2009). Human Reinforcement Learning Subdivides Structured Action Spaces by Learning Effector-Specific Values. *Journal of Neuroscience*, *29*(43), 13524–13531. https://doi.org/10.1523/JNEUROSCI.2469-09.2009

Gershman, Samuel J. (2015). Do learning rates adapt to the distribution of rewards? *Psychonomic Bulletin & Review*, *22*(5), 1320–1327. https://doi.org/10.3758/s13423-014-0790-3

Geschwind, N., Nicolson, N. A., Peeters, F., van Os, J., Barge-Schaapveld, D., & Wichers, M. (2011). Early improvement in positive rather than negative emotion predicts remission from depression after pharmacotherapy. *European Neuropsychopharmacology*, *21*(3), 241–247. https://doi.org/10.1016/j.euroneuro.2010.11.004

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology,* 617–638. https://doi.org/10.1037/0022-3514.75.3.617

Gillan, C. M., & Rutledge, R. B. (2021). Smartphones and the Neuroscience of Mental Health. *Annual Review of Neuroscience*, *44*(1), null. https://doi.org/10.1146/annurev-neuro-101220-014053

.Gillan, C. M., & Whelan, R. (2017). What big data can do for treatment in psychiatry. *Current Opinion in Behavioral Sciences*, *18*, 34–42. https://doi.org/10.1016/j.cobeha.2017.07.003

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, *108*(Supplement 3), 15647–15654. https://doi.org/10.1073/pnas.1014269108

Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. *NeuroImage*, *62*(1), 154–166. https://doi.org/10.1016/j.neuroimage.2012.04.024

Harmer, C. J., Shelley, N. C., Cowen, P. J., & Goodwin, G. M. (2004). Increased positive versus negative affective perception and memory in healthy volunteers following selective serotonin and norepinephrine reuptake inhibition. *The American Journal of Psychiatry*, *161*(7), 1256–1263. https://doi.org/10.1176/appi.ajp.161.7.1256

Hassan, M., & Geys, B. (2016). Expectations, realizations, and approval of tablet computers in an educational setting. *Journal of Educational Change*, *17*(2), 171–190. https://doi.org/10.1007/s10833-015-9270-4

Hayden, B., & Niv, Y. (2020). *The case against economic values in the brain*. PsyArXiv. https://doi.org/10.31234/osf.io/7hgup

Heller, A. S., Shi, T. C., Ezie, C. E. C., Reneau, T. R., Baez, L. M., Gibbons, C. J., & Hartley, C. A. (2020). Association between real-world experiential diversity and positive affect relates to hippocampal–striatal functional connectivity. *Nature Neuroscience*, *23*(7), 800–804. https://doi.org/10.1038/s41593-020-0636-4

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hochman, G., & Yechiam, E. (2011). Loss aversion in the eye and in the heart: The autonomic nervous system's responses to losses. *Journal of Behavioral Decision Making*, *24*(2), 140–156. https://doi.org/10.1002/bdm.692

Holmes, E. A., Lang, T. J., Moulds, M. L., & Steele, A. M. (2008). Prospective and positive mental imagery deficits in dysphoria. *Behaviour Research and Therapy*, *46*(8), 976–981. https://doi.org/10.1016/j.brat.2008.04.009

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, *4*, 97–110. https://doi.org/10.1177/0748730415597520

Huang, Y.-F., Soon, C. S., Mullette-Gillman, O. A., & Hsieh, P.-J. (2014). Pre-existing brain states predict risky choices. *NeuroImage*, *101*, 466–472. https://doi.org/10.1016/j.neuroimage.2014.07.036

Hudson, C. G. (2005). Socioeconomic Status and Mental Illness: Tests of the Social Causation and Selection Hypotheses. *American Journal of Orthopsychiatry*, *75*(1), 3–18. https://doi.org/10.1037/0002-9432.75.1.3

Husain, M., & Roiser, J. P. (2018). Neuroscience of apathy and anhedonia: A transdiagnostic approach. *Nature Reviews Neuroscience*, *19*(8), 470–484. https://doi.org/10.1038/s41583-018-0029-9

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, *19*(3), 404–413. https://doi.org/10.1038/nn.4238

Huys, Q. J. M., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, *3*(1), 12. https://doi.org/10.1186/2045-5380-3-12

Huys, Q. J. M., & Renz, D. (2017). A Formal Valuation Framework for Emotions and Their Control. *Biological Psychiatry*, *82*(6), 413–420. https://doi.org/10.1016/j.biopsych.2017.07.003

Hyman, S. E. (2007). Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience*, *8*(9), 725–732. https://doi.org/10.1038/nrn2218

Iigaya, K., Hauser, T. U., Kurth-Nelson, Z., O'Doherty, J. P., Dayan, P., & Dolan, R. J. (2019). *The value of what's to come: Neural mechanisms coupling prediction error and reward anticipation* [Preprint]. Neuroscience. https://doi.org/10.1101/588699

Iigaya, K., Story, G. W., Kurth-Nelson, Z., Dolan, R. J., & Dayan, P. (2016). The modulation of savouring by prediction error and its effects on choice. *ELife*, *5*, e13747. https://doi.org/10.7554/eLife.13747

Ingram, K. K., Ay, A., Kwon, S. B., Woods, K., Escobar, S., Gordon, M., Smith, I. H., Bearden, N., Filipowicz, A., & Jain, K. (2016). Molecular insights into chronotype and time-of-day effects on decision-making. *Scientific Reports*, *6*(1), 29392. https://doi.org/10.1038/srep29392

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Insel, T. R. (2017). Digital Phenotyping: Technology for a New Science of Behavior. *JAMA*, *318*(13), 1215–1216. https://doi.org/10.1001/jama.2017.11295

Isen, A. M. (1993). Positive affect and decision making. In *Handbook of emotions* (pp. 261–277). The Guilford Press.

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of Experienced Utility. *The Quarterly Journal of Economics*, *112*(2), 375–406. https://doi.org/10.1162/003355397555235

Kahneman, Daniel, & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kelly, M. A., & Clark, S. P. (2011). Returns in trading versus non-trading hours: The difference is day and night. *Journal of Asset Management*, *12*(2), 132–145. https://doi.org/10.1057/jam.2011.2

Keren, H., Zheng, C. Y., Jangraw, D. C., Chang, K., Vitale, A., Nielson, D., Rutledge, R. B., Pereira, F., & Stringaris, A. (2019). Timing matters: Primacy effects of experience on mood

revealed using a novel Mood-Machine-Interface. *BioRxiv*, 815944. https://doi.org/10.1101/815944

Kermer, D. A., Driver-Linn, E., Wilson, T. D., & Gilbert, D. T. (2006). Loss Aversion Is an Affective Forecasting Error. *Psychological Science*, *17*(8), 649–653. https://doi.org/10.1111/j.1467-9280.2006.01760.x

Killingsworth, M. A., & Gilbert, D. T. (2010). A Wandering Mind Is an Unhappy Mind. *Science*, *330*(6006), 932–932. https://doi.org/10.1126/science.1192439

Krizo, J. A., & Mintz, E. M. (2015). Sex differences in behavioral circadian rhythms in laboratory rodents. *Frontiers in Endocrinology*, *5*. https://doi.org/10.3389/fendo.2014.00234

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. D. (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain: A Journal of Neurology*, *131*(Pt 8), 2084–2093. https://doi.org/10.1093/brain/awn136

Lakshminarayanan, V. R., Chen, M. K., & Santos, L. R. (2011). The evolution of decision-making under risk: Framing effects in monkey risk preferences. *Journal of Experimental Social Psychology*, *47*(3), 689–693. https://doi.org/10.1016/j.jesp.2010.12.011

Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of Economic Behavior & Organization*, *17*(3), 377–400. https://doi.org/10.1016/S0167-2681(95)90015-2

Lawson, R. P., Nord, C. L., Seymour, B., Thomas, D. L., Dayan, P., Pilling, S., & Roiser, J. P. (2017). Disrupted Habenula Function in Major Depression. *Molecular Psychiatry*, *22*(2), 202–208. https://doi.org/10.1038/mp.2016.81

Lazzaro, S. C., Rutledge, R. B., Burghart, D. R., & Glimcher, P. W. (2016). The impact of menstrual cycle phase on economic choice and rationality. *PLOS ONE*, *11*(1), e0144080. https://doi.org/10.1371/journal.pone.0144080

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, *1*(4), 1–9. https://doi.org/10.1038/s41562-017-0067

Leibenluft, E. (1993). Do gonadal steroids regulate circadian rhythms in humans? *Journal of Affective Disorders*, *29*(2–3), 175–181. https://doi.org/10.1016/0165-0327(93)90031-E

Leibenluft, E., Hardin, T. A., & Rosenthal, N. E. (1995). Gender differences in seasonal affective disorder. *Depression*, *3*(1–2), 13–19. https://doi.org/10.1002/depr.3050030104

Levin, I. P., & Hart, S. S. (2003). Risk preferences in young children: Early evidence of individual differences in reaction to potential gains and losses. *Journal of Behavioral Decision Making*, *16*(5), 397–413. https://doi.org/10.1002/bdm.453

Levine, L. J., Lench, H. C., Kaplan, R. L., & Safer, M. A. (2012). Accuracy and artifact: Reexamining the intensity bias in affective forecasting. *Journal of Personality and Social Psychology*, *103*(4), 584–605. https://doi.org/10.1037/a0029544

Levine, L. J., Lench, H. C., Kaplan, R. L., & Safer, M. A. (2013). Like Schrödinger's cat, the impact bias is both dead and alive: Reply to Wilson and Gilbert (2013). *Journal of Personality and Social Psychology*, *105*(5), 749–756. https://doi.org/10.1037/a0034340

Levy, D. J., & Glimcher, P. W. (2011). Comparing Apples and Oranges: Using Reward-Specific and Reward-General Subjective Value Representation in the Brain. *Journal of Neuroscience*, *31*(41), 14693–14707. https://doi.org/10.1523/JNEUROSCI.2218-11.2011

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–1038. https://doi.org/10.1016/j.conb.2012.06.001

Lighthall, N. R., Mather, M., & Gorlick, M. A. (2009). Acute stress increases sex differences in risk seeking in the balloon analogue risk task. *PLOS ONE*, *4*(7), e6002. https://doi.org/10.1371/journal.pone.0006002

Liu, Y.-Y., Nacher, J. C., Ochiai, T., Martino, M., & Altshuler, Y. (2014). Prospect theory for online financial trading. *PLOS ONE*, *9*(10), 1–7. https://doi.org/10.1371/journal.pone.0109458

Loewenstein, G. (1987). Anticipation and the Valuation of Delayed Consumption. *The Economic Journal*, *97*(387), 666–684. https://doi.org/10.2307/2232929

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

Lorant, V., Kampfl, D., Seghers, A., Deliège, D., Closon, M.-C., & Ansseau, M. (2003). Socio-economic differences in psychiatric in-patient care. *Acta Psychiatrica Scandinavica*, *107*(3), 170–177. https://doi.org/10.1034/j.1600-0447.2003.00071.x

Maddox, W. T., Gorlick, M. A., Worthy, D. A., & Beevers, C. G. (2012). Depressive symptoms enhance loss-minimization, but attenuate gain-maximization in history-dependent decision-making. *Cognition*, *125*(1), 118–124. https://doi.org/10.1016/j.cognition.2012.06.011

Magnusson, A., & Partonen, T. (2005). The diagnosis, symptomatology, and epidemiology of seasonal affective disorder. *CNS Spectrums*, *10*(8), 625–634. https://doi.org/10.1017/S1092852900019593

Marinier, R. P., & Laird, J. E. (2009). Computational Modeling of Mood and Feeling from Emotion. *Cognitive Systems Research*, *10*(1), 48–69.

Marr, D., & Poggio, T. (1977). From understanding computation to understaning neural circuitry. *Neurosciences Research Program Bulletin*, *15*, 470–488.

Mason, L., Eldar, E., & Rutledge, R. B. (2017). Mood Instability and Reward Dysregulation—A Neurocomputational Model of Bipolar Disorder. *JAMA Psychiatry*, *74*(12), 1275. https://doi.org/10.1001/jamapsychiatry.2017.3163

Mateos, S. S., Sánchez, C. L., Paredes, S. D., Barriga, C., & Rodríguez, A. B. (2009). Circadian levels of serotonin in plasma and brain after oral administration of tryptophan in rats. *Basic & Clinical Pharmacology & Toxicology*, *104*(1), 52–59. https://doi.org/10.1111/j.1742-7843.2008.00333.x

Mathews, A., & MacLeod, C. (2005). Cognitive Vulnerability to Emotional Disorders. *Annual Review of Clinical Psychology*, *1*(1), 167–195. https://doi.org/10.1146/annurev.clinpsy.1.102803.143916

May, C. P. (1999). Synchrony effects in cognition: The costs and a benefit. *Psychonomic Bulletin & Review*, *6*(1), 142–147. https://doi.org/10.3758/BF03210822

Mazuski, C., Abel, J. H., Chen, S. P., Hermanstyne, T. O., Jones, J. R., Simon, T., Doyle, F. J., & Herzog, E. D. (2018). Entrainment of circadian rhythms depends on firing rates and neuropeptide release of VIP SCN neurons. *BioRxiv*, 322339. https://doi.org/10.1101/322339

McBee, M. (2010). Modeling Outcomes With Floor or Ceiling Effects: An Introduction to the Tobit Model. *Gifted Child Quarterly*, *54*(4), 314–320. https://doi.org/10.1177/0016986210379095

McElroy, T., & Dickinson, D. L. (2019). Thinking about complex decisions: How sleep and time-of-day influence complex choices. *Consciousness and Cognition*, *76*, 102824. https://doi.org/10.1016/j.concog.2019.102824

McMakin, D. L., Olino, T. M., Porta, G., Dietz, L. J., Emslie, G., Clarke, G., Wagner, K. D., Asarnow, J. R., Ryan, N. D., Birmaher, B., Shamseddeen, W., Mayes, T., Kennard, B., Spirito, A., Keller, M., Lynch, F. L., Dickerson, J. F., & Brent, D. A. (2012). Anhedonia Predicts Poorer Recovery Among Youth With Selective Serotonin Reuptake Inhibitor Treatment–Resistant Depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(4), 404–411. https://doi.org/10.1016/j.jaac.2012.01.011

McNab, F., Zeidman, P., Rutledge, R. B., Smittenaar, P., Brown, H. R., Adams, R. A., & Dolan, R. J. (2015). Age-related changes in working memory and the ability to ignore distraction. *Proceedings of the National Academy of Sciences*, 201504162. https://doi.org/10.1073/pnas.1504162112

McNamara, J. M., Mace, R. H., & Houston, A. I. (1987). Optimal daily routines of singing and foraging in a bird singing to attract a mate. *Behavioral Ecology and Sociobiology*, *20*(6), 399–405. https://doi.org/10.1007/BF00302982

Merikangas, K., Wicki, W., & Angst, J. (1994). Heterogeneity of Depression: Classification of Depressive Subtypes by Longitudinal Course. *The British Journal of Psychiatry : The Journal of Mental Science*, *164*, 342–348. https://doi.org/10.1192/bjp.164.3.342

Michely, J., Eldar, E., Erdman, A., Martin, I. M., & Dolan, R. J. (2020). SSRIs modulate asymmetric learning from reward and punishment. *BioRxiv*. https://doi.org/10.1101/2020.05.21.108266

Moore, M. T., & Fresco, D. M. (2012). Depressive realism: A meta-analytic review. *Clinical Psychology Review*, *32*(6), 496–509. https://doi.org/10.1016/j.cpr.2012.05.004

Morina, N., Deeprose, C., Pusowski, C., Schmid, M., & Holmes, E. A. (2011). Prospective mental imagery in patients with major depressive disorder or anxiety disorders. *Journal of Anxiety Disorders*, *25*(8), 1032–1037. https://doi.org/10.1016/j.janxdis.2011.06.012

Murphy, S. E., Longhitano, C., Ayres, R. E., Cowen, P. J., Harmer, C. J., & Rogers, R. D. (2008). The role of serotonin in nonnormative risky choice: The effects of tryptophan supplements on the "reflection effect" in healthy adult volunteers. *Journal of Cognitive Neuroscience*, *21*(9), 1709–1719. https://doi.org/10.1162/jocn.2009.21122

Naughton, M., Mulrooney, J. B., & Leonard, B. E. (2000). A review of the role of serotonin receptors in psychiatric disorders. *Human Psychopharmacology: Clinical and Experimental*, *15*(6), 397–415. https://doi.org/10.1002/1099-1077(200008)15:6<397::AID-HUP212>3.0.CO;2-L

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *Journal of Neuroscience*, *32*(2), 551–562. https://doi.org/10.1523/JNEUROSCI.5498-10.2012

Niv, Yael, Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–520. https://doi.org/10.1007/s00213-006-0502-4

Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta Earthquake. *Journal of Personality and Social Psychology*, *61*(1), 115.

Nusslock, R., & Alloy, L. B. (2017). Reward processing and mood-related symptoms: An RDoC and translational neuroscience perspective. *Journal of Affective Disorders*, *216*, 3–16. https://doi.org/10.1016/j.jad.2017.02.001

Orban, C., Kong, R., Li, J., Chee, M. W. L., & Yeo, B. T. T. (2020). Time of day is associated with paradoxical reductions in global signal fluctuation and functional connectivity. *PLOS Biology*, *18*(2), e3000602. https://doi.org/10.1371/journal.pbio.3000602

Otto, A. R., Fleming, S. M., & Glimcher, P. W. (2016). Unexpected but Incidental Positive Outcomes Predict Real-World Gambling. *Psychological Science*, *27*(3), 299–311. https://doi.org/10.1177/0956797615618366

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, *21*(6), 425–433. https://doi.org/10.1016/j.tics.2017.03.011

Parekh, P. K., & McClung, C. A. (2016). Circadian mechanisms underlying reward-related neurophysiology and synaptic plasticity. *Frontiers in Psychiatry*, *6*, 187. https://doi.org/10.3389/fpsyt.2015.00187

Paulus, M. P., Huys, Q. J. M., & Maia, T. V. (2016). A Roadmap for the Development of Applied Computational Psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(5), 386–392. https://doi.org/10.1016/j.bpsc.2016.05.001

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience*, *9*(12), 947–957. https://doi.org/10.1038/nrn2513

Paykel, E. S., Abbott, R., Jenkins, R., Brugha, T. S., & Meltzer, H. (2000). Urban–rural mental health differences in Great Britain: Findings from the National Morbidity Survey. *Psychological Medicine*, *30*(2), 269–280. https://doi.org/10.1017/S003329179900183X

Perlman, S. (2020). Another Decade, Another Coronavirus. *New England Journal of Medicine*, *382*(8), 760–762. https://doi.org/10.1056/NEJMe2001126

Pizzagalli, D. A. (2014). Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model. *Annual Review of Clinical Psychology*, *10*(1), 393–423. https://doi.org/10.1146/annurev-clinpsy-050212-185606

Pooseh, S., Bernhardt, N., Guevara, A., Huys, Q. J. M., & Smolka, M. N. (2018). Value-based decision-making battery: A Bayesian adaptive approach to assess impulsive and risky behavior. *Behavior Research Methods*, *50*(1), 236–249. https://doi.org/10.3758/s13428-017-0866-x

Prelec, D. (1998). The Probability Weighting Function. *Econometrica*, *66*(3), 497–527. https://doi.org/10.2307/2998573

Preston, S. D., Buchanan, T. W., Stansfield, R. B., & Bechara, A. (2007). Effects of anticipatory stress on decision making in a gambling task. *Behavioral Neuroscience*, *121*(2), 257–263. https://doi.org/10.1037/0735-7044.121.2.257

Pulcu, E., & Browning, M. (2017). *Affective Bias as a Rational Response to the Statistics of Rewards and Punishments*. https://doi.org/10.1101/114991

Pushkarskaya, H., Sobowale, K., Henick, D., Tolin, D. F., Anticevic, A., Pearlson, G. D., Levy, I., Harpaz-Rotem, I., & Pittenger, C. (2019). Contrasting contributions of anhedonia to obsessive-compulsive, hoarding, and post-traumatic stress disorders. *Journal of Psychiatric Research*, *109*, 202–213. https://doi.org/10.1016/j.jpsychires.2018.11.029

Rao, M. L., Gross, G., Strebel, B., Halaris, A., Huber, G., Bräunig, P., & Marler, M. (1994). Circadian rhythm of tryptophan, serotonin, melatonin, and pituitary hormones in schizophrenia. *Biological Psychiatry*, *35*(3), 151–163. https://doi.org/10.1016/0006-3223(94)91147-9

Richard, A., Rohrmann, S., Vandeleur, C. L., Schmid, M., Barth, J., & Eichholzer, M. (2017). Loneliness is adversely associated with physical and mental health and lifestyle factors: Results from a Swiss national survey. *PLOS ONE*, *12*(7), e0181442. https://doi.org/10.1371/journal.pone.0181442

Rogers, R. D., Tunbridge, E. M., Bhagwagar, Z., Drevets, W. C., Sahakian, B. J., & Carter, C. S. (2003). Tryptophan depletion alters the decision-making of healthy volunteers through altered processing of reward cues. *Neuropsychopharmacology*, *28*(1), 153–162. https://doi.org/10.1038/sj.npp.1300001

Rosenbaum, G. M., & Hartley, C. A. (2019). Developmental perspectives on risky and impulsive choice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1766), 20180133. https://doi.org/10.1098/rstb.2018.0133

Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Ruggeri, K., Alí, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., Davison, C., Demić, E., Esteban-Serna, C., Friedemann, M., Gibson, S. P., Jarke, H., Karakasheva, R.,

Khorrami, P. R., Kveder, J., Andersen, T. L., Lofthus, I. S., McGill, L., Nieto, A. E., … Folke, T. (2020). Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, *4*(6), 622–633. https://doi.org/10.1038/s41562-020-0886-x

Russek, E. M., Moran, R., McNamee, D., Reiter, A., Liu, Y., Dolan, R. J., & Huys, Q. J. M. (2020). Opportunities for emotion and mental health research in the resource-rationality framework. *Behavioral and Brain Sciences*, *43*. https://doi.org/10.1017/S0140525X19001663

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252–12257. https://doi.org/10.1073/pnas.1407535111

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2015). Dopaminergic modulation of decision making and subjective well-being. *Journal of Neuroscience*, *35*(27), 9811–9822. https://doi.org/10.1523/JNEUROSCI.0702-15.2015

Rutledge, Robb B., & Adams, R. A. (2017). Computational Psychiatry. In *Computational Models of Brain and Behavior* (pp. 29–42). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119159193.ch3

Rutledge, Robb B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., Lam, J., Skandali, N., Siegel, J. Z., Ousdal, O. T., Prabhu, G., Dayan, P., Fonagy, P., & Dolan, R. J. (2017). Association of neural and emotional impacts of reward prediction errors with major depression. *JAMA Psychiatry*, *74*(8), 790. https://doi.org/10.1001/jamapsychiatry.2017.1713

Rutledge, Robb B., Smittenaar, P., Zeidman, P., Brown, H. R., Adams, R. A., Lindenberger, U., Dayan, P., & Dolan, R. J. (2016). Risk taking for potential reward decreases across the lifespan. *Current Biology*, *26*(12), 1634–1639. https://doi.org/10.1016/j.cub.2016.05.017

Samanez-Larkin, G. R., & Knutson, B. (2015). Decision making in the ageing brain: Changes in affective and motivational circuits. *Nature Reviews Neuroscience*, *16*(5), 278–289. https://doi.org/10.1038/nrn3917

Scheier, M. E., & Carver, C. S. (1987). Dispositional Optimism and Physical Well-Being: The Influence of Generalized Outcome Expectancies on Health. *Journal of Personality*, *55*(2), 169–210. https://doi.org/10.1111/j.1467-6494.1987.tb00434.x

Schoenbaum, M. (1997). Do smokers understand the mortality effects of smoking? Evidence from the Health and Retirement Survey. *American Journal of Public Health*, *87*(5), 755–759. https://doi.org/10.2105/AJPH.87.5.755

Schulreich, S., Heussen, Y. G., Gerhardt, H., Mohr, P. N., Binkofski, F. C., Koelsch, S., & Heekeren, H. R. (2014). Music-evoked incidental happiness modulates probability weighting during risky lottery choices. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00981

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Shah, P., Harris, A. J. L., Bird, G., Catmur, C., & Hahn, U. (2016). A pessimistic view of optimistic belief updating. *Cognitive Psychology*, *90*, 71–127. https://doi.org/10.1016/j.cogpsych.2016.05.004

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), R941–R945. https://doi.org/10.1016/j.cub.2011.10.030

Sharot, T. (2012). *The Optimism Bias: Why we're wired to look on the bright side*. Hachette UK.

Sharot, T., & Garrett, N. (2016). Forming Beliefs: Why Valence Matters. *Trends in Cognitive Sciences*, *20*(1), 25–33. https://doi.org/10.1016/j.tics.2015.11.002

Silvia, P. J., Nusbaum, E. C., Eddington, K. M., Beaty, R. E., & Kwapil, T. R. (2014). Effort Deficits and Depression: The Influence of Anhedonic Depressive Symptoms on Cardiac Autonomic Activity During a Mental Challenge. *Motivation and Emotion*, *38*(6), 779–789. https://doi.org/10.1007/s11031-014-9443-0

Smittenaar, P., Rutledge, R. B., Zeidman, P., Adams, R. A., Brown, H., Lewis, G., & Dolan, R. J. (2015). Proactive and Reactive Response Inhibition across the Lifespan. *PLOS ONE*, *10*(10), e0140383. https://doi.org/10.1371/journal.pone.0140383

Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F., & Phelps, E. A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. *Proceedings of the*

National Academy of Sciences, *106*(13), 5035–5040. https://doi.org/10.1073/pnas.0806761106

Sokol-Hessner, Peter, & Rutledge, R. B. (2019). The psychological and neural basis of loss aversion. *Current Directions in Psychological Science*, *28*(1), 20–27. https://doi.org/10.1177/0963721418806510

Solomon, D. A. (2000). Multiple Recurrences of Major Depressive Disorder. *American Journal of Psychiatry*, *157*(2), 229–233. https://doi.org/10.1176/appi.ajp.157.2.229

Spencer, S., Falcon, E., Kumar, J., Krishnan, V., Mukherjee, S., Birnbaum, S. G., & McClung, C. A. (2013). Circadian genes Period 1 and Period 2 in the nucleus accumbens regulate anxiety-related behavior. *The European Journal of Neuroscience*, *37*(2), 242–250. https://doi.org/10.1111/ejn.12010

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). State-trait anxiety inventory manual. *Mind Garden, Inc*.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092. https://doi.org/10.1001/archinte.166.10.1092

Stopper, C. M., Tse, M. T. L., Montes, D. R., Wiedman, C. R., & Floresco, S. B. (2014). Overriding phasic dopamine signals redirects action selection during risk/reward decision making. *Neuron*, *84*(1), 177–189. https://doi.org/10.1016/j.neuron.2014.08.033

Story, G. W., Vlaev, I., Seymour, B., Winston, J. S., Darzi, A., & Dolan, R. J. (2013). Dread and the Disvalue of Future Pain. *PLOS Computational Biology*, *9*(11), e1003335. https://doi.org/10.1371/journal.pcbi.1003335

Strauss, G. P., & Gold, J. M. (2012). A New Perspective on Anhedonia in Schizophrenia. *American Journal of Psychiatry*, *169*(4), 364–373. https://doi.org/10.1176/appi.ajp.2011.11030447

Strunk, D. R., Lopez, H., & DeRubeis, R. J. (2006). Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behaviour Research and Therapy*, *44*(6), 861–882. https://doi.org/10.1016/j.brat.2005.07.001

Swiecicki, L., Zatorski, P., Bzinkowska, D., Sienkiewicz-Jarosz, H., Szyndler, J., & Scinska, A. (2009). Gustatory and olfactory function in patients with unipolar and bipolar depression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *33*(5), 827–834. https://doi.org/10.1016/j.pnpbp.2009.03.030

Taquet, M., Quoidbach, J., de Montjoye, Y.-A., Desseilles, M., & Gross, J. J. (2016). Hedonism and the choice of everyday activities. *Proceedings of the National Academy of Sciences*, *113*(35), 9769–9773. https://doi.org/10.1073/pnas.1519998113

Taquet, M., Quoidbach, J., Gross, J. J., Saunders, K. E. A., & Goodwin, G. M. (2020). Mood Homeostasis, Low Mood, and History of Depression in 2 Large Population Samples. *JAMA Psychiatry*, *77*(9), 944–951. https://doi.org/10.1001/jamapsychiatry.2020.0588

Teicher, M. H., Glod, C. A., Magnus, E., Harper, D., Benson, G., Krueger, K., & McGreenery, C. E. (1997). Circadian rest-activity disturbances in seasonal affective disorder. *Archives of General Psychiatry*, *54*(2), 124–130. https://doi.org/10.1001/archpsyc.1997.01830140034007

Teki, S., Kumar, S., & Griffiths, T. D. (2016). Large-Scale Analysis of Auditory Segregation Behavior Crowdsourced via a Smartphone App. *PLOS ONE*, *11*(4), e0153916. https://doi.org/10.1371/journal.pone.0153916

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, *26*(1), 24–36. JSTOR. https://doi.org/10.2307/1907382

Todd Maddox, W., & Markman, A. B. (2010). The Motivation–Cognition Interface in Learning and Decision Making. *Current Directions in Psychological Science*, *19*(2), 106–110. https://doi.org/10.1177/0963721410364008

Treffers, T., Koellinger, P. D., & Picot, A. (2016). Do Affective States Influence Risk Preferences? *Schmalenbach Business Review*, *17*(3), 309–335. https://doi.org/10.1007/s41464-016-0018-3

Tsanas, A., Saunders, K. E. A., Bilderbeck, A. C., Palmius, N., Osipov, M., Clifford, G. D., Goodwin, G. M., & De Vos, M. (2016). Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *Journal of Affective Disorders*, *205*, 225–233. https://doi.org/10.1016/j.jad.2016.06.065

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, Amos, & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

Tymula, A., Rosenberg Belmaker, L. A., Ruderman, L., Glimcher, P. W., & Levy, I. (2013). Like cognitive function, decision making across the life span shows profound age-related changes. *Proceedings of the National Academy of Sciences*, *110*(42), 17143–17148. https://doi.org/10.1073/pnas.1309909110

Uher, R., Perlis, R. H., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., Hauser, J., Dernovsek, M. Z., Souery, D., Bajs, M., Maier, W., Aitchison, K. J., Farmer, A., & McGuffin, P. (2012). Depression symptom dimensions as predictors of antidepressant treatment outcome: Replicable evidence for interest-activity symptoms. *Psychological Medicine*, *42*(5), 967–980. https://doi.org/10.1017/S0033291711001905

van den Bos, R., Homberg, J., & de Visser, L. (2013). A critical review of sex differences in decision-making tasks: Focus on the Iowa Gambling Task. *Behavioural Brain Research*, *238*, 95–108. https://doi.org/10.1016/j.bbr.2012.10.002

van den Bos, W., & Hertwig, R. (2017). Adolescents display distinctive tolerance to ambiguity and to uncertainty during risky decision making. *Scientific Reports*, *7*(1), 40962. https://doi.org/10.1038/srep40962

Verhagen, S. J. W., Hasmi, L., Drukker, M., van Os, J., & Delespaul, P. A. E. G. (2016). Use of the experience sampling method in the context of clinical trials: Table 1. *Evidence Based Mental Health*, *19*(3), 86–89. https://doi.org/10.1136/ebmental-2016-102418

Villano, W., Otto, A., Ezie, C., Gillis, R., & Heller, A. (2020). Temporal dynamics of real-world emotion are more strongly linked to prediction error than outcome. *Journal of Experimental Psychology: General*, *149*. https://doi.org/10.1037/xge0000740

Vinckier, F., Rigoux, L., Oudiette, D., & Pessiglione, M. (2018a). Neuro-computational account of how mood fluctuations arise and affect decision making. *Nature Communications*, *9*(1), 1708. https://doi.org/10.1038/s41467-018-03774-z

Vinckier, F., Rigoux, L., Oudiette, D., & Pessiglione, M. (2018b). Neuro-computational account of how mood fluctuations arise and affect decision making. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03774-z

Vrieze, E., Pizzagalli, D. A., Demyttenaere, K., Hompes, T., Sienaert, P., de Boer, P., Schmidt, M., & Claes, S. (2013). Reduced reward learning predicts outcome in major depressive disorder. *Biological Psychiatry*, *73*(7), 639–645. https://doi.org/10.1016/j.biopsych.2012.10.014

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology*, *104*(1), 3.

Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263–290. https://doi.org/10.1002/bdm.414

Weinstein, N. D. (1987). Unrealistic optimism about susceptibility to health problems: Conclusions from a community-wide sample. *Journal of Behavioural Medicine*, *10*(5), 481–500.

Werner-Seidler, A., & Moulds, M. L. (2011). Autobiographical memory characteristics in depression vulnerability: Formerly depressed individuals recall less vivid positive memories. *Cognition & Emotion*, *25*(6), 1087–1103. https://doi.org/10.1080/02699931.2010.531007

Wichers, M. C., Barge-Schaapveld, D. Q. C. M., Nicolson, N. A., Peeters, F., de Vries, M., Mengelers, R., & van Os, J. (2009). Reduced Stress-Sensitivity or Increased Reward Experience: The Psychological Mechanism of Response to Antidepressant Medication. *Neuropsychopharmacology*, *34*(4), 923–931. https://doi.org/10.1038/npp.2008.66

Wilson, T. D., & Gilbert, D. T. (2005). Affective Forecasting: Knowing What to Want. *Current Directions in Psychological Science*, *14*(3), 131–134. https://doi.org/10.1111/j.0963-7214.2005.00355.x

Wilson, T. D., Meyers, J., & Gilbert, D. T. (2003). “How Happy Was I, Anyway?” A Retrospective Impact Bias. *Social Cognition*, *21*(6), 421–446. https://doi.org/10.1521/soco.21.6.421.28688

Winterson, J. (2013). *Written on the Body*. Random House.

Yang, Y., Wang, H., Hu, J., & Hu, H. (2018). Lateral habenula in the pathophysiology of depression. *Current Opinion in Neurobiology*, *48*, 90–96. https://doi.org/10.1016/j.conb.2017.10.024

Yechiam, E., & Ert, E. (2011). Risk attitude in decision making: In search of trait-like constructs. *Topics in Cognitive Science*, *3*(1), 166–186. https://doi.org/10.1111/j.1756-8765.2010.01126.x

Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin*, *139*(2), 497–518. https://doi.org/10.1037/a0029383

Yuen, K. S. L., & Lee, T. M. C. (2003). Could mood state affect risk-taking decisions? *Journal of Affective Disorders*, *75*(1), 11–18. https://doi.org/10.1016/S0165-0327(02)00022-8