A big-data and experimental analysis of the relationship between time perspective, and anxiety and depression

Cole Robertson*

Radboud University

James Carney

The London Interdisciplinary School

Shane Trudell

TKTK

Additional authors: TBC

Additional affiliations: TBC

Author Note

*corresponding author: The Center for Language Studies, Radboud University, PO Box 9103 6500 HD, Nijmegen, cbjrobertson@gmail.com

Abstract

Anxiety and depression are mood disorders which negatively impact the lives of many. Future time horizons, or how "far" into the future people tend to think, are a significant marker of depression, while conflicting results suggest that how much people devalue delayed future rewards "temporal discounting" may be a significant marker of anxiety. At the same time, cross-linguistic differences in future time reference grammar, individual differences in future time reference language use, and temporal horizons have been found to predict temporal discounting. However, as far as we know, no one has investigated whether future time reference use is a significant marker of anxiety and/or depression. In study 1, we therefore analysed data from the social media website Reddit to explore any potential relationships between these variables. We found that users who had posted popular contributions to forums dedicated to anxiety and depression referenced the future and the past more often than controls, had more proximal future and past time horizons, and significantly differed in their linguistic future time reference patterns. They used fewer future tense constructions (e.g. *will*), more low-certainty constructions (e.g. *possibly, could*), more bouletic modal constructions (e.g. *hope, wish*) and more deontic modal constructions (e.g. *must, need to*). These findings motivated study 2, a survey-based mediation analysis in which we found participants who scored high in anxiety represented the future as more temporally distal and therefore temporally discounted to a greater degree. The same was not true of depression. We suggest such methods which combine descriptive big-data approaches with experimental paradigms may help identify novel markers of mental illness, including anxiety and depression, which can then aid in the development of new therapies and diagnostic criteria.

Keywords: future time reference, temporal discounting, mood disorders, anxiety, depression, big-data, time perspective

1 Introduction

Millions of people suffer from mental illness world wide. For example, in North America the prevalence of mental health disorders in adults may be as high as 25%. In addition to the impacts on well-being – important in their own right – the estimated economic costs of mental illness in North America are as much as US$83.1bn (Dewa & McDaid, 2011). At the same time, medical science has struggled to predict the onset of mental illness from physiological markers which has made preventative medicine difficult to implement (Kapur, Phillips, & Insel, 2012). A potentially promising avenue of research has been to use data from social media to predict various aspects of personality, including the onset and diagnosis of mental illness (Guntuku, Yaden, Kern, Ungar, & Eichstaedt, 2017).

Such approaches have met with success. For instance, past posts on the social media site Reddit were successfully used to predict whether a user would later post to a forum dedicated to mental illness (Thorstad & Wolff, 2019); tweets made by Twitter users who had self-diagnosed as suffering from depression or post traumatic stress disorder were successfully classified above chance levels (Coppersmith, Dredze, Harman, Hollingshead, & Mitchell, 2015); and posts to 11 different Reddit forums dedicated to mental illness were successfully differentiated with an average of 72% precision (Gkotsis et al., 2017). For general reviews of these approaches see (Guntuku et al., 2017; Goldstone & Lupyan, 2016).

Many predictive studies have tended to implement "black box" machine learning algorithms, which focus on predicting mental health outcomes from prior social media data. While predictive approaches are undoubtedly valuable, model parameters may not always be usefully interpreted by human practitioners (Rai, 2020). Another strand of research therefore seeks to complement predictive approaches with descriptive ones which attempt to generate psychological insights from the analysis of social media data. For instance, Thorstad and Wolff (2019) used per-word model weights to understand which words were most predictive of a user later posting to a Reddit forum focused on depression; and De Choudhury, Counts, Horvitz, and Hoff (2014) report language use differences on Facebook posts as a function of whether survey respondents had been clinically diagnosed with postpartum depression.

In the present study we therefore implemented a hybrid methodology, using a combination of machine-learning and descriptive natural language processing to explore the complex relationships between future time horizons, linguistic future time reference patterns, and anxiety and depression. The choice of these variables was motivated by previous research which suggests that how people construe future events – their future time perspective – may be a key marker of anxious and/or depressive tendencies (Dilling & Rabin, 1967; Pulcu et al., 2014; Steinglass et al., 2017).

Future time perspective encompasses how people tend think about and relate to future events. It includes questions such as: do people tend to save for the future or spend resources now, do they conceive of future events as proximal or distal, and do they tend to imagine temporally proximal or distal future events? Measures of future time perspective have been found to be significant markers of anxiety (Papastamatelou, Unger, Giotakos, & Athanasiadou, 2015), major depressive disorder (Pulcu et al., 2014) and social anxiety disorder (Steinglass et al., 2017).

Particularly, people suffering from depression have been found to have shorted "future time horizons" relative to controls (Dilling & Rabin, 1967). This is usually measured using the Wallace task (Wallace, 1956). In this task, participants are asked to think of 10 events that are likely to occur in their future lives, and what age they expect to be when these events will occur. Their future time horizon is then calculated by subtracting their current age from the ages they give. This approximates how far into the future participants tend to imagine themselves.

Pathological gamblers (Hodgins & Engel, 2002), alcoholics (Smart, 1968), and heroin addicts (Petry, Bickel, & Arnett, 1998) were all found to have shortened future time horizons compared with healthy controls. While these studies do not explicitly study depression, substance abuse disorders and depression co-occur often enough to warrant potential connections between these results and underlying depressive tendencies (Barrault & Varescon, 2013; Blaszczynski & McConaghy, 1989; Felton et al., 2020).

Since anxiety tends to involve excessive "anticipation of future threat" (American Psychiatric Association, 2015, p. 1), the subjective experience of future events seems almost inherently salient to the experience of anxiety. However, the relationship between anxiety and future time perspective seems less well understood then for depression.

Most research into anxiety and time perspective has used temporal discounting paradigms. Temporal discounting captures the extent to which people devalue delayed future rewards. In general, people tend to devalue delayed rewards as a function of wait time, but the extent of such devaluation differs between individuals (Green & Myerson, 2004). These differences are usually measured by giving participants a series of binary choices between immediate and delayed rewards, for instance $10 now vs. $20 in a month. "Present-oriented" people tend to devalue future outcomes and would therefore opt for the immediate $10, while "future-oriented" people tend to devalue less steeply and would therefore opt to wait for the $20. These measures are not unrelated to time horizons; shorter time horizons tend to be associated with increased discounting (Thorstad & Wolff, 2018; Thorstad, Nie, & Wolff, 2015), and researchers are increasingly understanding temporal discounting processes in terms of subjective experiences of future time (Kim & Zauberman, 2009; Zauberman, Kim, Malkoc, & Bettman, 2009).

While people suffering from major depressive disorder have been found to temporally discount more than healthy individuals (Felton et al., 2020; Pulcu et al., 2014), the results around anxiety have been mixed. Some findings suggest people suffering from social anxiety discount more (Rounds, Beck, & Grant, 2007), while others report the effect goes in the opposite direction (Steinglass et al., 2017), or that there is no significant relationship (Jenks & Lawyer, 2015). This suggests more research may be needed to untangle how future time perspective may relate to the experience of anxiety.

To add further complexity to this nexus of relationships, a growing body of research indicates that linguistic FTR may be an important factor in shaping future time perspective. FTR is a catch-all term used to refer to any linguistic statement which refers to future events, whether future-tensed or not (Dahl, 1985, 2000). For instance, “it could rain tomorrow, “it will rain tomorrow”, and “I hope it rains tomorrow” are all FTR. A growing body of research suggests that how languages differ in terms of their FTR grammars may impact speakers’ future time perspective as indicated by temporal discounting tendencies. Chen (2013) provides a framework for how this might occur. He hypothesized that languages which oblige speakers to use the future tense for FTR might cause speakers to perceive the future as farther away and therefore discount more. To get a feel for this, note that in English, speakers are obliged to use *will* in the sentence “it *will* rain tomorrow.” On the other hand, speakers of Dutch are free to use the present tense, ”Het regent morgen”, ‘it rains tomorrow.’ Chen (2013) theorized that consistently using the present tense for FTR would essentially collapse future into present time, effectively causing speakers to construe objective future dates as subjectively more proximal and therefore temporally discount less (i.e. be more willing to wait for subjectively more proximal future rewards). To support this hypothesis, he found that speakers of languages like Dutch were more likely to have saved each year, and retired with more assets. Since then, multiple other studies have found that speakers of languages like Dutch tend to behave as though they discount less (Chi, Su, Tang, & Xu, 2018; Falk et al., 2018; Fasan, Gotti, Kang, & Liu, 2016; Figlio, Giuliano, Özek, & Sapienza, 2016; Hübner & Vannoorenberghe, 2015b, 2015a; Liang, Marquis, Renneboog, & Sun, 2018; Mavisakalyan, Tarverdi, & Weber, 2018; Pérez & Tavits, 2017; Zhu, Hu, Wang, & Zheng, 2020).

However, resent typological research into cross-linguistic differences in FTR grammar has found that linguistic FTR may involve not just binary differences in the extent to which the future tense is obliged, but differences in degree in the extent to which the encoding of the future probability is obliged (Robertson & Roberts, 2020). For instance, it is not actually obligatory to use *will* in the sentence “it *will* rain tomorrow”. Rather English grammar obliges speakers to use one of “it *could/may/might/should* rain tomorrow.” This encoding of probability is referred to as “modality”. In unpublished results (chapters 3,4) we have found that the obligatory use of low-certainty FTR terminology in English may actually be driving observed differences in temporal discounting between English and Dutch speakers. These findings suggest that future time perspective may not simply involve how far into the future people tend to think, but how certain they tend to be about future events. Moreover, we have found (chapters 3,4) that individual-level differences in FTR usage is a strong predictor of temporal discounting behavior, indicating that FTR may reflect latent beliefs about the probability and/or temporal distance of future events.

Despite this, and evidence that future time perceptive may be related to anxiety and/or depression, the relationship between FTR habits and anxiety and depression has not been studied. We therefore undertook an exploratory study which sought to use big-data methodology to understand whether people who had posted to Reddit forums dedicated to anxiety and depression differed from a random sample in terms of their FTR habits and future time horizons.

2 Study 1

The basis of study 1 is an analysis of data downloaded from Reddit. Reddit is a social media platform with over 430m average monthly users (*https://www.redditinc.com/advertising/audience*, accessed December, 2020). The site is similar to a message board or traditional forum; users ("redditors"), communicate in forums ("subreddits") dedicated to specific topics. There are no character limits to user contributions ("posts"), entailing that users are free to comment in as much detail as they desire. Subreddits including topic focus and rules on acceptable behavior are created and self-monitored by moderator teams comprised of leading subreddit members. Comments are "upvoted" and "downvoted", and can be sorted by various criteria, for instance most upvoted of all time. The names of subreddits usually reflect their topical focus and are prefaced with *r/*, for instance, *r/news* is dedicated to sharing US and international news stories.

The largest subreddits dedicated to anxiety and depression are *r/Anxiety* and *r/depression*. In December, 2020, *r/Anxiety* had 417k members, and *r/depression* had 712k members (*reddit.com*, accessed December, 2020). Both are focused on providing support and advice for people dealing with issues related to anxiety and depression (respectively), and both allow posts by sufferers or those close to them.

2.1 Methods and materials

2.1.1 Data acquisition. Data were downloaded between TKTK and TKTK, TKTK, and ethical approval for the study was granted by the Brunel University Research Ethics Committee, ref. no. TKTK.

We were concerned that language used when posting to *r/Anxiety* and *r/depression* might differ from a random sample not due to the habits of speech of the posters, but due to rules of discourse imposed by the moderators. We therefore implemented a "snowball" sampling approach, first constructing a "seed" sample of redditors who had posted to *r/Anxiety* and *r/depression*, and then a final sample comprised of these same redditors' top posts in any subreddit.

Data were obtained using PRAW, the Reddit Python Reddit API Wrapper (`https://praw.readthedocs.io/en/latest/#`). This interface allows contributions and contribution metadata to be programmatically extracted by way of specific subreddits, Redditors, and comments. We extracted our data on October 30 2019 by taking the "hottest" 1,000 posts on the the anxiety and depression subreddits. "Hot", in this connection, is part of the Reddit proprietary algorithm for ranking content. It is calculated by taking the base 10 log of the net number of upvotes and adding it to the number of 12.5 hour periods that have passed since the first ever Reddit post. This means that, for each successive 12.5 hour period, a post must get ten times the number of upvotes than it had in the previous 12.5 hour period to be as "hot" as something 12.5 hours younger. The value of the "hot" metric is that it captures content specific to a given subreddit : measures like "top" that count total number of upvotes are disproportionately captured by stickied content that is permanently visible.

Specifically, we used PRAW, the Python Reddit API Wrapper, to download up to the top $1,000$ "hot" ranked submissions in each of *r/Anxiety* ($n = 910$ posts) and *r/depression* ($n = 902$ posts). (As scores for user-deleted submissions are still returned by the API, the number of usable posts was less than than 1,000 rate-limit cap.) "Hot", in this connection, is part of the Reddit algorithm for ranking content. It is calculated by taking the base 10 log of the net number of upvotes and adding it to the number of 12.5 hour periods that have passed since the first ever Reddit post. This means that, for each successive 12.5 hour period, a post must get ten times the number of upvotes than it had in the previous 12.5 hour period to retain the "hot" score it had when it was 12.5 hours younger. The value of the "hot" metric is that it captures topical content specific to a given subreddit: other measures like "top" that count total number of upvotes are disproportionately captured by stickied content that is permanently visible, and hence less topical.

We then downloaded up to the top $1,000$ most recent posts for each unique redditor in these data. This resulted in $n = 224,009$ posts made by redditors who had posted popular content in *r/Anxiety* (the "anxiety" condition), and $n = 182,555$ posts made by redditors who had posted popular content in *r/depression* (the "depression" condition). However, we found no differences between these conditions on any measure, and so combined these into a single "mental health" condition. We constructed a control sample of $n = 213,132$ posts by repeating an identical process with the *r/All* subreddit, which is a compendium of all content generated on Reddit.

This meant our final sample was comprised of $N = 619,696$ posts, drawn from $N = 13,651$ unique subreddits ($n = 6,978$ [mental health], $n = 2,781$ [control], and $n = 3,856$ crossovers), and $N = 4,205$ unique redditors ($n = 1,810$ [mental health], $n = 2,395$ [control], no crossovers). It should be noted that this process resulted in some posts from the original seed subreddits still being retained in the final sample ($n = 10,145$ posts from *r/Anxiety*, and $n = 17,319$ from *r/depression*). In unreported analyses, we found excluding these data did not substantively change any results, i.e. cause significance to cross critical thresholds or coefficients to change sign, so we did not exclude them.

2.1.2 Data preprocessing. Inspection of the data indicated that due to their unrestricted length a non-negligible number of posts appeared to contain references to both past and future time. We therefore split posts into constituent sentences using a non-monotonic transition system dependency parsing algorithm system (Honnibal & Johnson, 2015) implemented Python (Python Software Foundation, 2017) in the open source natural language processing package *spaCy* (Explosion AI, 2020). This resulted in $N = 2,029,956$ sentences ($n = 1,561,713$ [mental health], $n = 468,243$ [control]).

2.1.3 Temporal horizons. To estimate temporal horizons, we used the SUTime temporal tagger (Chang & Manning, 2012). SUTime is a rule-based deterministic natural language tagging system which uses regular expressions which combine both key-words (e.g. "yesterday") and rules (e.g. "[DATE] at [TIME]") to identify temporal expressions and convert them to numerical data. It processes absolute dates (e.g. November 7, 2012), relative dates (e.g. "this Friday"), and combinations of these (e.g. "tomorrow at 3pm"). For relative dates, we used time of posting (available from the Reddit API) as the reference date. For instance, if a redditor referenced "tomorrow" on 10-06-2020, SUTime would return 11-06-2020. To calculate time horizons (days), we subtracted time of posting from the SUTime reference values. In cases where there was more than one time reference in one sentence, we took the mean, i.e. $H(s) = \sum_r \frac{D(r)}{n}$; where $H(s)$ is the time horizon for a given s sentence, r is a temporal reference, $D(r)$ is the number of days from time of posting, and n is the number of temporal references in the sentence (Thorstad & Wolff, 2018). Such methods with data from Twitter have been found to positively and significantly correlate with participants' future time horizons as measured by the Wallace task, as well as predict both collective and individual intertemporal decision making (i.e. temporal discounting) (Thorstad & Wolff, 2018).

2.1.4 Time reference classification. To estimate whether sentences referred to the future (FTR) or past (PTR), we used as active learning paradigm implemented in *prodigy* (Montani & Honnibal, 2020) to train a machine learning "time reference" classifier. Active learning is a way of reducing the number of annotated examples machine learning models need to generate reliable predictions in supervised learning paradigms. In supervised learning, human raters annotate data and machine learning models are then trained to predict the annotations from the input data (in this case raw text). Active learning speeds up this process by keeping a machine learning model "in the loop" during the annotation process, and selecting the un-annotated data about which the model is most uncertain. This increases the number of annotations which are near the boundary criteria of the unknown function the model is attempting to approximate, and reduces the number of annotations the model is already able to confidently predict, thereby reducing the total number annotations needed. The authors used *prodigy* to annotate text examples for both FTR ($n = 4002$ [$n_{accept} = 1403$, $n_{reject} = 2406$, $n_{ignore} = 193$[1]]), and PTR ($n = 2382$ [$n_{accept} = 660$, $n_{reject} = 1704$, $n_{ignore} = 18$]). Because the active learning process resulted in some overlaps, this meant there were $N = 4,947$ unique text examples. For PTR, all annotations were drawn from our reddit data; for FTR annotations, were drawn from a combination of a Twitter dataset (see SM) and from our reddit data. Because we wanted to cast as wide a net as possible to better understand linguistic usage within the broad notional domains of future and past time reference, our annotation criteria were semantic rather then formal. Any statement which made FTR/PTR was classed positively, regardless of formal tensing. For instance, "they could win", "it will be ok", "it rains tomorrow", "I go in three weeks", "my train is arriving soon", and (in November) "Christmas is soon" would all be classed as FTR. Similarly, "it has rained", "I saw her", "it was tiresome", "I'm walking along the other day", and "I'm tired from my run this morning" would all be classed as PTR.

Using *prodigy*, we trained a model on these data using 80%/20% train/evaluate split, 10 iterations over the data, a dropout rate of 0.2, and exclusive categories. The *prodigy* text classification system wraps a *spaCy* ensemble model comprised bag of words model and a 1 dimensional convolution neutral net (CNN) where token vectors are calculated using a CNN, mean pooled and used as features in a feed forward network (Explosion AI, 2020). Our trained model can be loaded as a *spaCy* module in *Python* and is available at *URL TO COME*. Because not all the FTR annotations had been

[1]For indeterminable cases.

drawn from the reddit data, and because we wanted to make sure the model's out of sample performance was acceptable, we annotated two additional "test" datasets drawn exclusively from the reddit data ($N_{FTR} = 477$, and $N_{PTR} = 489$). Model performance on these withheld data was generally good, *Table 1.* In particular, we note that high precision indicates high agreement between positive predictions and ground truth.

2.1.5 FTR type classification. To establish FTR usage patterns, we used what we will refer to here as the FTR type classifier (Robertson, 2019). It is closed-vocabulary deterministic key-word based classification program the authors wrote in Python (Python Software Foundation, 2017). It classifies short future-referring text documents into 7 exclusive categories based on the presence of key-words which encode semantic domains most salient to FTR. One of these categories is further broken down according to the word classes of the keywords involved. See Robertson and Roberts (2020) for a full description and justification of these categories for English FTR. Each category is dichotomous, and coded with (1) to indicate a response is a positive example, otherwise (0).

Two categories are tense-based, and the other 5 are based on modal semantics. Modality generally involves quantifying what is possible and/or necessary relative to any of several common modal "bases" (Palmer, 2001). For instance, epistemic modality involves speakers expressing what they think is probable or unlikely relative to an epistemic "base", i.e. what they know and believe about the world (Nuyts, 2000); bouletic modality involves speakers quantifying what they think is possible or desirable relative to a desiderative base, i.e. what they hope or want to occur (Palmer, 2001); and deontic modality involves speakers quantifying what they think is obligatory or necessary relative to a normative base, i.e. what they think *must* or *should* occur (Palmer, 2001). With that in mind, the 7 categories are as follows:

THE FUTURE TENSE, for sentences which use the future tense without any other modal words, e.g. "tomorrow it *will/is going to/shall* rain" would be classed as future tense because it uses any of *will, is going to*, or *shall* and *not also* any modal class criterion words (see below).

THE PRESENT TENSE, for sentences which are FTR but fail to be classed in any other category, e.g. "tomorrow it rains" would be classed as present tense because it does not use the future tense or any other modal class criterion words (see below).

LOW-CERTAINTY EPISTEMIC MODALITY, for sentences which express low-certainty about future events. This category is further broken down depending on whether a modal verb or some other key-word class is used, e.g. "it *could/may/might/should* rain tomorrow" would all be classed as verbal low-certainty because they use the low-certainty modal verbs *could, may, might* or *should*, while "it will *possibly/probably/potentially* rain tomorrow" and "I *think* it will rain tomorrow" would be classed as lexical low-certainty because they express low-certainty without using a modal verb. The reason for this subdivision is that our typological research (Robertson & Roberts, 2020) as well as review of the literature (Nuyts, 2000), suggests that modal verb use is more constrained by features of English grammar, so we reasoned that individual differences in FTR usage would be more likely to manifest in the lexical categories which are freer to reflect speaker beliefs and intentions.

HIGH-CERTAINTY EPISTEMIC MODALITY, for sentences which express high-certainty. In English FTR, this domain is typically encoded using high-certainty epistemic adjectival or adverbial modifiers, i.e. "it will *definitely/certainty/absolutely* rain tomorrow" would be classed as lexical high-certainty.

DEONTIC MODALITY, for sentence which express necessity and/or obligation. In English FTR, this is usually accomplished using the modal verb *must* and/or *have to/need to* constructions (Nuyts, 2000), e.g. "you *must* come tomorrow" and "I *have to/need to* pick up more groceries tomorrow" would classed as deontic because they use the deontic modal verb *must*, and deontic *have to/need to* constructions.

BOULETIC MODALITY, for responses which express desires and hopes for the future, e.g. "I *hope* it rains tomorrow, "I *want* it wo rain tomorrow", and "I *wish* it would rain tomorrow" would all be classed as bouletic modality because they use the bouletic modal verbs *hope*, *want*, and *wish*.

IRREALIS, for responses which express counterfactual realities, e.g. "*If* it rains, I'll go out", "*if* he *would* only go out with me, I *would* be happy" would both be classed as irrealis because they express non-factual realities using the conditional *if*, and *would*.

Since the FTR type classifier is intended to classify based on semantics, responses in the present tense which expressed modal semantics (e.g. "tomorrow it could rain") were classed according to the modal notions, and not formal tense categories. Similarly, responses in the future tense but which expressed modal notions (e.g. "tomorrow it *will possibly* rain") were classed according to the modal notions and not the tense category. This is because it seems obvious to us that e.g. "it will possibly rain" and "it will rain" express different semantics and we wanted this to be reflected by our classification scheme.

To check FTR type classification accuracy on our Reddit data, we had trained assistants hand-annotate an evaluation dataset. To select data, we randomly selected from sentences for which $p(FTR) > 50\%$, $n = 50$ from each FTR type category, plus $n = 50$ uncoded sentences, and $n = 50$ randomly selected sentences (no overlaps). After exclusions for negations this resulted in $N = 447$ sentences. We checked predictions against this evaluation dataset, and updated the FTR type classification code to fix any discrepancies. In order to check that this had not resulted in over-fitting to the evaluation dataset, we repeated this process with a second test dataset, this time randomly selecting $n = 30$ sentences as described above ($N = 299$ total sentences). Accuracy metrics against this test dataset are presented in *Table 1*.

2.2 Results

2.2.1 Analytic approach. The data contained multiple sentences from individual posts, multiple posts from individual redditors, and multiple redditors across various subreddits. Using ordinary regressions which assume observations are independent might therefore have risked underestimating actual standard errors and increasing Type I error rates. Our analytic solution to this issue was to calculate the "design effect" (*deff*) for each potential source of within-group variation (posts, redditor, subreddits) for each dependent variable, and use this to ascertain whether it was necessary to employ mixed regression techniques to address these data structures. Kish (1965) defined the *deff* of some sample statistic as “the ratio of the actual variance of a sample to the variance of a simple random sample of the same number of elements” (p. 258). In mixed regression modeling, *deff* can be calculated using the ratio of within-group to total variance, (the Intraclass Correlation Coefficient ICC), and the average cluster size, c, such that (Muthén & Satorra, 1995):

$$deff = 1 + (c - 1) \times ICC \tag{1}$$

Deff is sensitive to either high $ICCs$ or high group sizes; a widely accepted cut-off criteria is that for $deff < 2$, error bias will be within 10% and no mixed modeling is necessary (Lai & Kwok, 2015). However, recent research suggests lower cut offs may be appropriate (Lai & Kwok, 2015).

*Deff*s for posts and subreddits were low, $1.01 \leq deff_{posts} \leq 1.12$, $1.03 \leq deff_{subred} \leq 1.27$. According to *Table 3* in Lai and Kwok (2015), this meant were justified in ignoring group-wise correlation by posts and subreddits and could expect standard errors to be biased by no more than $\approx 10\%$. However, *deff*s for redditors tended to be higher, $1.10 \leq deff_{red} \leq 4.48$. We therefore estimated multilevel regressions of the following general form, with random intercepts clustered by redditor:

$$y_{ij} = \beta_0 + \beta_1 m.heal_{ij} + u_j + e_{ij} \tag{2}$$

where the error, $e_{ij} \sim N(0, \sigma_e^2)$, and random terms, $u_j \sim N(0, \sigma_u^2)$, are assumed to be drawn from the normal distributions with means of 0 and standard deviations drawn from the sample. Our predictor of interest, $m.heal$, was a dummy for mental health (1) vs. control (0) condition.

2.2.2 Frequency of time reference. The first thing we were interested in was how much redditors made non-present time reference. To answer this question, we created a unified time reference variable based the time reference classifier; observations where $p(FTR/PTR) > 50\%$ were treated as positive (1) examples, otherwise (0). We then used mixed regressions with a logistic link function and random intercepts clustered by redditor to regress the resultant binary time reference variables over our condition dummy:

$$ln(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_0 + \beta_1 m.heal_{ij} + u_j \tag{3}$$

where π_{ij} is the probability that $y_{FTR/PTR} = 1$ for observation i and redditor j. For mixed effects models like these, standard calculation of R^2 is complicated by the fact that there are multiple variance components explained by the model (i.e. fixed: $\beta_0 + \beta_1 m.heal_{ij}$ and random: u_j). We therefore report $R^2_{marginal}$, which represents the variance explained by the fixed effects, and $R^2_{conditional}$, which represents the variance explained by the random effects + the fixed effects (Nakagawa, Johnson, & Schielzeth, 2017). For the FTR model, $R^2_m = 0.001$, $R^2_c = 0.019$ (where more than one method to calculate these quantities was available, we report the mean). This indicates that approximately 0.1% of the variance was explained by mental health condition. For the PTR model $R^2_m = 0.007$, $R^2_c = 0.05$, indicating approximately 0.7% of variance was explained.

We nonetheless found that redditors in the mental health condition were significantly more likely to make FTR, $e^\beta = 1.23$, $SE = 0.02$, $z = 10.41$, $p < .001$, and PTR, $e^\beta = 1.58$, $SE = 0.03$, $z = 17.11$, $p < .001$ (βs are exponentiated so represent changes in odds ratios). This indicates that redditors who had posted popular contributions to *r/Anxiety* and *r/depression* had approximately 23% higher odds of referencing the future and approximately 58% higher odds of referencing the past, see *Fig. 1*.

2.2.3 Future and past time horizons. Next we investigated whether there were differences in temporal horizons. We excluded observations with no temporal expressions according to SUTime (leaving $n = 35,248$ sentences). We then created a dichotomous "time reference" dummy for whether time horizons were positive for FTR (0) or negative for PTR (1). Temporal horizons were highly right skewed, $skew = 13.44$, and truncated at zero so log transformation was appropriate. We therefore regressed the natural log of the absolute value of time horizon over our predictor variables and their interaction, in case effects of condition differed by time reference:

$$ln(H)_i = \beta_0 + \beta_1 m.heal_{ij} + \beta_2 t.ref_{ij} + \beta_3 m.heal_{ij} t.ref_{ij} + e_{ij} + u_j \quad (4)$$

Inspection of regression residuals indicated the log transformation had resulted in a better approximation of regression assumptions of error normality than untransformed time horizons, so we proceeded. We found that the fixed components of the model explained approximately 8% of the variance, $R^2_m = 0.08,\ R^2_c = 0.16$, and that redditors in the mental health condition had significantly shorter time horizons, $(e^\beta - 1) \times 100 = -24.42,\ SE = 0.08,\ t(2791.38) = -3.56,\ p < .001$ (βs are transformed as indicated so represent percentage change in y). This indicates that time horizons were approximately 24% shorter in the the mental health condition, *Fig. 2a.* There was also a significant effect of time reference. Compared with FTR, time horizons for PTR were approximately 4 times longer, $(e^\beta - 1) \times 100 = 400.28,\ SE = 0.07,$ $t(35247.58) = 24.69,\ p < .001$. The interaction term was not significant, $(e^\beta - 1) \times 100 = -1.98,\ SE = 0.07,\ t(35242.89) = -0.27,\ p = .786$.

2.2.4 FTR type analysis. We next wanted to understand whether there were any differences in how redditors were referring to the future, i.e. in FTR type. We excluded all sentences where $p(FTR) \leq 50\%$ according to the time reference classifier. This left $n = 145,130$. However, because of the key-word methods it employs, the FTR type classifier cannot handle negations. For instance, “rain tomorrow is not possible” expresses high-certainty but would be classed as low-certainty because of the presence of the low-certainty key-word “possible”. We therefore detected the presence of negations using an averaged perceptron tagger following Collins (2002) but with Brown cluster features as described by Koo, Carreras, and Collins (2008) and using greedy decoding (implemented in *spaCy* (Explosion AI, 2020)). We excluded any sentences which used negations ($n = 24,399$), which left a final sample of $n = 120,731$.

We report FTR type usage proportions in *Fig. 3.* These suggest that, rather than the future tense marking the preponderance of FTR statements, FTR marking in

English tends to be fairly equitably distributed across the different semantic categories estimated by the FTR type classifier.

To investigate effect of condition, we regressed each FTR type over mental health condition with a logit link function and random intercepts clustered by redditor, i.e. equation 3 but for FTR type. Results are reported in *Table 2* and *Fig. 4*.

We found that redditors in the mental health condition were less likely to use the future tense, *Fig. 4a*, lexical high-certainty constructions, *Fig. 4e*, the present tense (though only at trend levels, $p < .1$), *Fig. 4b*. At the same time, they were more likely to use low-certainty terminology (whether lexical or verbal, *Fig. 4c-d*), more likely to use deontic modal terms, which encode future obligations and notions of permission and necessity, *Fig. 4f*; and more likely to use of bouletic modal terms, which relate to hopes, plans, and intentions for the future, *Fig. 4g*. We found no differences for irrealis terms, *Fig. 4g*.

2.3 Discussion

The semantic domain encoded by the future tense, the present tense, and lexical high-certainty terms is high-certainty (Giannakidou & Mari, 2018; Robertson & Roberts, 2020; Salkie, 2010). In combination with higher use of low-certainty terms, lower use of these FTR types suggests that the experience of anxiety and/or depression may be characterized in increased uncertainty about future events. Additionally higher use of bouletic terminology suggests increased desire for future change.

We also found they were more likely to make non-present time reference (past and future), but that their time horizons were shortened relative to control. This suggests that anxiety and depression may be characterized by heightened salience of future and past events, but within a shorter temporal window.

However, we found no differences between the anxiety and depression conditions. We reasoned that self-selecting to post in *r/Anxiety* and *r/depression* was a weak diagnostic criteria, especially given high comorbidity rates between the two conditions (Groen et al., 2020). We therefore undertook to differentiate the relationships between anxiety, depression, temporal horizons, and temporal discounting in a survey paradigm over which we could exert a greater degree of control.

3 Study 2

Our goals for study 2 were twofold. Firstly, reasoning that individuals who only tend to imagine themselves in proximal future event might be doing so because they perceived such proximal events as subjectively distal, we hypothesized future time horizons might be negatively correlated with the subjective representation of future distance. In other words, if sufferers of anxiety and depression tend to think of more temporally proximal objective dates, this might be because they represent such dates as more subjectively distal. Since distal representations of subjective temporal distance have been found to be a consistent predictor of increased temporal discounting (Thorstad et al., 2015; Kim & Zauberman, 2009; Zauberman et al., 2009), we hypothesized and that such perceptual biases might actually be driving previously reported tendencies for sufferers of anxiety to temporally discount to a greater degree. Secondly, we wanted to disambiguate whether such cognitive biases were characteristic of anxiety, depression, or both.

3.1 Materials and methods

3.1.1 Participants. Participants who failed attention checks during study 2 were ejected immediately (see SM). After these ejections, a sample of $N = 202$ participants ($n = 103$ females, $n = 98$ males, and $n = 1$ other) completed study 2. Data were collected in March, 2020. All participants were English speakers residing in the United Kingdom and were 18 or over. Participants were recruited from Amazon Mechanical Turk Prime and completed the survey online. Ethical approval for the study was granted by the Brunel University Research Ethics Committee, ref. no. 22348-MHR-Jan/2020-24407-1. All participants were remunerated.

3.1.2 Materials and procedures. The survey was hosted on Qualtrics. It had a within-subjects correlational design, and consisted of 3 tasks. Firstly, participants completed an intertemporal choice task designed to establish temporal discounting, then they completed the "slider" task designed to establish subjective representations of future time (Kim & Zauberman, 2009; Zauberman et al., 2009), and finally the Depression Anxiety and Stress Scales 21 item measure (DASS-21) to establish depression, and anxiety levels (Lovibond & Lovibond, 1995). For all 3 tasks, item order was randomized and one item was displayed per page.

In the intertemporal choice task, participants made repeated binary decisions between a larger reward (the "larger-later reward") offered after a delay and a smaller reward (the "smaller-sooner reward") offered immediately; for instance, "Would you prefer $50 now of $100 in six months?" The smaller-sooner reward was always smaller than the larger-later reward, which was fixed at $100. The factors of this task were the amounts of the smaller-sooner reward ($50-$95 by increments of $5), and the delays of the larger-later reward (later today, tomorrow, one week, one month, two months, three months, six months, one year, two years, five years, and ten years). If participants chose the larger-later reward, this was scored with (1), otherwise (0). Amounts and delays were fully crossed to produce a test battery of $10_{amounts} \times 11_{delays} = 110_{items}$. Prior to starting, participants were told to "Try to answer quickly and intuitively, without thinking about it too much."

To be able to model linear relationships between discounting (which is non-linear over time (Green & Myerson, 2004)) and participant-level variables (anxiety, depression), we needed to calculate a participant-level measure of discounting. Research has shown that the following hyperbolic function fits real and hypothetical delay discounted value very well (Mazur, 1987; Kirby, Petry, & Bickel, 1999):

$$V = \frac{A}{1 + kD} \tag{5}$$

where V is the subjective value of a delayed reward A at a given delay D, and k is a scaling parameter which captures individual differences in discounting. Higher values of k imply lower V, i.e. more discounting. To derive participant-level ks, we followed Kirby et al. (1999). This involved calculating hypothetical values of k and retaining for each participant some k which best predicted empirical choices. Specifically, we calculated all k_{1-n} at indifference between LLR and SSR for each D and SSR, i.e. all ks such that $SSR = LLR/(1 + k_i D)$ for all values of SSR and D under the study. We then predicted hypothetical intertemporal choices for each k_{1-n} and retained k_i for each participant p_j which had the highest proportion of matches against p_j empirical choices. When more than one k had an equal number of matches, we took the geometric mean (Kirby et al., 1999). Since k tends to be non-normally distributed, we used $log_e(k)$ as our participant-level measure of discounting (Kirby et al., 1999).

Participants then completed the slider task. Rather than establishing how far into the future participants tend to think, it was designed to establish how subjectively distant participants felt given objective time distances to be. Following Zauberman et al. (2009) and Kim and Zauberman (2009), participants used a slider to rate whether they construed a given objective temporal distance as "close to now" (0) or "far from now" (100). The only factor of this task was the objective distances, which matched those in the intertemporal choice task. For each item, participants were directed to "indicate with the slider how far away from NOW the given time feels to you." Prior to starting the task, participants were given a training example involving a past time reference (9 months ago), and were told to "try to answer quickly and intuitively, without thinking about it too much." To increase a sense of subjectivity, numbered slider increments were not displayed.

Finally, participants completed the DASS-21 (Lovibond & Lovibond, 1995). The DASS-21 is a self-report questionnaire in which participants rated to what extent a given statement applied to them between "did not apply to me at all" (0) and "applied to me very much or most of the time" (3). For each item, participants were instructed to "Please read each statement use the scale provided to indicate how much the statement applied to you over THE PAST WEEK. There are no right or wrong answers. Do not spend too much time on any statement." The DASS-21 is broken down into 3 dimensions, each comprising 7 items, designed to separately measure depression, anxiety, and stress. Example items from each dimension are as follows: "I felt I wasn't worth much as a person" (depression), "I felt I was close to panic" (anxiety), and "I found it difficult to relax" (stress). See SM for the full questionnaire. It is not a diagnostic tool *visa-vi* the discrete diagnostic categories of, for instance, the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013). Rather, it is based on research which suggests that the differences in depression, anxiety, and stress experienced by healthy and clinical populations are a matter of degree (Lovibond & Lovibond, 1995). It is therefore designed to span the boundaries between clinical and healthy populations. Dependent variables are the the sum of scores for each dimension multiplied by 2, and therefore theoretically range between (0) and (42) (in our data the max score was 40).

3.2 Results

Since we wanted to understand the conditional relationship of anxiety and depression on discounting via subjective temporal distance, we conducted a mediation analysis. In mediation analyses, a predictor X is assumed to effect an outcome Y via a mediating variable M. In mediation terminology, the $X \rightarrow M$ path is refereed to as α, the $M \rightarrow Y$ path is referred to as β, and the $X \rightarrow Y$ path is referred to as τ'. The "indirect effect", $X \rightarrow M \rightarrow Y$, can be estimated as the product of the paths involved, $\alpha\beta$, and captures the effect of X on Y via M while controlling for the "direct effect", τ'. Similarly, the direct effect captures the effect of X on Y while controlling for the indirect effect. The "total effect" is the sum of $\alpha\beta$ and τ', and captures the total effect of X on Y, as in a normal regression (Yuan & MacKinnon, 2009).

We used $log_e(k)$ as our outcome variable, anxiety and depression scores from the DASS-21 as our predictor variables, and participant-level mean subjective temporal distance scores as our mediating variable. The model therefore took the following form:

$$subj.dist_i = \lambda_1 + \alpha_1 anx_i + \alpha_2 dep_i + \gamma_{k1} X_{ki} + e_{1i} \tag{6}$$

$$log_e(k)_i = \lambda_2 + \tau_1' anx_i + \tau_2' dep_i + \beta subj.dist_i + \gamma_{j2} X_{ji} + e_{2i} \tag{7}$$

where $\lambda_{1,2}$ are intercepts, $\gamma_{j1,2}$ are parameters for a vector X_j of covariates (demographic measures, see SM), and $\alpha_{1,2}$, $\tau'_{1,2}$, and β are slope coefficients as described.

We used a Bayesian approach to estimation. Bayesian statistics are well-suited to mediation analyses; as well as making no assumptions about the normality of sampling statistics, they allow for straight-forward inferences about any transformation of model parameters (i.e. path products) through simply carrying out the desired operation on posterior probability distributions (Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2019). We therefore used the *brms* package (Bürkner, 2017) to estimate model parameters using a no U-Turn Hamiltonian Monte Carlo sampling procedure (Hoffman & Gelman, 2014; Stan Development Team, 2020) with uninformative priors (i.e. $Unif(-\infty\ -\ \infty)$), 4 chains, $16,000$ iterations, a burn in discard of $4,000$, and a max tree-depth of 15. Inspection of caterpillar plots and $\hat{R}$s indicated that the estimation procedure had converged (Vehtari et al., 2019).

We report a conceptual diagram including point estimates and 95% Credibility Intervals (CIs), *Fig. 5b*, and path products of interest, *Fig. 5a.* Since we had made the directional hypotheses that anxiety and/or depression would be associated with more distal subjective temporal distances and therefore increased discounting (higher $log_e(k)$), we tested the one-tailed hypotheses that these effects would be positive. As we had predicted for anxiety, we found significant positive indirect, $Est. = 0.04$, $Er. = 0.02,\ 90\%CI = [0.02, 0.07],\ pp > .999$, direct, $Est. = 0.07,\ Er. = 0.03$, $90\%CI = [0.02, 0.13],\ pp = .982$, and total, $Est. = 0.11,\ Er. = 0.03$, $90\%CI = [0.06, 0.17],\ pp > .999$, effects. However, we found no significant effects for depression (indirect: $Est. = -0.01,\ Er. = 0.01,\ 90\%CI = [-0.02, 0],\ pp = .158$, direct: $Est. = -0.02,\ Er. = 0.02,\ 90\%CI = [-0.06, 0.01],\ pp = .146$, total: $Est. = -0.03$, $Er. = 0.02,\ 90\%CI = [-0.07, 0.01],\ pp = .103$). If anything, effects tend to go in the opposite direction, though not significantly. In these parameter summaries, $Est.$ and $Er.$ are the mean and standard deviation of the posterior probability distribution, and"*pp*" reports the probability that the posterior estimate of the effect matches the direction of the hypothesis. In contrast to frequentist p values, $pp > .95$ indicates that evidence that the parameter matched the prediction exceeds 95% (Bürkner, 2017).

3.2.1 Discussion.

These findings suggest that anxious participants tended to construe future events more distally, and that this in turn caused them to temporally discount to a greater degree. The significant direct effect also suggests anxiety may be associated with increased temporal discounting independently of temporal construals – or at least our measurements of them. On the other hand, we found no significant relationships between depression and subjective temporal distance or temporal discounting.

4 General discussion

In study 1, we found that language used by redditors who had posted to *r/Anxiety* and *r/depression* differed significantly from control. Firstly, we found they were more likely to make non-present time reference, both past and future. This suggests that sufferers of anxiety and/or depression may spend more time thinking about future and past events rather than present ones. It may be important to note that we did not build a tri-class past/present/future time reference classifier, so the reference category in these analyses is not present time reference. Rather it is anything that is not future/past time reference, for instance statements of ability "it can rain a lot here"; statements of fact, "the sun is hot"; or present time reference, "I am hot". In light of this, it would be incorrect to infer that redditors in the mental health condition spent less time making present time reference. Rather, they spent more time making past and future time reference.

While relationships between future time perspective and anxiety and depression have been established, such research does not illuminate the extent to which the future is salient to sufferers from these mood disorders. As far as we know, ours is the first study to find that the experience of anxiety and depression may be characterized by an increased amount of time spent thinking about the future and the past.

We also found significant (though small) differences in FTR type. This suggests that habits of thinking about the future may differ between healthy populations and sufferers of anxiety and/or depression. The largest differences seemed to be driven by lower use of the future tense and higher use of bouletic language relative to control, *Fig. 3*. The semantic domain which bouletic modal language encodes in desiderative, having to do with hopes, desires, and plans for change. This suggest that future thinking in sufferers of anxiety and/or depression tends to be characterized by desire for future change rather than simply statements of fact.

The other major difference was that we found higher use of low-certainty language relative to control; this suggests that sufferers of anxiety and/or depression may construe the future more uncertainly then healthy populations. This raises the possibility that probability discounting may be a relevant marker of anxiety and/or depression. Probability discounting is similar to delay discounting, but involves the devaluation of uncertain rather than delayed rewards (Green & Myerson, 2004). It is measured in a similar way, by giving participants binary choices, but rather than immediate and delayed, the choice is between certain and uncertain rewards; for instance a 100% chance of receiving $10 vs. a 50% chance of receiving $20. People who probabilistically discount more would tend to choose high-certainty options because they devalue low-certainty ones to a greater degree. Recent research involving rewards which are both probabilistic and delayed, e.g. a 100% chance of receiving $10 vs. a 50% chance of receiving $20 in three months, has found that these factors interactively predict discounting rates (Vanderveldt, Green, & Myerson, 2015). This suggests that real-world intertemporal decision making (in which people must balance present vs. future options as in delay discounting paradigms) which usually involves some degree of uncertainty about the future may be driven by both probabilistic and temporal discounting rates.

The relationship between probability discounting and anxiety and/or depression is understudied relative to temporal discounting. While one study found probability discounting was not a significant marker of major depressive disorder (Hart, Brown, Roffman, & Perlis, 2019), our results suggest that anxiety and/or depression may be characterized by low-certainty construals of future events, which may warrant further investigation into potential relationships between anxiety, depression, and low-certainty future construals which may impact probabilistic intertemporal decision making processes.

Additionally, we found higher use of deontic modal language relative to control. The primary deontic modal verb in English is *must*, and the other criterion key-words in the FTR type classifier in this category are *have to*, *need to*, and variants thereof. This suggests that redditors in the mental health condition were more likely to encode *obligation* and *necessity*, and that therefore their experience of future events may feel simultaneously lower-certainty and more constrained by future obligations.

It is unclear whether these patterns are causal or descriptive. In other words, habitually construing future events in low-certainty, highly-constrained, bouletic terms may actually contribute to the development and maintenance of anxiety and/or depression. On the other hand, such FTR usage differences may simply be symptomatic of anxiety and/or depression. In either case, we appear to have isolated isolated new markers of these mood disorders vis-à-vis FTR. A functionalist perspective may help clarify this particular pattern of results.... SHANE?

Finally, in both study 1 and 2, we found differences in terms of time horizons. In study 1, we found contracted time horizons relative to control. In other words, redditors in the mental health condition referenced more proximal future and past times. In study 2, we operationalized time horizons in a slightly different way; rather then using the Wallace task to establish how far into the future people tend to imagine (i.e. time horizon) we used a slider paradigm to establish far into the future participants felt to be a given objective distance. We found that anxious participants tended to rate future events as farther away. If the study populations are comparable, this suggests that time horizons and subjective future distance may be negatively related, as we had hypothesized. In other words, shortened objective time horizons may arise because objective future dates are represented as subjectively more distant. An intriguing possibility is that over subjective time future time horizons are constant. More research using slider type tasks as well as the Wallace task might explore whether this is the case.

We also found that distal future distance representations in anxious individuals mediated a relationship between anxiety and temporal discounting. This suggests that subjective representations of future time may be partially driving reported links between anxiety and temporal discounting (Rounds et al., 2007), and adds evidence to the inference that the relationship is significant and positive (contra: Jenks & Lawyer, 2015; Steinglass et al., 2017).

On the other hand, we found no relationship between our measures of discounting and depression, despite the fact that depression has been previously linked to shorted future time horizons (Dilling & Rabin, 1967), and increased temporal discounting (Felton et al., 2020; Pulcu et al., 2014). This may be driven by symptom severity and/or reward size effects. Reward size has a well-established effect on discounting rates; larger future rewards are discounted less, i.e. people are more willing to wait when future rewards are large even when present rewards are proportionally identical (Green & Myerson, 2004). Different reward sizes may therefore be responsible for differences in reported findings.

Indeed, reward size may affect discounting in depressive individuals differently to health ones. Pulcu et al. (2014) compared discounting between clinically diagnosed present sufferers of Major Depressive Disorder (MDD), remitted MMD (rMDD), and healthy controls, with small (£30), medium (£55) and large (£80) rewards. They found no significant differences between MDD participants and controls, and only found significant differences between MDD and controls for large rewards. Since we used a comparable reward size ($100), this suggests that depressive symptom severity in our sample was not high enough to detect an effect. On the other hand, the fact we found no relationship between subjective temporal representations and depression despite previously reported links between depression and shortened time horizons (Dilling & Rabin, 1967) suggests more work might be done using slider-type subjective distance task and the Wallace task in sufferers of depression.

4.1 Construal level theory

How are we to make sense of the overall pattern of results across both studies? Construal Level Theory (CLT) offers one theoretical framework which may help explain the seemingly disparate findings. In this connection, CLT posits that there exists a relationship between psychological distance and abstraction. Specifically, the claim is that events that are perceived as distant on the dimensions of space, time, likelihood, and social familiarity are perceived as more abstract, with nearer events being perceived as more concrete. Reciprocally, it is hypothesized that abstraction cues expectations of psychological distance, and concreteness expectations of psychological proximity (Trope & Liberman, 2010). Thus, far-future or culturally alien scenarios are imagined as more abstract than temporally or culturally proximal scenarios. The value of CLT for our purposes is that its claims provide an empirically supported link between temporal distance and areas of cognition like moral orientation, self-perception, emotionality, tolerance, conformity, and creativity.

The first finding the CLT might explain is the discovery that redditors in *r/Anxiety* and *r/depression* are more likely to make non-present time reference.

4.2 Therapeutic applications

How might our results be leveraged to better inform therapeutic approaches to the treatment of anxiety and depression.... SHANE?

4.3 Conclusions

Anxiety and depression continue to affect the lives of many (Dewa & McDaid, 2011). In combining descriptive big-data analyses with survey-based paradigms we used the information available on social media to develop novel insights into the habits of speech and thought which may characterize anxiety and depression. We were able to show that redditors who has once posted to forums dedicated to discussion of anxiety and depression made more frequent past and future time reference, and used a different distribution of FTR type; they used more bouletic, more low-certainty, more deontic, and fewer future tense constructions. While these effects were small, they might be further developed as markers of these disorders. Additionally, we found once posters to mental health forums exhibited contracted time horizons relative to control. This motivated a mediation analysis in study 2 in which we found that subjective representations of future temporal distance mediated a relationship between anxiety and temporal discounting; anxious participants construed future events as more distant, and therefore were less likely to wait for future rewards.

If progress is to be made in understanding and treating these disorders, new markers need to be explored which may help develop more accurate diagnostic criteria. Additionally, predictive approaches to online social data need to be complemented by descriptive ones which allow insights from the analysis of social media data to be tested in controlled environments, and applied in novel therapies. Our results suggest that such approaches may yield fruitful insights. In particular they suggest a multimodal approach to future time perceptive may be useful and that ongoing research should attempt to disentangle how these dimensions may contribute to the etiology of anxiety and depression.

5 Tables

Table 1
Time reference and FTR type classifier performance metrics

classifier	category	sub-category	accuracy	precision	recall	f1
time reference	FTR	-	.91	.93	.73	.82
	PTR	-	.81	.83	.35	.49
FTR-type	future tense	-	.84	.81	.92	.86
	present tense	-	.93	.94	.99	.96
	low-certainty	verbal	.99	$>$ .99	.99	.99
		lexical	.96	$>$ .99	.97	.98
	high-certainty	verbal	.98	$>$ .99	.98	.99
		lexical	.98	.99	.98	.99
	bouletic modality	-	.94	.99	.94	.97
	irrealis	-	.89	.99	.89	.94

Accuracy is defined as $a = (tp + tn)/(tp + fp + fn + tn)$ where tp is the number of true positives, tn is the number of true negatives, fp is the number of false positives and fn is the number of false negatives. Accuracy captures the classifier's performance without prioritizing either positive or negative examples. Precision is $p = tp/(tp+fp)$; it captures the model's likelihood of being correct if it makes a positive prediction and is therefore sensitive to the model's type I error rate. Conversely, denominator in recall is the false negatives, $r = tp/(tp + fn)$; it therefore captures whether the model tends to miss true examples, and is sensitive to the model's type II error rate. F1 is the harmonic mean of recall and precision, $F1 = (2rp)/(r + p)$, and attempts to balance the two.

Table 2
FTR type by condition on Reddit (study 1)

DV	IV	e^{β}	SE	*z-score*	
future tense	(Intercept)	0.37	0.02	-41.96	***
$R^2_m < .001,\ R^2_c = .023$	mental health cond.	0.92	0.03	-3.11	**
present tense	(Intercept)	0.19	0.02	-67.42	***
$R^2_m < .001,\ R^2_c = .016$	mental health cond.	0.95	0.03	-1.89	$\cdot$
lexical low-certainty	(Intercept)	0.11	0.03	-70.7	***
$R^2_m = .001,\ R^2_c = .028$	mental health cond.	1.22	0.04	5.56	***
verbal low-certainty	(Intercept)	0.09	0.03	-74.01	***
$R^2_m < .001,\ R^2_c = .025$	mental health cond.	1.14	0.04	3.71	***
lexical certainty	(Intercept)	0.04	0.05	-69.4	***
$R^2_m < .001,\ R^2_c = .042$	mental health cond.	0.9	0.05	-2.07	*
deontic	(Intercept)	0.03	0.04	-80.02	***
$R^2_m < .001,\ R^2_c = .021$	mental health cond.	1.14	0.05	2.71	**
bouletic modality	(Intercept)	0.14	0.03	-63.61	***
$R^2_m = .007,\ R^2_c = .041$	mental health cond.	1.57	0.04	12.72	***
irrealis	(Intercept)	0.15	0.03	-65.91	***
$R^2_m < .001,\ R^2_c = .025$	mental health cond.	1.03	0.03	0.88	

Coefficients represent the change in odds ratio of each predictor compared to the intercept.
$^{***}\ p < .001$; $^{**}\ p < .01$; $^{*}\ p < .05$; $\cdot\ p < .1$

6 Figures

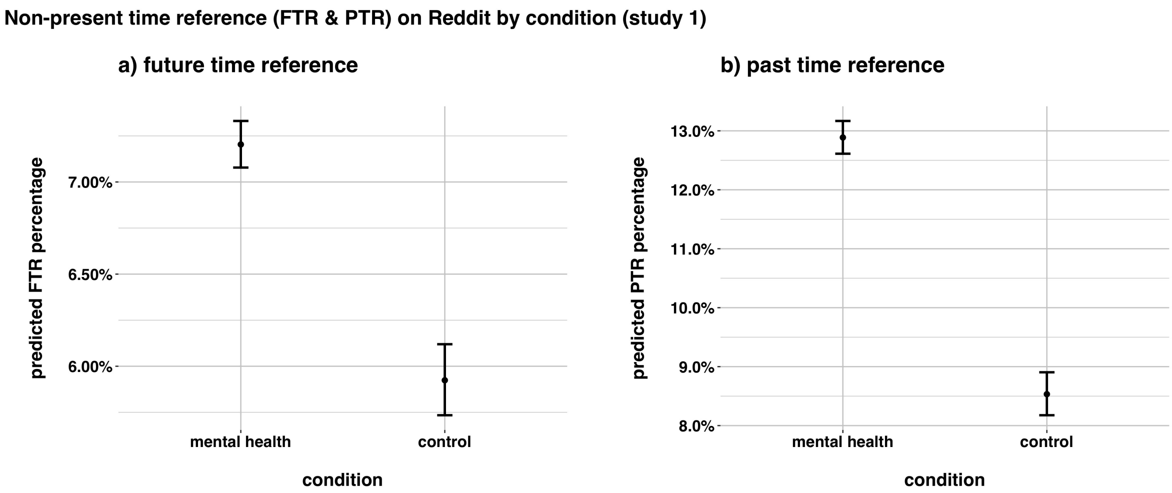

Figure 1. Differences in probability of making FTR (a) and PTR (b) as a function of condition. Redditors in the mental health condition exhibited higher probability of referencing the past and future relative to control. Above and in *Figs. 2,4*, confidence intervals account for random effect variance by matrix-multiplying a predictor X by the parameter vector B to get the predictions, then extracting the variance-covariance matrix V of the parameters and computing XVX' to get the variance-covariance matrix of the predictions. The square-root of the diagonal of this matrix represents the standard errors of the predictions, which are then multiplied by ±1.96 (Lüdecke, 2019).

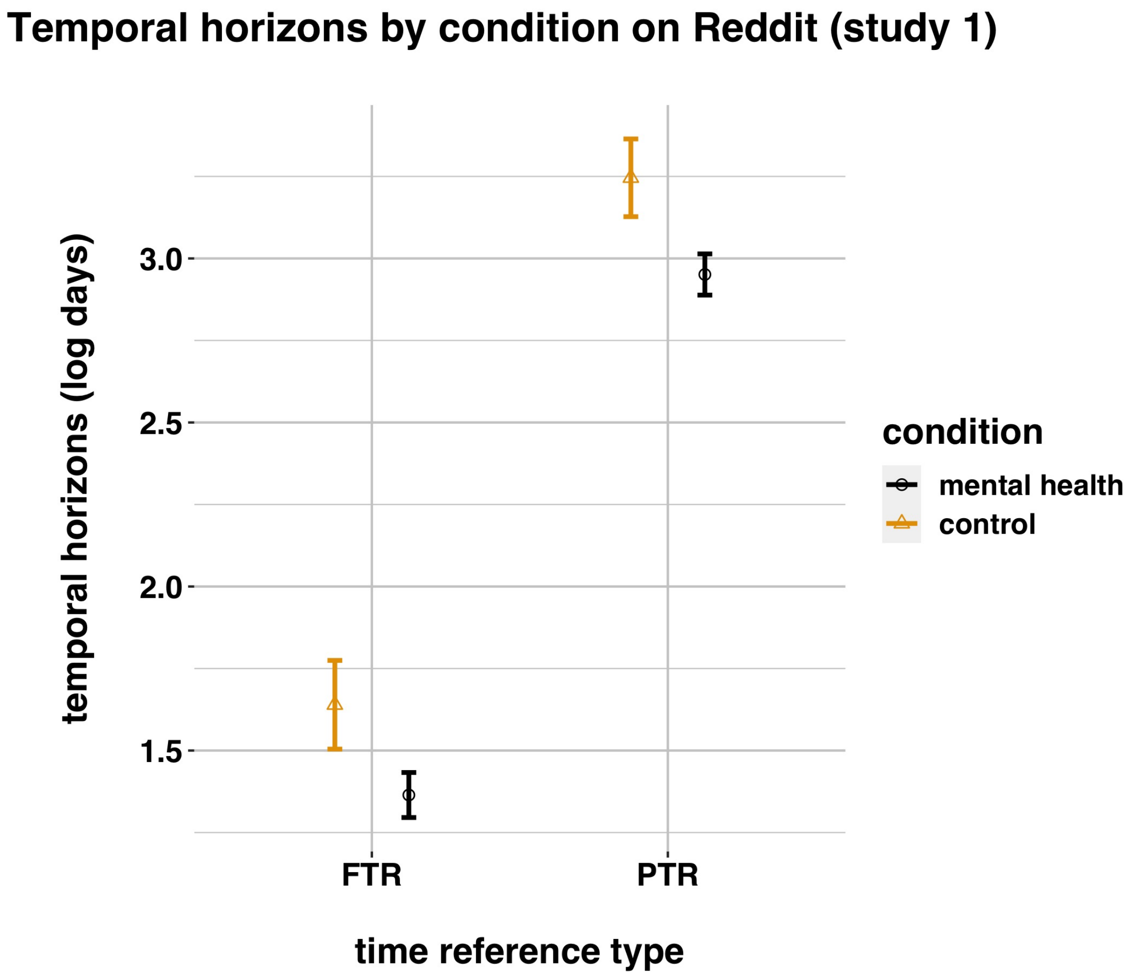

Figure 2. Differences in temporal horizons as a function of time reference type and condition. Redditors in the mental health condition exhibited contracted time horizons relative to control.

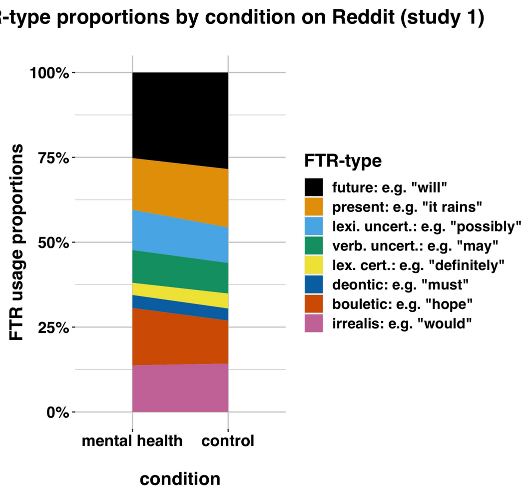

Figure 3. Plotted values are the proportional breakdown of FTR type, i.e. the proportions per type normalized to sum to 1, $x_k = ((\sum_{i=1}^{n_k} x_{ik})/n_k)/\sum_{i=1}^{k} x_k$ for k FTR-types. This suggests FTR is spread across multiple semantic domains, not just tense (results are similar even when modality does not dominate tense (see SM).

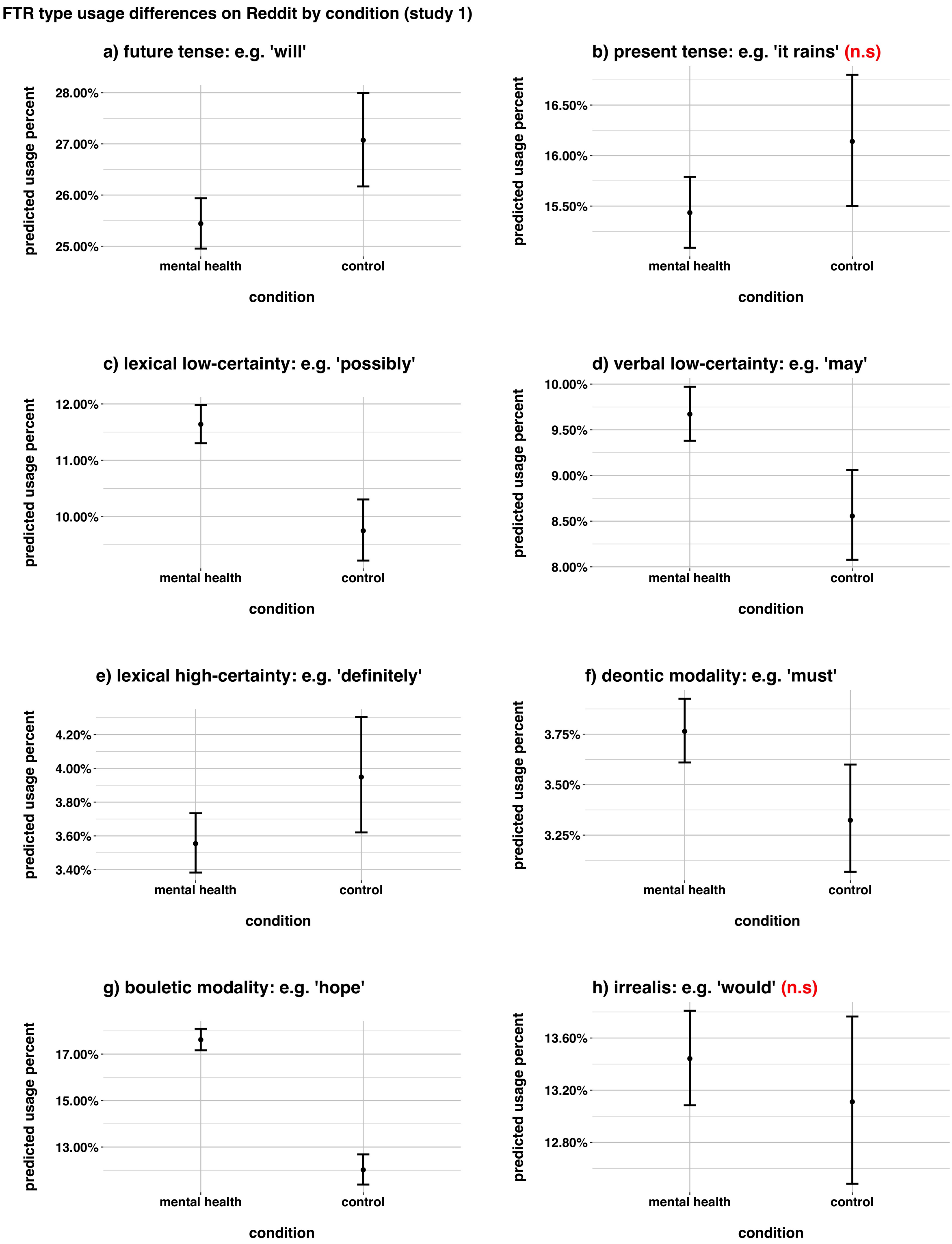

Figure 4. Differences in probability of using different FTR types when referring to the future as a function of condition. Redditors in the mental health condition exhibited higher probability of using low-certainty language, bouletic language, and verbal deontic language, and lower probability of using the future tense and lexical high-certainty language.

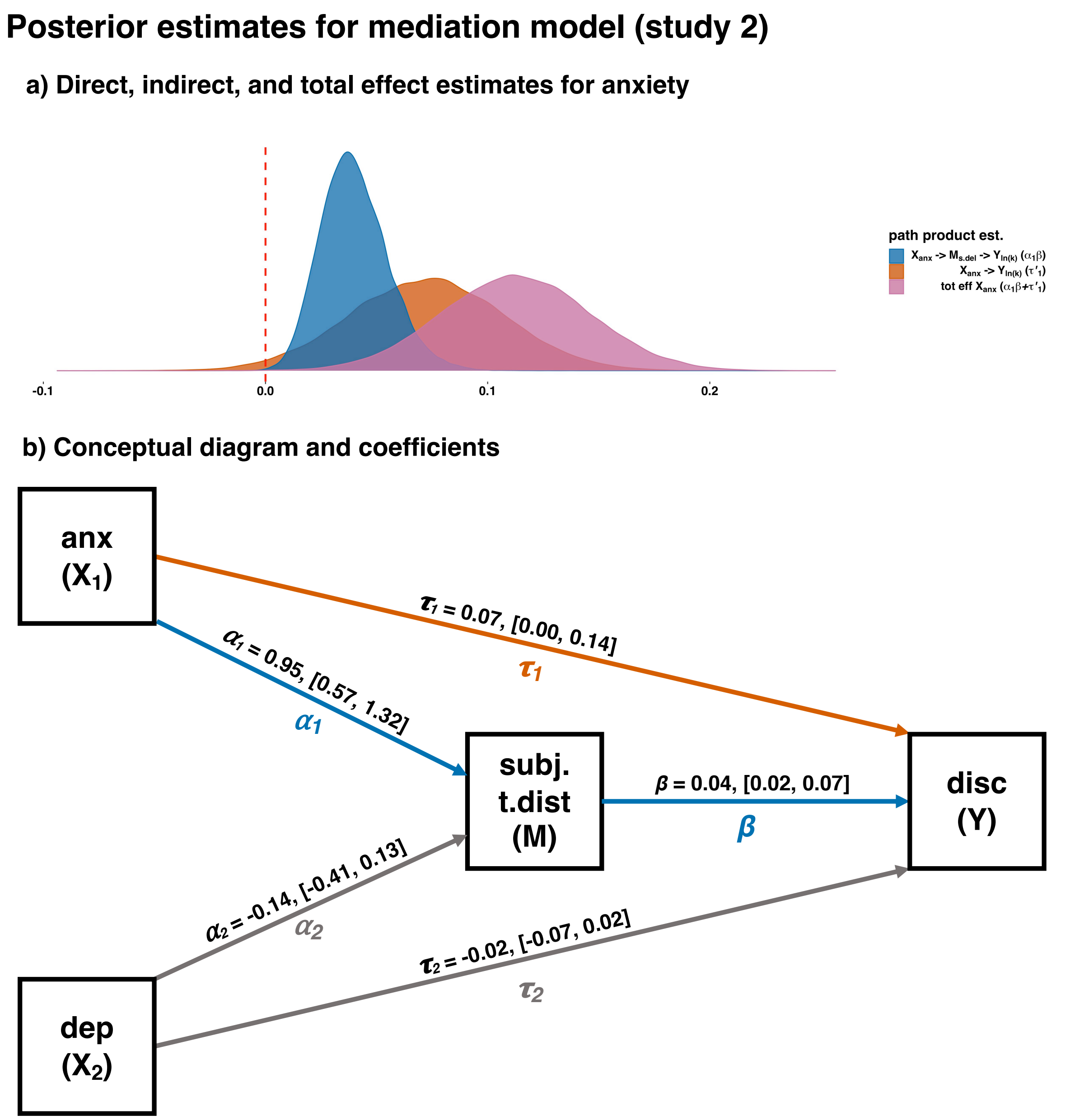

Figure 5. a) posterior estimates for paths and path products for the effects of anxiety on discounting $ln(k)$ via mean ratings of subjective future temporal distance. No effects of depression were significant so for easier viewing are not depicted.
b) Conceptual diagram, with parameter point estimates and Bayesian 95% Credibility Intervals (CI). If the CI does not contain zero, parameters are considered to be significantly different from zero.

Acknowledgments

This research was supported TKTK. Additionally, CR is supported by a Natural Sciences and Engineering Council of Canada doctoral fellowship (PGSD2 - 517110 - 2018) and by EU's Horizon 2020 FET Open RIA 662725 IBSEN project. We thank Arran Davis and Bronwyn Tarr for editing and advice on manuscript structure.

7 References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5.* Washington: American Psychiatric Publishing.

American Psychiatric Association. (2015). *Anxiety Disorders: DSM-5 Selections.* Philadelphia: American Psychiatric Publishing.

Barrault, S., & Varescon, I. (2013). Cognitive distortions, anxiety, and depression among regular and pathological gambling online poker players. *Cyberpsychology, Behavior, and Social Networking*, *16*(3), 183–188. doi: 10.1089/cyber.2012.0150

Blaszczynski, A., & McConaghy, N. (1989). Anxiety and/or depression in the pathogenesis of addictive gambling. *Substance Use and Misuse*, *24*(4), 337–350. doi: 10.3109/10826088909047292

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *10*(1), 1–28. doi: 10.18637/jss.v080.i01

Chang, A. X., & Manning, C. D. (2012). SUTIME: A library for recognizing and normalizing time expressions. In *Proceedings of the 8th international conference on language resources and evaluation, lrec 2012* (pp. 3735–3740).

Chen, K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, *103*(2), 690–731. doi: 10.1257/aer.103.2.690

Chi, J. D., Su, X., Tang, Y., & Xu, B. (2018). Is language an economic institution? Evidence from R&D investment. *SSRN Electronic Journal*, 1–48. doi: 10.2139/ssrn.3262822

Collins, M. (2002). Discriminative training methods for hidden markov models: theory and experiments with perceptron algorithms. In *Proceedings of the conference on empirical methods in natural language processing (emnlp)* (pp. 1–8). Philadelphia: Association for Computational Linguistics.

Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., & Mitchell, M. (2015). CLPsych 2015 Shared Task: Depression and PTSD on Twitter. In *Proceedings ofthe 2nd workshop on computational linguistics and clinical psychology: From* *linguistic signal to clinical reality* (pp. 31–39). Denver: Association for Computational Linguistics. doi: 10.3115/v1/w15-1204

Dahl, Ö. (1985). *Tense and Aspect Systems.* Oxford: Blackwell.

Dahl, Ö. (2000). The grammar of future time reference in European languages. In Ö. Dahl (Ed.), *Tense and aspect in the languages of europe* (pp. 309–328). Berlin/New York: Mouton de Gruyter.

De Choudhury, M., Counts, S., Horvitz, E. J., & Hoff, A. (2014). Characterizing and predicting postpartum depression from shared facebook data. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 625–637. doi: 10.1145/2531602.2531675

Dewa, C. S., & McDaid, D. (2011). Investing in the Mental Health of the Labor Force: Epidemiological and Economic Impact of Mental Health Disabilities in the Workplace. In I. Z. Schultz & E. Sally Rogers (Eds.), *Work accommodation and retention in mental health* (pp. 33–51). Berlin: Springer Science+Business Media. doi: 10.1007/978-1-4419-0428-7

Dilling, C. A., & Rabin, A. I. (1967). Temporal experiences in depressive states and schizophrenia. *Journal of Counsulting Psychology*, *31*(6), 604–608.

Explosion AI. (2020). *spaCy.* Berlin: Explosion AI. Retrieved from `https://spacy.io/`

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, *133*(4), 1645–1692. doi: 10.1093/qje/qjy013

Fasan, M., Gotti, G., Kang, T., & Liu, Y. (2016). Language FTR and earnings management: International evidence. *SSRN Electronic Journal*, 1–49. doi: 10.2139/ssrn.2763922

Felton, J. W., Strutz, K. L., McCauley, H. L., Poland, C. A., Barnhart, K. J., & Lejuez, C. W. (2020). Delay Discounting Interacts with Distress Tolerance to Predict Depression and Alcohol Use Disorders among Individuals Receiving Inpatient Substance Use Services. *International Journal of Mental Health and Addiction*, *18*(5), 1416–1421. doi: 10.1007/s11469-019-00163-5

Figlio, D., Giuliano, P., Özek, U., & Sapienza, P. (2016). Long-term orientation and educational performance. *National beureau of economic research working paper series*(10147). doi: 10.3386/w22541

Giannakidou, A., & Mari, A. (2018). A unified analysis of the future as epistemic modality: The view from Greek and Italian. *Natural language and linguistic Theory*, *36*(1), 85–129.

Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T. J., Dobson, R. J., & Dutta, R. (2017). Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific Reports*, *7*(March). doi: 10.1038/srep45141

Goldstone, R. L., & Lupyan, G. (2016). Discovering Psychological Principles by Mining Naturally Occurring Data Sets. *Topics in Cognitive Science*, *8*(3), 548–568. doi: 10.1111/tops.12212

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin*, *130*(5), 769–792. doi: 10.1037/0033-2909.130.5.769

Groen, R. N., Ryan, O., Wigman, J. T., Riese, H., Penninx, B. W., Giltay, E. J., ... Hartman, C. A. (2020). Comorbidity between depression and anxiety: Assessing the role of bridge mental states in dynamic psychological networks. *BMC Medicine*, *18*(1), 1–17. doi: 10.1186/s12916-020-01738-z

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, *18*, 43–49. doi: 10.1016/j.cobeha.2017.07.005

Hart, K. L., Brown, H. E., Roffman, J. L., & Perlis, R. H. (2019). Risk tolerance measured by probability discounting among individuals with primary mood and psychotic disorders. *Neuropsychology*, *33*(3), 417–424. doi: 10.1037/neu0000506

Hodgins, D. C., & Engel, A. (2002). Future time perspective in pathological gamblers.

Journal of Nervous and Mental Disease, *190*(11), 775–780. doi: 10.1097/00005053-200211000-00008

Hoffman, M. D., & Gelman, A. (2014). The No-U-turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(2008), 1593–1623.

Honnibal, M., & Johnson, M. (2015). An Improved Non-monotonic Transition System for Dependency Parsing. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1373—-1378). Lisbon, Portugal: Association for Computational Linguistics. Retrieved from https://aclweb.org/anthology/D/D15/D15-1162

https://www.redditinc.com/advertising/audience. (n.d.). Retrieved 2020-12-22, from https://www.redditinc.com/advertising/audience

Hübner, M., & Vannoorenberghe, G. (2015a). Patience and Inflation. *Munich Personal RePEc Archive*(65811). doi: 10.1227/01.NEU.0000349921.14519.2A

Hübner, M., & Vannoorenberghe, G. (2015b). Patience and long-run growth. *Economics Letters*, *137*, 163–167. doi: 10.1016/j.econlet.2015.10.011

Jenks, C. W., & Lawyer, S. R. (2015). Using delay discounting to understand impulsive choice in socially anxious individuals: Failure to replicate. *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 198–201. doi: 10.1016/j.jbtep.2014.10.010

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it. *Molecular Psychiatry*, *17*(12), 1174–1179. doi: 10.1038/mp.2012.105

Kim, B. K., & Zauberman, G. (2009). Perception of Anticipatory Time in Temporal Discounting. *Journal of Neuroscience, Psychology, and Economics*, *2*(2), 91–101. doi: 10.1037/a0017686

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General*, *128*(1), 78–87. doi: 10.1037/0096-3445.128.1.78

Kish, L. (1965). *Survey sampling.* New York: John Wiley & Sons Ltd.

Koo, T., Carreras, X., & Collins, M. (2008). Simple Semi-supervised Dependency Parsing. In *Proceedings ofacl-08: Hlt* (pp. 595–603). Columbus, Ohio: Association for Computational Linguistics.

Lai, M. H., & Kwok, O. M. (2015). Examining the rule of thumb of not using multilevel modeling: The "design effect smaller than two" rule. *Journal of Experimental Education*, *83*(3), 423–438. doi: 10.1080/00220973.2014.907229

Liang, H., Marquis, C., Renneboog, L., & Sun, S. L. (2018). Future-time framing: The effect of language on corporate future orientation. *Organization Science*, *29*(6), 1093–1111. doi: 10.1287/orsc.2018.1217

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety and Stress Scales* (2nd ed.). Sydney: Psychology Foundation.

Lüdecke, D. (2019). *Ggeffects: Marginal effects of regression models.* Retrieved from `https://strengejacke.github.io/ggeffects/articles/ggeffects.html`

Mavisakalyan, A., Tarverdi, Y., & Weber, C. (2018). Talking in the present, caring for the future: Language and environment. *Journal of Comparative Economics*, *46*(4), 1370–1387. doi: 10.1016/j.jce.2018.01.003

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *The effect of delay and of intervening events on reinforcement value: Quantitative analyses of behavior, vol. 5* (pp. 55–73). Hillsdale, NJ: Erlbaum.

Montani, I., & Honnibal, M. (2020). *Prodigy: A new annotation tool for radically efficient machine learning.* Berlin: Explosion AI. Retrieved from `https://prod.gy`

Muthén, B. O., & Satorra, A. (1995). Complex Sample Data in Structural Equation Modeling. *Sociological Methodology*, *25*, 267–316. doi: 10.2307/271070

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134).

doi: 10.1098/rsif.2017.0213

Nuyts, J. (2000). *Epistemic modality, language, and conceptualization.* Amsterdam: John Benjamins. doi: 10.1075/hcp.5

Palmer, R. F. (2001). *Mood and modality* (2nd ed.). Cambridge: Cambridge University Press.

Papastamatelou, J., Unger, A., Giotakos, O., & Athanasiadou, F. (2015). Is Time Perspective a Predictor of Anxiety and Perceived Stress? Some Preliminary Results from Greece. *Psychological Studies*, *60*(4), 468–477. doi: 10.1007/s12646-015-0342-6

Pérez, E. O., & Tavits, M. (2017). Language shapes people's time perspective and support for future-oriented policies. *American Journal of Political Science*, *00*(0), 1–13. doi: 10.1111/ajps.12290

Petry, N. M., Bickel, W. K., & Arnett, M. (1998). Shortened time horizons and insensitivity to future consequences in heroin addicts. *Addiction*, *93*(5), 729–738. doi: 10.1046/j.1360-0443.1998.9357298.x

Pulcu, E., Trotter, P. D., Thomas, E. J., McFarquhar, M., Juhasz, G., Sahakian, B. J., ... Elliott, R. (2014). Temporal discounting in major depressive disorder. *Psychological Medicine*, *44*(9), 1825–1834. doi: 10.1017/S0033291713002584

Python Software Foundation. (2017). *Python language reference.* Retrieved from `http://www.python.org`

Rai, A. (2020). Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141. doi: 10.1007/s11747-019-00710-5

reddit.com. (2020).

Robertson, C. (2019). *FTR-classifier.* Retrieved from `https://pypi.org/project/ftr-classifier/`

Robertson, C., & Roberts, S. G. (2020). Not when but if: Modality and future time reference in English, Dutch and German. *PsyArXiv.* doi: 10.31234/osf.io/gdr8e

Rounds, J. S., Beck, J. G., & Grant, D. M. M. (2007). Is the delay discounting paradigm useful in understanding social anxiety? *Behaviour Research and*

Therapy, *45*(4), 729–735. doi: 10.1016/j.brat.2006.06.007

Salkie, R. (2010). Will: Tense or modal or both? *English Language and Linguistics*, *14*(2), 187–215. doi: 10.1017/S1360674310000055

Smart, R. G. (1968). Future Time Perspectives in Alcoholics and Social Drinkers. *Journal of Abnormal Psychology*, *73*(1), 81–83. doi: 10.1037/h0025449

Stan Development Team. (2020). *Stan modeling language users guide and reference manual.* Retrieved from `https://mc-stan.org`

Steinglass, J. E., Lempert, K. M., Choo, T. H., Kimeldorf, M. B., Wall, M., Walsh, B. T., … Simpson, H. B. (2017). Temporal discounting across three psychiatric disorders: Anorexia nervosa, obsessive compulsive disorder, and social anxiety disorder. *Depression and Anxiety*, *34*(5), 463–470. doi: 10.1002/da.22586

Thorstad, R., Nie, A., & Wolff, P. (2015). Representations of Time Affect Willingness to Wait for Future Rewards. In *Proceedings of the annual meeting of the cognitive science society* (pp. 2392–2397).

Thorstad, R., & Wolff, P. (2018). A big data analysis of the relationship between future thinking and decision-making. *Proceedings of the National Academy of Sciences*, *E1741*. doi: 10.1073/pnas.1706589115/

Thorstad, R., & Wolff, P. (2019). Predicting future mental illness from social media: A big-data approach. *Behavior Research Methods*, *51*(4), 1586–1600. doi: 10.3758/s13428-019-01235-z

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. doi: 10.1037/a0018963

Vanderveldt, A., Green, L., & Myerson, J. (2015). Discounting of monetary rewards that are both delayed and probabilistic: Delay and probability combine multiplicatively, not additively. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(1), 148–162. doi: 10.1037/xlm0000029

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2019). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *arXiv*, 1–27. doi: 10.1214/20-ba1221

Wallace, M. (1956). Future time perspective in schizophrenia. *Journal of Abnormal and Social Psychology*, *52*(2), 240–245. doi: 10.1037/h0039899

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*(4), 301–322. doi: 10.1037/a0016972

Zauberman, G., Kim, B. K., Malkoc, S. A., & Bettman, J. R. (2009). Discounting time and time discounting: Subjective time perception and intertemporal preferences. *Journal of Marketing Research*, *46*(4), 543–556. doi: 10.1509/jmkr.46.4.543

Zhu, J., Hu, S., Wang, J., & Zheng, X. (2020). Future orientation promotes climate concern and mitigation. *Journal of Cleaner Production*, *262*, 1–10. doi: 10.1016/j.jclepro.2020.121212