# The Effect of User Psychology on the Content of Social Media Posts

**Citation for published version:**
Chen, LL, Magdy, W & Wolters, M 2020, 'The Effect of User Psychology on the Content of Social Media Posts: Originality and Transitions Matter', *Frontiers in Psychology*, vol. 11, 526. https://doi.org/10.3389/fpsyg.2020.00526

**Digital Object Identifier (DOI):**
10.3389/fpsyg.2020.00526

**Link:**
Link to publication record in Edinburgh Research Explorer

**Document Version:**
Peer reviewed version

**Published In:**
Frontiers in Psychology

# The Effect of User Psychology on the Content of Social Media Posts: Originality and Transitions Matter

**Lucia Lushi Chen** [1,*], **Walid Magdy** [1] **and Maria Wolters** [1]

[1] *School of Informatics, University of Edinburgh, Edinburgh, United Kingdom*

Correspondence*:
Lucia Lushi Chen
lushi.chen@ed.ac.uk

## ABSTRACT

Multiple studies suggest that frequencies of affective words in social media text are associated with the user's personality and mental health. In this study, we re-examine these associations by looking at the transition patterns of affect. We analyzed the content originality and affect polarity of 4086 posts from 70 adult Facebook users contributed over two months. We studied posting behaviour including silence periods when the user does not post any content. Our results show that more extrovert participants tend to post positive content continuously, and that more agreeable participants tend to avoid posting negative content. We also observe that participants with stronger depression symptoms posted more non-original content. We recommend that transitions of affect pattern derived from social media text and content originality should be considered in further studies on mental health, personality, and social media.

**Keywords: affect, social media, emotion, Facebook, personality traits, depression, mental health**

## 1 INTRODUCTION

Many people express rich moods and emotions in their social media posts. Psychologists use the word "affect" to describe these experiences of feelings and emotions. Affect plays an important role in cognition (Gross et al., 1998) and wellbeing (Silvera et al., 2008). Therefore, affective expressions on social media text have emerged as a key variable for making inferences about users' personality traits (Bachrach et al., 2012; Golbeck et al., 2011; Farnadi et al., 2013) or mental health (De Choudhury et al., 2013; Coppersmith et al., 2014; De Choudhury and De, 2014; Bazarova et al., 2015).

Existing studies formulate the associations between affect and wellbeing based on the frequencies of affective words used in social media text (Schwartz et al., 2013; Yarkoni, 2010; Golbeck et al., 2011; Park et al., 2015; Chen et al., 2020). However, patterns of affect are an important class of symptoms of affective disorders (Rottenberg, 2005; Frijda, 1993; Bylsma et al., 2011; Sheppes et al., 2015; Thompson et al., 2012; Houben et al., 2015; Carlo et al., 2012). Personality may also predispose individuals to specific moods (Rusting and Larsen, 1995; Rusting, 1998). With this in mind, we examined how patterns of affect expressed in social media text is related to with users' mental health and personality.

While non-original content has been extensively studied in opinion mining (Agarwal et al., 2011; Balahur et al., 2009), it has been comparatively neglected in the study of psychological interpretations of social media data. However, social media users often use lyrics or quotes to communicate their emotions. Such content comes from other media, such as literature, videos, films, or music, which can evoke strong emotional experiences (Juslin and Laukka, 2004; Scherer et al., 2001; Scherer, 2004). Since the affect of the non-original content may be different from the social media users' affect when they are post this content, we differentiated between original and non-original content in our analysis.

This pilot study was designed to examine the following research questions:

1. **Changes in Affect:** To what extent do changes in the affect of social media posts correlate with users' personality traits and mental wellbeing?
2. **Originality:** To what extent does the use of non-original material in their posts correlate with users' personality traits and mental wellbeing?

Following best practice in sentiment analysis and opinion mining, we distinguish between positive, negative, neutral, and mixed (both positive and negative) affect (Moilanen and Pulman, 2007; Rosenthal et al., 2015; Agarwal et al., 2011).

We used a well known dataset, myPersonality (Bachrach et al., 2012; Youyou et al., 2015), that enriches Facebook posts with many validated psychological measures. In MyPersonality, positive mental wellbeing is measured using the Satisfaction with Life Scale (Diener et al., 1985, 1999), while the presence of depressive symptoms is assessed using the Centre for Epidemiologic Studies Depression scale (CES-D) (Radloff, 1977). Personality traits are established following the OCEAN model(McCrae and John., 1992), which consists of the five traits Openness to Experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism.

We included all 70 adult users who provided sufficient, regular Facebook data for two months before completion of the CES-D questionnaire, and corrected for multiple comparisons in our statistical analysis. We find that the transitions from one affective state to another expressed in social media posts give us a highly nuanced view of personality traits. While the amount of non-original posts in ones' social media status updates is closely linked to depression symptoms, this link is mediated by neuroticism.

## 2 BACKGROUND

Affect refers to both mood and emotion. Moods are slow-moving states that can be influenced by people, objects or situations, whereas emotions are quick reactions to stimuli (Rottenberg and Gross, 2003; Watson, 2000), and highly situation- or object-specific (Bylsma et al., 2008). Mood influences the probability of having emotions of the same valence—negative mood facilitates negative emotions, and positive mood makes positive emotions more likely (Rottenberg, 2005; Fredrickson, 1998). Affect is is an important predictor of mental wellbeing, including a person's overall satisfaction with life (Headey et al., 1993; Singh and Jha, 2008; Chen et al., 2017), and the level of symptoms of depression (Tsugawa et al., 2015; Coppersmith et al., 2015; Resnik et al., 2015).

Personality also predisposes people to certain affective states (Rothbart et al., 2000). While neuroticism is associated with negative affect (Pishva et al., 2011), positive affect is strongly linked to extroversion (Watson and Clark, 1997; Fujita et al., 1991). Extroverts experience more positive affect because they engage in more social situations (Diener and Emmons, 1984; Ryan and Deci, 2001). Individuals who score high on agreeableness have a greater ability to regulate negative affect (Meier et al., 2006; Haas et al., 2007). This relationship between affect and personality is also reflected in social media studies (Lin et al., 2017; Golbeck et al., 2011; Schwartz et al., 2013; Pennebaker and King, 1999). For example, people who use negative affective words in their social media posts tend to have lower conscientiousness, lower agreeableness (Golbeck et al., 2011), and higher neuroticism (Pennebaker and King, 1999).

In psychology, quantitative representations of affect are typically multidimensional (Russell, 1980). In this study, we focus on valence, which is represented in many classic affect models. Traditional measures, such as the Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988), report the strength of positive and negative valence. Mixed valence can occur when people experience 'dialectic' emotion, which is a mix of positive and negative emotions (Russell, 2003; Schimmack et al., 2002).

The personality trait measurements in myPersonality are based on Costa and McCrae's well-validated OCEAN model (McCrae and John., 1992). The model consists of five dimensions: extroversion, agreeableness, conscientiousness, neuroticism, and openness to experience. Neuroticism refers to the degree of emotional stability. Openness reflects the degree of creativity and curiosity. Conscientious individuals tend to be careful and diligent. Extroversion refers to a tendency to be energetic and friendly. Agreeableness reflects the tendency to be compassionate and to cooperate with others (Digman, 1990). The five-factor structure has proved to be robust in both self and peer ratings (McCrae and John., 1992), children and adult (Mervielde et al., 1995), across different cultures (McCrae and Allik, 2002), and stable over time (McCrae and John., 1992).

## 3 DATA AND METHODOLOGY

The myPersonality data set (Bachrach et al., 2012; Youyou et al., 2015) contains more than 180,000 Facebook users, enriched with a variety of additional validated scales (Bachrach et al., 2012). The collection of myPersonality data complied with the terms of Facebook service, informed consent for research use was obtained from all users, and researchers had to seek permission to use the dataset. Permission for the use of this database was obtained before it closed for new studies in 2018. The study was granted Ethical Approval by the Ethics Committee of the School of Informatics, University of Edinburgh.

### 3.1 Choice of Scales

From the extensive data collected within myPersonality, we chose two scales for quantifying mental wellbeing, the *Center for Epidemiologic Studies Depression Scale (CES-D)* and the *Satisfaction with Life Scale (SWL)*. The CES-D scale measures a key aspect of mental health, the presence of depression symptoms (Radloff, 1977). The scale has high internal consistency, test-retest reliability (Radloff, 1977; Orme et al., 1986; Roberts, 1980), and validity (Orme et al., 1986). Following previous social media studies (De Choudhury et al., 2013; Park et al., 2012), we adopt a score of 22 or higher as a cut-off value for likely depressive disorder (maximum score: 60). The 5-item SWL scale has been tested across different cultures and age groups (Pavot and Diener, 2009) and has been found to have high internal consistency and temporal reliability (Diener et al., 1985). Personality traits were measured using a 100 item scale using items from the open source International Personality Item Pool (Goldberg et al., 2006) that were validated against the the NEO-PI-R (Schwartz et al., 2013) instrument.

### 3.2 Selection of Participants

The data set was originally designed for a study of the effect of mental wellbeing and values on social media disclosure. We therefore selected only those participants who had completed the CES-D scale, the SWL scale, and the Schwartz Value survey (Schwartz, 1992) in addition to the full personality questionnaire. 301 participants in myPersonality provided full data for all four scales.

To ensure we had enough posts to assess the frequency of affect transitions, we only included users in our sample that regularly updated their public Facebook feed (*regular users*). We defined regular users as individuals who posted on average twice a week or more. We estimated posting frequency using the average post count per day during the sampling frame. If an individual had a post count per day of 0.3, this individual made around 110 posts in 365 days, which was roughly equivalent to an average of 2 posts per week. Of the original 301 participants, 122 (40.5%) were regular users.

Since the CES-D asks about symptoms in the past week, we excluded a further 31 users who had not posted any content in the week before completing the CES-D scale. We then focused on a 60-day span (two months) before CES-D completion, to ensure we had sufficient data to track the development of users' moods. We removed 14 users who contributed less than 20 posts during that time. Finally we removed four users who were under 18 year old and three users with more than 20% of the posts written in a language other than English, because English was the common language of the annotation team. The final sample consisted of 4086 posts from 70 users.

## 3.3 Corpus Annotation

### 3.3.1 Social Media Affect

For the purpose of this study, we refer to the affect shown in social media posts as *social media affect*. In this study, we operationalize valence as the post author's attitude towards a primary target of opinion, following (Mohammad, 2016). We refer to the 'dialectic' affective state as *mixed valence*. If there is no clear trend towards positive or negative affect, the associated valence is *neutral*.

After extensive piloting, we created an annotation guideline (available as part of the supplementary material) that was largely based on Mohammad (2016)'s work on defining the valence of a social media post. Each post is assigned one of four affect polarities: + (positive), - (negative), $\pm$ (mixed), or 0 (neutral). We used manual annotation since this is commonly used in computational linguistics to create a baseline gold standard data set for further analysis (Teufel, 1999).

Out of the 4086 posts, 2698 (66%) were annotated by a team of six trained annotators and 1185 (29%) by the first author. 5% of all posts were annotated by all seven annotators to establish inter-rater reliability, which was measured using Cohen's $\kappa$ (Gamer et al., 2012). Average inter-rater reliability between the first author and the annotators is 0.88, and 0.78 among the six annotators.

After annotation, most of the posts were of positive valence (N= 1588, 39%), followed by negative valence (N=1164, 28%), neutral valence (N=982, 24%) and mixed (N=312, 8%). 40 posts were excluded from analysis, since they did not contain English text.

### 3.3.2 Originality

We define posts that consist of quotes from sources such as song lyrics, books, or movies as non-original content; all other content was defined as original. Since non-original content might not directly reflect the user's moods or emotions, annotators were instructed to annotate such posts according to the likely emotions of the author. For example, if a post consists of an uplifting motivational quote, annotators considered the underlying valence to be positive.

In order to establish the originality of a post, we retrieved the first page of results obtained by searching for the post text using the Google API. For each web page on the first page of results, we computed the cosine similarity between the the post content and the page content. Posts with a cosine similarity greater than 0.96 were labeled as non-original, and posts with a cosine similarity between 0.92 and 0.96, where

**Table 1.** Affect and originality representation for a sample week

| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
|---|---|---|---|---|---|---|---|
| *Affect* | | | | | | | |
| Post-Level | + - - | + - + | + + | S | ± | 0 | + - |
| Day-Level | - | + | + | S | ± | 0 | ± |
| *Originality* | | | | | | | |
| Post-Level | O N O | O O N | N N | S | O | O | N N |

Note: $\leftrightarrow$, negative valence: $-$, positive valence: $+$, mixed valence:$\pm$, $S$: silence day, original content: O, non-original content: N

the website links or website names included the words 'lyrics' or 'quote' were labeled as potentially non-original. Posts with a cosine similarity lower than 0.92 were labeled as original. The cutoff points were determined based on a sample of 300 posts manually annotated for originality by the first author. On these posts, the classifier yields 100% recall, 81% precision, and an F1-score of 0.89. In our data set, 287 (7%) of all posts were identified as non-original.

### 3.4 Modelling Affect Transitions

We examine two types of transitions:

- **Post-Level versus Day-Level:** *Post-level* transitions focus on changes in affect between subsequent social media posts, whereas *day-level* transitions focus on changes in overall dominant affect between subsequent days.
- **Silence versus Non-Silence:** Not all users post every day. In our *default* models, these silent days are ignored, whereas in our *with-silence* models, days without posts are explicitly modelled as *Silence*.

The post-level social media affect is likely to be influenced by emphunderlying emotions, which change more quickly, whereas the day-level social media affect is likely to be influenced by *underlying mood* during the day. Day-level affect was calculated as follows. If the majority of the posts $p_{ij}$ on day $d_j$ have the same affect $a$, then the affect of day $d_j$ is set to $a$. If there is an equal number of positive (+) and negative (-) posts, or if the number of mixed affect ($\pm$) posts is equal to the number of posts with other types of affect, affect is set to $\pm$ (mixed). For transitions between original and non-original posts, we only consider the post-level representation. Table 1 shows an example of the affect and originality representations.

### 3.5 Statistical Analysis

Demographic differences between users above and below the CES-D cut-off score for probable depression were assessed using Wilcoxon-Mann-Whitney tests (R-package 'Stats').

We used Pearson correlation coefficients to assess the significance of correlations between social media data on one hand and personality traits and mental wellbeing on the other hand. Due to the small sample size and the number of correlations computed, all correlation coefficients were estimated using a permutation approach (Higgins, 2003) as implemented in the R Package `jmuOutlier` (Garren, 2017). Correlations that reach $p < 0.01$ or better are reported as significant; correlations that reach $p < 0.05$ are reported as trends in the data. For all correlations reported in the paper, we give the estimated correlation coefficient, the bootstrap 95% confidence interval, and the corresponding coefficient of determination $r^2$.

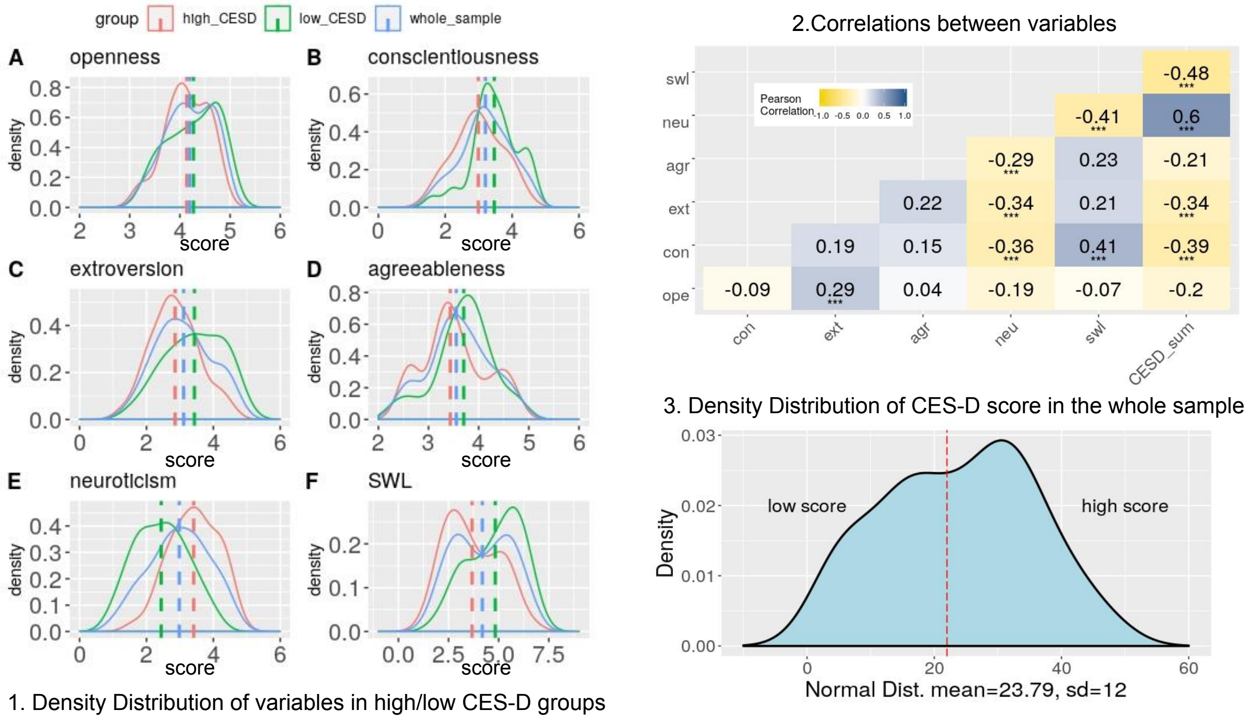


**Figure 1.** Basic statistics for personality trait scores, SWL and CES-D scores. Plot 1 shows density plots of the distribution of of personality traits and SWL for all participants, participants with CES-D $>=22$ (high CES-D), and participants with CES-D $< 22$ (low CES-D). The dotted line shows the median. Plot 2 is a heat map of correlations between personality traits, SWL and CES-D scores (***: $p < 0.001$). Plot 3 illustrates the distribution of the CES-score in the entire sample (N = 70). The dotted line indicates the cutoff score of 22.

# 4 RESULTS

## 4.1 Demographics and Baseline Statistics

Table 2 shows the basic statistics of our sample. Our data predominantly comes from single female Caucasian young adults. The average CES-D score is above the cut-off for possible depressive disorder.

Thirty-nine (56%) participants had a CES-D score of 22 or higher (mean: 33, SD: 6.5), which means that it is possible that that they have depressive disorder, and 31 (44%) had a score of 21 or lower (mean: 12, SD: 6). Figure 1 Plot 1 shows the density distributions of personality trait and SWL scores for three groups, the full sample, people above the cut-off, people below the cut-off..

Participants with possible depressive disorder are less extroverted (Z = 375, p<0.005), have higher levels of neuroticism (Z = 990, p<0.001), lower levels of conscientiousness (Z = 375, p<0.001), and lower satisfaction with life (Z = 323, p<0.001). Detailed results are reported in Figure 1 Plot 2.

All scales are normally distributed (Shapiro-Wilks test), except for openness to experience (W=0.96, $p < 0.05$), and satisfaction with life (W=0.95, $p < 0.05$), which are bimodal. Figure 1 Plot 2 shows the correlations between different personality dimensions. As expected, the five personality dimensions are not orthogonal.

**Table 2.** Demographics of the sample.

| Variable | N (%) | Variable | Mean (SD) |
|---|---|---|---|
| *Gender* | | *Age* | |
| - Female | 49 (70%) | - Female | 23.52 (6.56) |
| - Male | 21 (30%) | - Male | 22.84 (7.13) |
| *Ethnicity* | | *Personality* | |
| - Caucasian | 54 (75%) | - Openness to Experience | 4.19 (0.46) |
| - Black | 3 (4%) | - Conscientiousness | 3.20 (0.75) |
| - Asian | 5 (7%) | - Extraversion | 3.11 (3.83) |
| - other | 8 (14%) | - Agreeableness | 3.55 (0.68) |
| | | - Neuroticism | 2.98 (0.89) |
| *Living Status* | | *Mental Wellbeing* | |
| - Living with partner | 8 (10%) | - SWL | 4.18 (1.44) |
| - Single | 54 (77%) | - CES-D | 23.79 (11.86) |
| - Married | 5 (7%) | | |
| - Unknown | 3 (4%) | | |

Note: Caucasian includes White people of American, British, and other origins; Black includes African Americans and Black people from Europe. SWL: score for Satisfaction with Life Scale. CES-D: Center for Epidemiologic Studies Depression Scale,

## 4.2 Social Media Affect: Frequencies versus Transitions

For **overall frequencies of affect category**, the only clear correlation is between extroversion and positive content. Overall, more extroverted participants are more likely to have days where they make predominantly positive posts (r=0.29, $p < 0.01$, 95%CI = (-0.15, 0.32), $r^2$ = 0.08). In addition, participants who score higher on agreeableness tend to post fewer negative posts and have fewer days with predominantly negative posts (both r=-0.26, $p < 0.05$, 95%CI = (-0.48, -0.04), $r^2$ = 0.07).

When we look at **transitions between affect categories**, however, a more nuanced picture emerges. Table 3 summarises the correlations between personality, well being and transition types. Significant correlations are summarised in Table 4. Due to the number of correlations presented, we choose a cut-off of $p < 0.01$, which is stricter than the normal $p < 0.05$.

Several transition types are correlated positively and negatively with Extroversion and Agreeableness. Neuroticism, conscientiousness, and SWL show interesting trends ($p < 0.05$) that do not reach significance (c.f. Table 3

More extroverted participants are more likely to post predominantly positive content several days in a row (*day-level*, +↔+, r=0.30, $p < 0.001$, 95% CI = (0.06, 0.54), $r^2$=0.09). They have more transitions to or from a silence day with a positive post (*post-level with-silence*, S↔+, r=0.29, $p < 0.01$, 95% CI = (-0.01, 0.46), $r^2$=0.08). This pattern fits well the overall predominance of posts with positive affect. Extroverts are also less likely to alternate between days with neutral and days with non-neutral content (*day-level*, for both 0↔+ and 0↔-, r=-0.28, $p < 0.01$, 95% CI = (-0.52, -0.09), , $r^2$=0.08).

People who score higher on agreeableness are less likely to follow a post with negative affect with another negative affect post (-↔-, *post-level with-silence*: r=-0.37, $p < 0.001$, 95% CI = (-0.50, -0.06) , $r^2$=0.14); This tendency is much less pronounced on the day-level (-↔-, r=-0.22, $p < 0.1$, 95% CI = (-0.44, -0.02) , $r^2$=0.04). On top of that, they are more likely to alternate between days with mixed valence

**Table 3.** Correlations between personality, SWL, and CES-D scores and affect transitions. Number of participants N=70

| | S$\leftrightarrow$S | -$\leftrightarrow$- | +$\leftrightarrow$+ | $\pm\leftrightarrow\pm$ | 0$\leftrightarrow$0 | +$\leftrightarrow$- | $\pm\leftrightarrow$+ | $\pm\leftrightarrow$- | $\pm\leftrightarrow$0 | 0$\leftrightarrow$+ | 0$\leftrightarrow$- | S$\leftrightarrow$+ | S$\leftrightarrow$- | $\pm\leftrightarrow$S | S$\leftrightarrow$0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Post-level representation (Post Plus Silence) | | | | | | | | | | | | | | | |
| $N_{Occ}$ | 1238 | 346 | 542 | 29 | 230 | 599 | 143 | 134 | 100 | 424 | 414 | 641 | 384 | 137 | 211 |
| ope | 0.09 | -0.17 | -0.17 | -0.16 | -0.05 | -0.14 | -0.07 | -0.08 | 0.11 | 0.01 | 0.03 | 0.17 | 0.00 | 0.13 | 0.03 |
| con | -0.06 | 0.01 | 0.09 | -0.09 | -0.15 | 0.11 | 0.00 | -0.01 | -0.14 | -0.07 | -0.08 | 0.16 | 0.00 | 0.15 | -0.15 |
| ext | 0.04 | -0.12 | 0.16 | -0.10 | -0.19 | -0.06 | -0.03 | -0.12 | -0.09 | -0.09 | -0.17 | **0.29**$^{**}$ | -0.04 | 0.00 | -0.18 |
| agr | 0.14 | **-0.37**$^{***}$ | 0.03 | 0.02 | -0.15 | **-0.22**$^{*}$ | 0.08 | 0.04 | 0.04 | -0.04 | **-0.23**$^{*}$ | **0.23**$^{*}$ | -0.04 | **0.29**$^{**}$ | -0.13 |
| neu | -0.07 | 0.19 | 0.18 | 0.18 | -0.03 | **0.23**$^{*}$ | 0.11 | 0.04 | 0.02 | 0.05 | -0.05 | **-0.22**$^{*}$ | -0.03 | **-0.23**$^{*}$ | -0.13 |
| swl | 0.04 | -0.10 | -0.13 | -0.10 | 0.06 | -0.03 | 0.02 | -0.05 | -0.04 | 0.02 | -0.08 | 0.02 | 0.16 | -0.02 | 0.18 |
| CESD | -0.04 | 0.19 | 0.08 | 0.09 | 0.00 | 0.04 | 0.15 | 0.07 | 0.03 | -0.06 | 0.11 | -0.20$\cdot$ | 0.00 | -0.11 | -0.03 |
| Post-level representation (Post only), N = 70 | | | | | | | | | | | | | | | |
| $N_{Occ}$ | | 396 | 694 | 34 | 313 | 728 | 188 | 166 | 142 | 547 | 502 | | | | |
| ope | | -0.16 | -0.05 | -0.06 | -0.02 | -0.05 | 0.06 | -0.01 | 0.14 | 0.09 | 0.13 | | | | |
| con | | -0.07 | 0.18 | -0.07 | **-0.23**$^{*}$ | 0.08 | 0.14 | 0.10 | -0.11 | -0.13 | -0.12 | | | | |
| ext | | -0.04 | **0.33**$^{***}$ | 0.04 | **-0.24**$^{*}$ | 0.05 | 0.08 | -0.10 | -0.15 | -0.16 | -0.20$\cdot$ | | | | |
| agr | | **-0.28**$^{**}$ | 0.18 | 0.00 | -0.16 | -0.10 | **0.26**$^{*}$ | **0.28**$^{**}$ | 0.13 | 0.03 | **-0.26**$^{*}$ | | | | |
| neu | | 0.14 | 0.00 | 0.11 | -0.02 | 0.16 | -0.14 | -0.09 | -0.08 | 0.01 | -0.12 | | | | |
| swl | | 0.00 | -0.12 | -0.11 | 0.11 | 0.02 | 0.09 | 0.09 | -0.02 | 0.08 | -0.04 | | | | |
| CESD | | 0.14 | -0.04 | 0.03 | 0.04 | -0.03 | -0.06 | -0.11 | 0.04 | -0.11 | 0.13 | | | | |
| Day-level representation, N = 70 | | | | | | | | | | | | | | | |
| $N_{Occ}$ | 228 | 281 | 271 | 267 | 304 | 287 | 303 | 296 | 298 | 261 | 311 | 242 | 259 | 261 | 261 |
| ope | 0.12 | -0.17 | -0.11 | -0.05 | -0.02 | -0.08 | 0.00 | -0.14 | -0.07 | -0.01 | -0.02 | 0.12 | -0.02 | 0.19 | 0.13 |
| con | -0.06 | -0.03 | **0.25**$^{*}$ | 0.05 | -0.01 | 0.03 | -0.03 | -0.04 | -0.16 | -0.19 | -0.12 | 0.08 | 0.10 | 0.06 | -0.07 |
| ext | 0.06 | -0.11 | **0.30**$^{***}$ | -0.03 | -0.14 | 0.04 | 0.14 | -0.13 | 0.01 | **-0.28**$^{**}$ | **-0.28**$^{**}$ | **0.24**$^{*}$ | -0.08 | 0.02 | -0.17 |
| agr | 0.11 | -0.22$\cdot$ | 0.15 | -0.05 | 0.08 | -0.12 | 0.16 | -0.06 | 0.11 | -0.08 | -0.17 | 0.15 | -0.07 | **0.28**$^{**}$ | -0.09 |
| neu | -0.08 | 0.16 | 0.00 | 0.19 | -0.17 | **0.21**$^{*}$ | 0.09 | 0.11 | -0.01 | 0.12 | 0.08 | -0.14 | -0.12 | **-0.26**$^{*}$ | -0.03 |
| swl | 0.02 | -0.08 | -0.01 | -0.08 | **0.25**$^{*}$ | -0.03 | -0.06 | -0.10 | 0.03 | -0.06 | -0.04 | -0.02 | 0.12 | 0.06 | 0.08 |
| CESD | -0.03 | 0.11 | -0.10 | 0.08 | -0.18 | 0.02 | 0.10 | 0.08 | 0.08 | -0.01 | 0.21$\cdot$ | -0.18 | 0.03 | -0.16 | 0.05 |

Note: Pearson correlation P-value (permutation testing): $\cdot < 0.1$, $* < .05$, $** < .01$, $*** < .001$, bidirectional transition types: $\leftrightarrow$, negative valence: $-$, positive valence: $+$, mixed valence: $\pm$, neutral: 0, silence day: $S$, $N_{Occ}$: number of occurrences of each transition type, ope: openness, con: conscientiousness, ext: extraversion, agr: agreeableness, neu: neuroticism, swl: Satisfaction with Life Scale, CESD: Center for Epidemiologic Studies Depression Scale

**Table 4.** Summary of the significant correlations between transition states and the five personality traits ($p < 0.01$

| | Transitions | Post-Level (with-silence) | Post-Level (without-silence) | Day-Level |
|---|---|---|---|---|
| Extraversion | S $\leftrightarrow$ + | $\uparrow$ | — | — |
| | 0 $\leftrightarrow$ + | — | — | $\downarrow$ |
| | 0 $\leftrightarrow$ - | — | — | $\downarrow$ |
| | + $\leftrightarrow$ + | — | $\uparrow$ | $\uparrow$ |
| Agreeablness | - $\leftrightarrow$ - | $\downarrow$ | $\downarrow$ | — |
| | $\pm \leftrightarrow$ S | $\uparrow$ | — | $\uparrow$ |
| | $\pm \leftrightarrow$ - | — | $\uparrow$ | — |

Note: $\downarrow$ indicates a significant negative correlation at $p < 0.01$ or better, $\uparrow$ indicates a significant positive correlation at $p < 0.01$ or better. — indicates that the correlation is not significant at this level. Bidirectional transition types: $\leftrightarrow$, negative valence: $-$, positive valence: $+$, mixed valence:$\pm$ , neutral: 0, silence day: $S$.

and silence (*day-level*, $\pm\leftrightarrow$S, r=0.28, $p < 0.01$, 95% CI = (-0.01, 0.46), $r^2$=0.08, post-level with-silence, $\pm\leftrightarrow$S, r=0.29, $p < 0.01$, 95% CI = (0.08, 0.52), $r^2$=0.08) .

Participants with higher neuroticism tend to alternate between positive and negative content, but this is only evident when we take silence into account (+$\leftrightarrow$-, *post-level with-silence*: r=0.23, $p < 0.05$, 95% CI = (0.00, 0.47), $r^2$=0.04, *post-level without-silence*: r=0.16, 95% CI = (-0.08, 0.41), $r^2$=0.025, *day-level*: r=0.21, $p < 0.05$, 95% CI = (-0.46, -0.10), $r^2$=0.04).

There are interesting differences in transition patterns that incorporate information about silence days and those that do not. When disregarding silence days, we observe that people with higher conscientiousness or extroversion are slightly less likely to follow a neutral post with another neutral post (*post-level without-silence*, conscientiousness, 0$\leftrightarrow$0, r= -0.23, $p < 0.05$, 95% CI = (-0.41, -0.04), $r^2$=0.07; extroversion, 0$\leftrightarrow$0, r = -0.24, $p < 0.05$, 95% CI = (-0.41, -0.04), $r^2$=0.07).

When we take into account silence days for computing transitions, we find several more interesting trends. People who are more satisfied with life are more likely to follow a neutral post with another neutral post (0$\leftrightarrow$0, *day-level*: r=0.25, $p < 0.05$, 95% CI = (-0.01, 0.44), $r^2$=0.06). In addition, people with higher neuroticism are more likely to alternate between positive and negative posts (0$\leftrightarrow$-, *day-level*: r=0.21, $p < 0.05$, 95% CI = (-0.01, 0.40), $r^2$=0.04), but less likely to make a positive post after a period of one or more silence days (S$\leftrightarrow$+, *post-level with-silence*: r=-0.22, $p < 0.05$, 95% CI = (-0.48, 0.00), $r^2$=0.04). We found that silence to silence transitions are not correlated with personality or mental health.

### 4.3 Post Originality

High CES-D scores are significantly correlated with posting non-original content (r=0.29, $p < 0.01$, 95% CI = (0.10, 0.46), $r^2$=0.08). There is a similar tendency for participants with higher neuroticism scores (r=0.25, $p < 0.05$, 95% CI = (0.06, 0.43), $r^2$=0.07). Examining transitions between post originality shows that these effects stem from slightly different posting patterns. Users with higher CES-D scores tend to follow non-original content with non-original content (N$\leftrightarrow$N, *post-level with-silence*, r=0.26, $p < 0.05$, 95% CI = (0.07, 0.43), $r^2$=0.07) or to alternate between original and non-original content (N$\leftrightarrow$O *post-level with-silence*, r=0.27, $p < 0.05$, 95% CI = (0.08, 0.44), $r^2$=0.07). Users with higher neuroticism scores tend to post sequences of non-original content (N$\leftrightarrow$N, post-level with-silence, r=0.25, $p < 0.05$, 95% CI = (0.06, 0.43), $r^2$=0.05), and are less likely to post original content before or after a period of silence (O$\leftrightarrow$S, *post-level with-silence*, r=0.28, $p < 0.05$, 95% CI = (0.09, 0.45), $r^2$=0.08).

Since neuroticism is closely linked to depression symptoms, we also computed a partial correlation between content originality and CES-D while controlling for neuroticism. The resulting correlation was no longer significant (r=0.14, $p = 0.22$, $r^2$=0.02). Therefore, the association between content originality and depression symptoms might be moderated by neuroticism.

## 5 DISCUSSION

### 5.1 Main Findings

Many studies have found associations between the frequency of affective words used in social media text and personality. However, existing studies often see affect as static and only focused on the strength of bipolar valence (positive/negative). Instead, our work focuses on affect patterns. We encode posting behaviour, transitions between affect states, and content originality. From a practical point of view, our technique can supplement experience sampling techniques (Myin-Germeys et al., 2018) to help clinicians and patients develop a more comprehensive view of a person's affect patterns, arrive at a better substantiated diagnosis, and make improved treatment decisions. However, this depends on whether the patient is willing to share information from their social media feed with their therapist.

Overall, the correlations seen between affect transitions and personality traits are in line with the consensus in the early literature (Gross et al., 1998). Extroverts tend to produce sequences of positive posts. This behaviour fits well with the positive emotional core in extroverts stipulated in (Watson and Clark, 1997). Participants with higher agreeableness are less likely to post sequences of negative posts. This could be due to their ability to regulate negative affect (Meier et al., 2006; Haas et al., 2007).

Although the psychology literature suggests a strong association between negative mood states and neuroticism (Rusting and Larsen, 1995), we did not find this in our data. Our results are in line with previous studies of verbal cues to personality traits in social media (Schwartz et al., 2013; Yarkoni, 2010; Golbeck et al., 2011; Park et al., 2015). Golbeck et al. (2011) found social media users who were more likely to talk about anxiety were on the higher end of the neuroticism scale. We speculate that self-presentation bias may influence how social media users regulate their expression of negative emotions in their public posts. The only relevant association we found was that social media users on the high end of neuroticism are more likely to switch between posting positive and negative affective content. This finding aligns well with the fact that high neuroticism is associated with high emotional instability (Costa and McCrae, 1992).

The link between posting non-original content and elevated depression symptoms appears to be moderated by neuroticism. This suggests that high levels of neuroticism predispose users both to depressive symptoms, and to an indirect disclosure of emotions through quotes and lyrics.

In our sample, the prevalence of depressive symptoms is higher than what would be expected in general population. In the original CES-D paper, Radloff (1977) proposed three levels of depression severity: low (0-15), mild-to-moderate (16-22), and high (23-60). They found that only 21% of the general population scored above the low symptom level. In contrast, in our sample, nearly half of the participants exhibit a high level of symptoms (>22). Within the context social media studies of depression, however, our data set is not exceptional. For many studies in the area, high symptom individuals account for nearly half of the data set (Reece et al., 2017; Tsugawa et al., 2015; Nadeem, 2016; De Choudhury et al., 2013; Orabi et al., 2018).

Our results support the claim that affect expressed in social media data text is associated with social media users' affect patterns in real life. However, the data set used in this study is from the early 2010's, and only covers the well established social media platform Facebook. The associations found in this study are likely to be slightly different from those found in another social network (e.g., Instagram) or in a new data set collected ten years later.

### 5.2 Limitations.

Due to the restrictions imposed by the need for sufficient Facebook updates to allow analysis, our final sample is relatively small. Given the size of the significant effects we found in the data, power calculations indicate that a well-powered study should include data from around 200 users (Schönbrodt and Perugini, 2013). It also skews heavily towards younger female Caucasians with relatively low satisfaction with life and strong depression symptoms. It is possible that other groups of users (e.g., non-Caucasians, males) are less likely to disclose personal information about mood and emotions on their public Facebook (Dosono et al., 2017; McDonald et al., 2019).

## 6 CONCLUSION AND FUTURE WORK

In this pilot study, we demonstrated the benefits of detailed representations of social media affect for unpacking the relationship between personality, mental wellbeing, and the content posted on social media. Importantly, our representations include non-binary affect categories (positive, negative, mixed, neutral), and take into account content originality. As a consequence, we were able to obtain a more detailed picture of the link between patterns of affect and depressive symptoms.

In future work, we plan to enrich our data set with more in-depth analyses of original versus non-original content, extend coverage by including a larger sample of the myPersonality data set, and construct statistical models that allow us to observe long-term trends in posting patterns. Future studies should also examine the extent to which affect expressed in non-original content is aligned with the users' affect when they post the material.

## ACKNOWLEDGEMENTS

We thank Michael Kosinski and David Stilwell for permission to use myPersonality, and our six undergraduate Research Assistants from the Psychology Department of the University of Edinburgh for their hard annotation work. Walid Magdy's and Maria Wolters' work on this paper was partly funded by The Alan Turing Institute (EPSRC, EP/N510129/1).

## REFERENCES

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)* (Portland, Oregon). 30–38.

Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell. 2012. Personality and patterns of Facebook usage. In *Proceedings of the 4th annual ACM web science conference* (New York). ACM, 24–32.

Alexandra Balahur, Ralf Steinberger, Erik van der Goot, Bruno Pouliquen, and Mijail Kabadjov. 2009. Opinion mining on newspaper quotations. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 03* (Milano, Italy). IEEE Computer Society, 523–526.

Natalya N Bazarova, Yoon Hyung Choi, Victoria Schwanda Sosik, Dan Cosley, and Janis Whitlock. 2015. Social sharing of emotions on Facebook: Channel differences, satisfaction, and replies. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing* (Vancouver, BC, Canada). ACM, 154–164.

Lauren M Bylsma, Bethany H Morris, and Jonathan Rottenberg. 2008. A meta-analysis of emotional reactivity in major depressive disorder. *Clinical psychology review* 28, 4 (2008), 676–691.

Lauren M Bylsma, April Taylor-Clift, and Jonathan Rottenberg. 2011. Emotional reactivity to daily events in major and minor depression. *Journal of abnormal psychology* 120, 1 (2011), 155.

Gustavo Carlo, Maria Vicenta Mestre, Meredith M McGinley, Paula Samper, Ana Tur, and Deanna Sandman. 2012. The interplay of emotional instability, empathy, and coping on prosocial and aggressive behaviors. *Personality and Individual Differences* 53, 5 (2012), 675–680.

Lushi Chen, Christopher Hon Kwong Cheng, and Tao Gong. 2020. Inspecting Vulnerability to Depression from Social Media Affect. *Frontiers in Psychiatry* 11 (2020), 54.

Lushi Chen, Tao Gong, Michal Kosinski, David Stillwell, and Robert L Davidson. 2017. Building a profile of subjective well-being for social media users. *PloS one* 12, 11 (2017), e0187278.

Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying mental health signals in Twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality* (Denver, Colorado). 51–60.

Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. 2015. CLPsych 2015 shared task: Depression and PTSD on Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. 31–39.

Paul T Costa and Robert R McCrae. 1992. *Neo Pi-R*. Psychological Assessment Resources Odessa, FL.

Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth International AAAI Conference on Weblogs and Social Media* (Ann Arbor, Michigan,). 21–30.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media* (Cambridge, Massachusetts, USA). 170–185.

Ed Diener and Robert A Emmons. 1984. The independence of positive and negative affect. *Journal of personality and social psychology* 47, 5 (1984), 1105.

Ed Diener, Eunkook M Suh, Richard E Lucas, and Heidi L Smith. 1999. Subjective well-being: Three decades of progress. *Psychological bulletin* 125, 2 (1999), 276.

E. D. Diener, Robert A. Emmons, Randy J. Larsen, and Sharon Griffin. 1985. The satisfaction with life scale. *Journal of personality assessment* 49, 1 (1985), 71–75.

John M Digman. 1990. Personality structure: Emergence of the five-factor model. *Annual review of psychology* 41, 1 (1990), 417–440.

Bryan Dosono, Yasmeen Rashidi, Taslima Akter, Bryan Semaan, and Apu Kapadia. 2017. Challenges in Transitioning from Civil to Military Culture: Hyper-Selective Disclosure Through ICTs. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (Dec. 2017), 41:1–41:23. `https://doi.org/10.1145/3134676`

Golnoosh Farnadi, Susana Zoghbi, Marie-Francine Moens, and Martine De Cock. 2013. Recognising personality traits using Facebook status updates. In *Seventh International AAAI Conference on Weblogs and Social Media* (Cambridge, Massachusetts, USA). 154–164.

Barbara L Fredrickson. 1998. What good are positive emotions? *Review of general psychology* 2, 3 (1998), 300–319.

Nico H Frijda. 1993. Moods, emotion episodes, and emotions. *Handbook of emotions* 12, 2 (1993), 155.

Frank Fujita, Ed Diener, and Ed Sandvik. 1991. Gender differences in negative affect and well-being: the case for emotional intensity. *Journal of personality and social psychology* 61, 3 (1991), 427.

Matthias Gamer, Jim Lemon, Ian Fellows, and Puspendra Singh. 2012. irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84. *Internet resource: http://CRAN. R-project. org/package= irr](Verified April 10, 2013)* (2012).

Steven T Garren. 2017. Permutation Tests for Nonparametric Statistics Using R. *Asian Research Journal of Mathematics* (2017), 1–8.

Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. 2011. Predicting personality from twitter. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE, 149–156.

L R Goldberg, J A Johnson, H W Eber, R Hogan, M C Ashton, C R Cloninger, and H G Gough. 2006. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality* 40 (2006), 84–96.

James J Gross, Steven K Sutton, and Timothy Ketelaar. 1998. Relations between affect and personality: Support for the affect-level and affective-reactivity views. *Personality and social psychology bulletin* 24, 3 (1998), 279–288.

Brian W Haas, Kazufumi Omura, R Todd Constable, and Turhan Canli. 2007. Is automatic emotion regulation associated with agreeableness? A perspective using a social neuroscience approach. *Psychological Science* 18, 2 (2007), 130–132.

Bruce Headey, Jonathan Kelley, and Alex Wearing. 1993. Dimensions of mental health: Life satisfaction, positive affect, anxiety and depression. *Social indicators research* 29, 1 (1993), 63–82.

J.J. Higgins. 2003. *Introduction to Modern Non-Parametric Statistics.*

Marlies Houben, Wim Van Den Noortgate, and Peter Kuppens. 2015. The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological bulletin* 141, 4 (2015), 901.

Patrik N Juslin and Petri Laukka. 2004. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of new music research* 33, 3 (2004), 217–238.

Junjie Lin, Wenji Mao, and Daniel D Zeng. 2017. Personality-based refinement for sentiment classification in microblog. *Knowledge-Based Systems* 132 (2017), 204–214.

Robert R. McCrae and Juri Allik (Eds.). 2002. *The five-factor model of personality across cultures.* Springer Science and Business Media.

Robert R. McCrae and Oliver P. John. 1992. An introduction to the five-factor model and its applications. *Journal of personality* 60, 2 (1992), 175–215.

James McDonald, Kate Lockwood Harris, and Jessica Ramirez. 2019. Revealing and Concealing Difference: A Critical Approach to Disclosure and an Intersectional Theory of "Closeting". *Communication Theory* (2019). `https://doi.org/10.1093/ct/qtz017`

Brian P Meier, Michael D Robinson, and Benjamin M Wilkowski. 2006. Turning the other cheek: Agreeableness and the regulation of aggression-related primes. *Psychological Science* 17, 2 (2006), 136–142.

Ivan Mervielde, Veerle Buyst, and Filip De Fruyt. 1995. The validity of the Big-Five as a model for teachers' ratings of individual differences among children aged 4–12 years. *Personality and Individual Differences* 18, 4 (1995), 525–534.

Saif Mohammad. 2016. A practical guide to sentiment annotation: Challenges and solutions. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (San Diego, California). 174–179.

Karo Moilanen and Stephen Pulman. 2007. Sentiment composition. In *Proceedings of the Recent Advances in Natural Language Processing International Conference* (Borovets, Bulgaria). 378–382.

Inez Myin-Germeys, Zuzana Kasanova, Thomas Vaessen, Hugo Vachon, Olivia Kirtley, Wolfgang Viechtbauer, and Ulrich Reininghaus. 2018. Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry* 17, 2 (2018), 123–132. `https://doi.org/10.1002/wps.20513` arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/wps.20513

Moin Nadeem. 2016. Identifying depression on Twitter. *arXiv preprint arXiv:1607.07384* (2016).

Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, and Diana Inkpen. 2018. Deep learning for depression detection of twitter users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic* (New Orleans, LA). 88–97.

John G. Orme, Janet Reis, and Elicia J. Herz. 1986. Factorial and discriminant validity of the center for epidemiological studies depression (CES-D) scale. *Journal of clinical psychology* 42, 1 (1986), 28–33.

Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin EP Seligman. 2015. Automatic personality assessment through social media language. *Journal of personality and social psychology* 108, 6 (2015), 934.

Minsu Park, Chiyoung Cha, and Meeyoung Cha. 2012. Depressive moods of users portrayed in Twitter. In *Proceedings of the ACM SIGKDD Workshop on healthcare informatics (HI-KDD)*, Vol. 2012. ACM New York, NY, 1–8.

William Pavot and Ed Diener. 2009. Review of the satisfaction with life scale. *Psychological assessment* 5, 2 (2009), 164.

James W Pennebaker and Laura A King. 1999. Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology* 77, 6 (1999), 1296.

Nooshin Pishva, Maryam Ghalehban, Afsane Moradi, and Leila Hoseini. 2011. Personality and happiness. *Procedia-Social and Behavioral Sciences* 30 (2011), 429–432.

Lenore Sawyer Radloff. 1977. The CES-D scale: a self-report depression scale for research in the general population. *Applied psychological measurement* 1, 3 (1977), 385–401.

Andrew G Reece, Andrew J Reagan, Katharina LM Lix, Peter Sheridan Dodds, Christopher M Danforth, and Ellen J Langer. 2017. Forecasting the onset and course of mental illness with Twitter data. *Scientific reports* 7, 1 (2017), 13006.

Philip Resnik, William Armstrong, Leonardo Claudino, Thang Nguyen, Viet-An Nguyen, and Jordan Boyd-Graber. 2015. Beyond LDA: Exploring supervised topic modeling for depression-related language in Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (Denver, Colorado). 99–107.

Robert E. Roberts. 1980. Reliability of the CES-D scale in different ethnic contexts. *Psychiatry research* 2, 2 (1980), 125–134.

Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. Semeval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)* (Denver, Colorado). 451–463.

Mary K Rothbart, Stephan A Ahadi, and David E Evans. 2000. Temperament and personality: origins and outcomes. *Journal of personality and social psychology* 78, 1 (2000), 122.

Jonathan Rottenberg. 2005. Mood and emotion in major depression. *Current Directions in Psychological Science* 14, 3 (2005), 167–170.

Jonathan Rottenberg and James J Gross. 2003. When emotion goes wrong: Realizing the promise of affective science. *Clinical Psychology: Science and Practice* 10, 2 (2003), 227–232.

James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

James A Russell. 2003. Core affect and the psychological construction of emotion. *Psychological review* 110, 1 (2003), 145.

Cheryl L Rusting. 1998. Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. *Psychological bulletin* 124, 2 (1998), 165.

Cheryl L Rusting and Randy J Larsen. 1995. Moods as sources of stimulation: Relationships between personality and desired mood states. *Personality and individual differences* 18, 3 (1995), 321–329.

Richard M Ryan and Edward L Deci. 2001. On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual review of psychology* 52, 1 (2001), 141–166.

Klaus R Scherer. 2004. Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of new music research* 33, 3 (2004), 239–251.

Klaus R Scherer, Marcel R Zentner, et al. 2001. Emotional effects of music: Production rules. *Music and emotion: Theory and research* 361, 2001 (2001), 392.

Ulrich Schimmack, Shigehiro Oishi, and Ed Diener. 2002. Cultural influences on the relation between pleasant emotions and unpleasant emotions: Asian dialectic philosophies or individualism-collectivism? *Cognition & Emotion* 16, 6 (2002), 705–719.

Felix D Schönbrodt and Marco Perugini. 2013. At what sample size do correlations stabilize? *Journal of Research in Personality* 47, 5 (2013), 609–612.

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), 234.

Shalom H Schwartz. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*. Vol. 25. Elsevier, 1–65.

Gal Sheppes, Gaurav Suri, and James J Gross. 2015. Emotion regulation and psychopathology. *Annual review of clinical psychology* 11 (2015), 379–405.

David H Silvera, Anne M Lavack, and Fredric Kropp. 2008. Impulse buying: the role of affect, social influence, and subjective wellbeing. *Journal of Consumer Marketing* 25, 1 (2008), 23–33.

Kamlesh Singh and Shalini Duggal Jha. 2008. Positive and negative affect, and grit as predictors of happiness and life satisfaction. *Journal of the Indian Academy of Applied Psychology* 34, 2 (2008), 40–45. `https://doi.org/10.1038/s41746-020-0233-7`

Simone Teufel. 1999. *Argumentative Zoning: Information Extraction from Scientific Articles*. Ph.D. Dissertation. Centre for Cognitive Science, University of Edinburgh.

Renee J Thompson, Jutta Mata, Susanne M Jaeggi, Martin Buschkuehl, John Jonides, and Ian H Gotlib. 2012. The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of abnormal psychology* 121, 4 (2012), 819.

Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. Recognizing depression from twitter activity. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 3187–3196.

David Watson. 2000. *Mood and temperament*. Guilford Press.

David Watson and Lee Anna Clark. 1997. Extraversion and its positive emotional core. In *Handbook of personality psychology*. Elsevier, 767–793.

David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063.

Tal Yarkoni. 2010. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality* 44, 3 (2010), 363–373.

Wu Youyou, Michal Kosinski, and David Stillwell. 2015. Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences* 112, 4 (2015), 1036–1040.