



Valence norms for 3,600 English words collected during the COVID-19 pandemic: Effects of age and the pandemic

Aki-Juhani Kyröläinen^{1,2} · Javon Luke² · Gary Libben² · Victor Kuperman¹

Accepted: 3 November 2021 / Published online: 16 December 2021
© The Psychonomic Society, Inc. 2021

Abstract

The topic of affective development over the lifespan is at the forefront of psychological science. One of the intriguing findings in this area is superior emotion regulation and increased positivity in older rather than younger adults. This paper aims to contribute to the empirical base of studies on the role of affect in cognition. We report a new dataset of valence (positivity) ratings to 3,600 English words collected from North American and British English-speaking younger (below 65 years of age) and older adults (65 years of age and older) during the COVID-19 pandemic. This dataset represents a broad range of valence and a rich selection of semantic categories. Our analyses of the new data pitted against comparable pre-pandemic (2013) data from younger counterparts reveal differences in the overall distribution of valence related both to age and the psychological fallout of the pandemic. Thus, we found at the group level that older participants produced higher valence ratings overall than their younger counterparts before and especially during the pandemic. Moreover, valence ratings saw a super-linear increase after the age of 65. Together, these findings provide new evidence for emotion regulation throughout adulthood, including a novel demonstration of greater emotional resilience in older adults to the stressors of the pandemic.

Keywords Valence · Aging · COVID-19 · Affect · Emotion regulation

Introduction

Language is a powerful and precise tool for identifying the overall emotional well-being of a person or a group, as well as their emotional responses to specific phenomena, events, and objects in their environment. This notion underlies a thriving psychological and linguistic literature on how language communicates emotion and also constructs its perception (see e.g., Boyd and Schwartz (2021); Davitz (1969); Lindquist (2017); Majid (2012)). Yet neither language use nor its emotional tenor remains stable over the lifetime of an individual; both change. Proposed reasons for these changes include physiological, psychological, and cognitive aspects of aging, accumulation of experience in both linguistic and

affective domains, as well as changes in the person's physical and social environment Panksepp and Miller (1996); Ramscar et al. (2014, (2017); Urry and Gross (2010); Wahl and Lang (2003). The dynamism on both sides of the language-emotion connection has given rise to a question which is at the core of the present paper. Does language faithfully reflect changes in the emotional tenor of an individual that come with aging? If this question is answered in the affirmative, researchers can use linguistic data as a window into the affective state of an aging person. The goal of the present paper is to highlight emotional characteristics of aging as they emerge in language data obtained from older individuals. Often, this characterization relies on comparisons of affective language in older vs younger cohorts. We pursue this goal and contribute to the broader research inquiry in two ways. We make available a novel database of psychological valence (positivity) ratings to 3,600 English words produced by younger (below 65 years of age) and older adults (65 years old and above) and collected during the COVID-19 pandemic in 2020-2021. Furthermore, we analyze the new data against pre-pandemic comparable data from a younger population (2013), putting an emphasis on distributional characteristics of the language data in which age plays the most salient role.

✉ Aki-Juhani Kyröläinen
akkyro@gmail.com

¹ Department of Linguistics and Languages, McMaster University, Togo Salmon Hall 513, 1280 Main Street West, Ontario, L8S 4M2 Hamilton, Canada

² Department of Applied Linguistics, Brock University, 1812 Sir Isaac Brock Way, Ontario, L2S 3A1, St. Catharines, Canada

Perhaps the most robust and intriguing finding at the intersection of emotion and aging is that older individuals self-report and demonstrate higher levels of emotional well-being than their younger counterparts Carstensen et al. (2006); Urry and Gross (2010). This is despite the typically observed increasing constraints on many aspects of life in aging individuals. An oft-cited explanation for this observation is that older adults allocate more attention to emotion regulation, i.e., strategies people use to “influence which emotions they have, when they have them, and how they experience and express them” Gross (1998)[p. 271]. The advantage in emotion regulation is evident in that older adults are less likely to either experience or express negative emotions than positive ones Mather and Carstensen (2005): lower levels of negative affect and higher levels of positive affect in aging have been confirmed both in cross-sectional and longitudinal studies Cacioppo et al. (2008); Charles et al. (2001); Stawski et al. (2008). In the domain of language, the proposal of age-related differences in emotion regulation has seen experimental support as well. For instance, older adults showed a better ability to preferentially select, allocate more attention to, and show better memorization for positive (linguistic and other) stimuli rather than negative ones, compared to their younger counterparts (see reviews by Carstensen and DeLiema (2018); Reed et al. (2014)).

Strong support for the persistence of the language-emotion connection throughout the lifespan has come from a large-scale study by Kyröläinen et al. (2021). They analyzed word frequency distributions in English texts produced as Facebook updates by over 20,000 individuals throughout adulthood. With advancing age, writers showed an increasingly more frequent use of positive words and also used a greater variety of positive words. Indeed, the positivity bias seems to be inherent in human languages across all age groups (e.g., Dodds et al. (2015); Warriner and Kuperman (2015)). The finding of a gradually increasing positivity bias in language use as a function of age is compatible with the notion that aging translates into a growing experience with a gamut of emotional states and grants one an increasing ability to modulate what emotions to experience the most strongly, when and how.

While the prior literature corroborates the notion that language use is reflective of age-related affective changes, the research field can benefit both from larger-scale and more nuanced data. One of the reasons is that all existing studies that base their analyses on frequency distributions of emotion-laden words across age groups, including Kyröläinen et al. (2021) and Schwartz et al. (2013), share the same methodological limitation. Namely, they assume that positive words are equally positive and negative ones are equally negative for all age groups. While it is likely to be true as an overall tendency (e.g., the word

ice-cream elicits a more positive emotion than the word *rapist* in any age), aging may lead to subtle emotional shifts in the entire lexicon or in specific semantic fields. Also, age-driven changes in one’s physiological, mental, and social life may lead to different emotional responses to words particularly strongly related to certain stages of life. For instance, Schwartz et al. (2013) demonstrated that words most diagnostic of older age primarily concern family and health, those particularly strongly associated with middle-age relate to jobs, careers and children, while those identifying teenagers contained a larger number of swear-words and words related to school and education. It stands to reason that semantic fields of particular interest to older vs younger individuals will elicit different emotional responses in the respective groups. Analyses of frequency distributions cannot tap into this valuable information about the change in affect over age.

Our study remedies this limitation by recruiting an additional source of relevant evidence. If indeed older adults have a stronger positivity bias thanks to better emotion regulation, this bias may not only manifest itself in a preference towards positive expressions or stimuli. It may also surface in more positive evaluations of lexical stimuli, relative to valence judgments of younger counterparts. Our dataset of valence ratings and associated analyses are among the first to explore this possibility (see also Liu et al. (2021)).

Emotion regulation and a related concept of emotional resilience – the ability to adapt to stressors in the face of adversity – came to the forefront of psychological research in the times of the COVID-19 pandemic, during which our data were collected (e.g., Killgore et al. (2020); Kyröläinen and Kuperman (submitted); Shanahan et al. (2020)). In all age groups, the unprecedented disruption caused by the pandemic and protective measures, including the global lockdown, has led to an increased incidence of mental illnesses, suicides, and feelings of anxiety, depression, loneliness and social isolation Cullen et al. (2020); Kontoangelos et al. (2020); Lwin et al. (2020); Pfefferbaum and North (2020). Not only are older adults at the greatest medical risk of severe illness from COVID-19, this population has also been in the epicentre of protective measures since the beginning of the pandemic, experiencing longer and more restrictive periods of physical isolation and limited social interactions. Thus, it is possible that the pandemic-related stressors are not distributed uniformly over the lifespan and that older adults have been living in particularly psychologically taxing environments. Whether older adults respond by exhibiting particularly lower levels of positivity in their judgments during COVID-19 relative to their younger peers, or whether they show a greater emotional resilience and thus compensate for the harmful impact of the pandemic is one of the questions this paper asks.

The present study

This paper pursues two goals: first, we aim to present the research community with a new source of linguistic data on aging, i.e, a new dataset of valence (positivity) ratings for 3,600 English words collected from younger (< 65 y.o.) and older (≥ 65 y.o) English-speaking participants during the COVID-19 pandemic. We are aware of only one similar resource that targets older individuals specifically Liu et al. (2021) and none in English. The second goal of the paper is to examine the link between language and emotion as a function of age and the psychological fallout of the COVID-19. We do so by comparing younger and older cohorts of the present resource with existing data from a younger population Warriner et al. (2013), as is detailed below.

The focus of our second goal is on whether valence ratings reflect hypothesized age-related differences in emotion regulation as well as differences in emotional response to the pandemic. Most previous studies have only tested either younger or older participants (e.g., Kyröläinen and Kuperman (submitted); Shanahan et al. (2020), or only examined responses to a small number of either positive or negative words, and also used words denoting emotional states (e.g., *anger*, *joy*). We present participants with selections of words that occupy the entire range of valence and while many of the words are emotionally charged, very few denote emotions (see details below). This approach enables us to test the ability of older individuals for emotion regulation when confronted with emotionally positive, neutral, and negative stimuli. It is possible that the stronger positivity bias observed by Kyröläinen et al. (2021) in word frequency distributions of texts produced by older adults may also emerge in their ratings of psychological valence: e.g., older participants would provide

higher average ratings to the same words when compared to younger respondents. Yet, unlike in natural written productions studied in Kyröläinen et al. (2021) where writers are in control of their topics and word choices, the present task taps into strategies that individuals develop both for their preferred positive emotional states and for negative ones. Thus, it is possible that a comparison of valence ratings from older and younger raters will elicit a more nuanced picture, different for the negative and positive sub-ranges of valence. Analyses that tackle the possibilities presented in this study contribute to the understanding of emotions and emotion regulation over the lifespan and during the time of crises.

Method

Participants

A total of 1446 participants were recruited via the Prolific crowdsourcing website of which 554 were assigned to the older adult group and 892 to the younger adult group: after trimming (described in the Results section below), 545 participants remained in the older adult dataset and 886 remained in the younger adult dataset. Respondents were restricted to those who self-identified as being (i) of or over the age of 65 in the case of the older adult group and add age limit in-between 18 and 65 for the younger group, (ii) both born in and a current resident of the USA, UK, or Canada, (iii) a native speaker of English, and (iv) without language impairment. The mean age was 70 for the older adult subset of new data (SD 4, range = 65–89) and 34 for the younger adult subset (SD 9.5, range = 18–62). For gender, three small categories of responses were merged into a single level labeled as other, namely missing ($n = 8$), explicit no

Table 1 Distribution of participants across levels of education and gender

Variable	Gender			Total
	Female	Male	Other	
Education				
Doctorate	23 (1.6%)	28 (2.0%)	0 (0%)	51 (3.6%)
Master's degree	104 (7.3%)	123 (8.6%)	3 (0.2%)	230 (16%)
Bachelor's degree	244 (17%)	296 (21%)	7 (0.5%)	547 (38%)
Associate degree	46 (3.2%)	50 (3.5%)	4 (0.3%)	100 (7.0%)
Some college no degree	151 (11%)	143 (10.0%)	7 (0.5%)	301 (21%)
High school graduate	82 (5.7%)	79 (5.5%)	0 (0%)	161 (11%)
Some high school	17 (1.2%)	13 (0.9%)	0 (0%)	30 (2.1%)
Unknown	4 (0.3%)	4 (0.3%)	3 (0.2%)	11 (0.8%)
Total	671 (47%)	736 (51%)	24 (1.7%)	1,431 (100%)

response ($n = 4$ and other ($n = 12$). The distribution of participants across levels of education and gender is provided in Table 1.

This study has received clearance from the McMaster Research Ethics Board (protocol 3670).

Materials

We set the number of words to be included in the dataset of valence ratings to 3,600. This number was determined based on financial considerations and a relatively limited number of older (65+) participants available for recruitment via Prolific. Nevertheless, this number of stimuli is substantial and covers about one one-third of the 11,200 word families that an average North American 20-year old speaker of English is estimated to know Brysbaert et al. (2016). Also, we established that this number of distinct words provides an adequate coverage both of the range of valence observed in English words by Warriner et al. (2013) and of multiple semantic categories of potential interest, see below.

We aimed to ensure that any given word list in our study was not biased towards either extreme of valence ratings, thus removing the possibility of a positivity bias in the stimuli, rather than the responses. To this end, all word stimuli in the present study were sampled from a set of roughly 14,000 English words in Warriner et al. (2013). We excluded words that were not reported in the frequency list of the 51-million token SUBTLEX-US corpus of subtitles to US films and media Brysbaert and New (2009). The total of 3,600 words in the stimulus list were divided into 40 lists, with 90 words each. The words represented in Warriner et al. (2013) were divided into three groups representing low, mid and high valence, based on the tertiles of the valence distribution. To create a 90-word list for our study, we randomly sampled 30 words from each bin, without replacement. Each iteration of this procedure resulted in a list containing 90 words covering the whole range of valence. This procedure was repeated 40 times yielding the final stimulus set of 3,600 words.

To obtain information about semantic categories that show a difference in valence as a function of age, we made use of Linguistic Inquiry and Word Count (LIWC) curated dictionaries provided by Pennebaker and Francis (1999), which assign words into thematically and semantically coherent groups such as BODY, AFFECT OR ANGER. A given word may be attributed to one or more LIWC categories, for example the word *abdomen* is associated with the semantic categories of BIO and BODY. To maximize representation of words that contain independently established semantic information in our data, we sampled 28 lists specifically from the subset of words included in one or more semantic categories of the LIWC database. Then we sampled the remaining 12 lists from the remainder of Warriner et al.'s dataset. In total we sampled 40 lists of 90 words each, with every list representing the entire valence range, as outlined above.

Our final set included 3,600 words, of which 23.3% are primarily used as adjectives, 60% as nouns, 15.9% as verbs and 0.8% as other or unspecified parts of speech. The mean word frequency of the set was 3135 ($SD = 16156$, range: 15–314232, Mdn 270), based on SUBTLEX_US Brysbaert and New (2009). Words with LIWC semantic information account for 2552 words or 70.9% of the data covering 66 semantic categories. On average each word was assigned to 2.6 semantic categories ($SD = 1.4$, range: 1–8).

Following Warriner et al. (2013)'s procedure, we appended 10 calibrator words to the beginning of each list. These words represented both the extremes of the valence range (e.g., negative *jail* and *invader*, and positive *joke*, *free*) and the relatively neutral mid-range (e.g., *icebox*, *hat*). The purpose of the calibrators was to give participants a sense of the full range of the valence dimension that they would encounter. Thus, the resulting lists contained 100 words each. We also included two validation questions in each experimental list to ensure the learner was not typing in random numbers. The two validation questions were "Type the number 5" and "Type the number 9". These attention checks were used to filter out unreliable respondents.

When performing the task, participants always saw the randomly shuffled calibrator words first. After the calibrator words, one list of 90 words was randomly selected and combined with the two verification questions. The list was randomly shuffled and presented one item at a time to the participant.

Procedure

First, participants read and approved a consent statement by clicking the "I agree" button. Then, participants were asked to indicate their age, gender, place of birth, age, country/state resided in most between birth and age 7, country/state currently reside in, native language(s), and education level.

Afterwards, participants were presented with an instruction page. The instructions informed the participant that the purpose of the study was to investigate emotion. The instructions asked participants to "respond to different types of words, by providing a rating on a scale of 1 (unhappy) to 9 (happy) of how you felt while reading each word. If you feel completely neutral you should rate a 5". Subsequently, participants were presented the list of experimental stimuli, beginning with 10 calibrator words and 90 target words, combined and randomly shuffled with the two validation questions. A similar procedure was used in Warriner et al. (2013) ensuring the comparability of these studies.

Participants were expected to press the keys 1–9 to indicate their rating of the valence of each word. It is important to note that Warriner et al. (2013) used a flipped rating scale of 1 indicating happy and 9 indicating unhappy: our rating scale went from 1 (unhappy) to 9 (happy). In case the

word was unknown to them, they were instructed to press the letter ‘n’. We also measured the reaction time of each keyboard press. Once finished, participants clicked “Submit” to complete the study. The entire experiment took less than 20 minutes. Data collection was completed in November 2020 - March 2021.

Valence ratings across age groups

Many analyses below involved a comparison between the present dataset collected from mature adults and an existing dataset of valence ratings Warriner et al. (2013) representing the entire range of adult age. Warriner et al.’s study collected norms of valence, arousal, and dominance from 1,827 North American native speakers of English in the age range from 16 to 87. While technically this range overlaps with the age range in this study, only 20 (less than 1%) of participants in Warriner et al. (2013) fell into the age group targeted in this study (65 years of age or older). For this reason, we consider the present collection of data to be non-overlapping in terms of participant age with Warriner et al.’s (2013) dataset. We label the cohorts used in our analyses by age and data collection: the newly reported datasets “older 2021”, and “younger 2021”, and Warriner et al.’s data “younger 2013”, respectively. One implication of using Warriner et al.’s (2013) dataset is that data from the younger cohort in the new resource have a pre-pandemic counterpart for a direct comparison, but the data from the older cohort do not. We discuss this circumstance below.

We estimated and reported the difference in valence ratings between all pairs of cohorts (younger 2021, older 2021, younger 2013) for each word as a standardized effect size, using the present dataset and Warriner et al.’s (2013) data respectively. This was achieved by calculating Cohen’s *d* Cohen (1992) using means, standard deviations, and the number of responses in both the older and younger cohorts and applying a correction for smaller samples:

$$d = \frac{\mu_1 - \mu_2}{0.5 * \sqrt{s_1^2 + s_2^2}} * \frac{n - 3}{n - 2.25} * \sqrt{\frac{n - 2}{n}}. \quad (1)$$

where μ_1 and μ_2 are mean valence ratings in the older and younger cohorts; s_1 and s_2 are standard deviations of valence ratings in the older and younger cohorts; and n is the total number of ratings for the word summed over both cohorts. A positive value of Cohen’s *d* indicates a higher valence rating in the older compared to the younger cohort, and a negative value indicates that the rating in the younger cohort was higher than in the older one.

All analyses were carried out in R, version 4.0.3. R Core Team (2020). Linear mixed-effects models were fitted using the lme4 package in R, version 1.1-27.1 Bates et al. (2015).

For these models, the Satterthwaite’s degrees of freedom method Giesbrecht and Burns (1985); Fai and Cornelius (1996); Satterthwaite (1946) was used to estimate the statistical significance of a variable, implemented in the R package lmerTest, version 3.1-3 Kuznetsova et al. (2017).

Results

Descriptive statistics for the present 2021 database

The total number of responses was 156,520, including the older 2021 group (77,490) and the younger 2021 group (79,030). The number of participants was 1446, with 554 unique participants in the older 2021 group and 892 in the younger 2021 group. The average number of ratings per word for the older 2021 group was 21.46 ($SD = 1.9$, range = 13–26), while for the younger 2021 group the average was 21.95 ($SD = 2.62$, range = 12–31).

Data trimming procedures were executed as follows: We removed a total of 11 trials (e.g., sets of ratings to 90 words) that individual raters took more than once, i.e., older 2021 group (11), younger 2021 group (0). We removed a total of 7 trials where individual raters answered both verification questions incorrectly, i.e., older 2021 group (6), younger 2021 (1). We removed a total of 14 participants where the self-reported participant’s age was either not between 65 and 100 years old in older 2021 group (9) or greater than 65 years old in younger 2021 group (5). We manually cleaned 5 values of user ages that appeared to be entered incorrectly: older 2021 group (5), younger 2021 group (0). We removed a total of 2 trials in which participants had more than 15% of words with z-scores of 3 or above: z-scores were calculated based on ratings per given word, older group 2021 (2), younger group 2021 (0). This step aimed to remove participants who reversed their rating scale in their responses.

The older 2021 group finished with 861 out of 889 lists (97% of its original data pool) and 545 out of 554 (98%) unique participants. The younger 2021 group finished with 886 out of 892 (99% of its original data pool) and 886 out of 892 (99% of its original data pool) unique participants. The total number of “n” responses (indicating that the word was not known) was 935: older 2021 group (225) and younger 2021 group (710). The words with the most “word unknown” responses for the older and younger adults were ‘gad’ (12 responses) and ‘succubus’ (7 responses), respectively. All analyses below were based on numeric responses only.

Reliability of valence ratings was estimated at the word level using the split-half method. Specifically, we considered each word ($N = 3,585$) that received 16 or more numeric ratings and randomly drew two samples (without replacement), each containing 8 of these ratings. Mean ratings were

calculated for both samples for each word in the stimulus list, resulting in two sets of 3585 mean ratings. The correlation between two sets gave a point-wise estimate of split-half reliability. The distribution of these estimates obtained via 1000 iterations of random sampling indicated a very high split-half reliability of valence ratings at the word level both for the older 2021 group: $r = 0.899$, 95% CI [0.89, 0.91] and the younger 2021 group: $r = 0.871$, 95% CI [0.86, 0.88].

With age as one of the critical variables of interest, we conducted several analyses of whether the participants' age affected individual or group patterns of responses in the valence judgment task. Each of the analyses is reported in a separate section below.

Age effect on valence ratings in the present 2021 dataset

The first analysis tested the effect of age on valence ratings in the present dataset that included two complementary

cohorts, younger 2021 and older 2021. On average, the older 2021 cohort gave higher ratings to the 3,600 target words ($M=5.19$, $Mdn=5.00$, $IQR=3$) than their younger counterparts ($M=5.04$, $Mdn=5.00$, $IQR=2$). The difference was relatively small (Cohen's $d = 0.07$, 95% CI 0.06-0.08), suggesting a very substantial overlap between distributions of valence (97.2%). We provide more detailed analyses of the group differences in the next section.

To examine the effect of age on the participant level, a linear mixed-effects model was fitted to valence ratings with age as a predictor, word frequency as a control covariate, and by-item and by-participant intercepts. To avoid outliers, we only considered age values represented by at least 4 participants: this restricted the range of age to 18–81 years old. As indicated by the preliminary data inspection, the effect of age on valence rating was nonlinear and we modeled it using a quadratic polynomial. The summary information of the fitted model is provided in Table 2 and the estimated age effect is visualized in Figure 1.

Table 2 Summary information of the fitted linear mixed-effects model with the coefficients and standard error

Effect	Group	Term	Estimate	Std.error	Statistic	Df	P-value
fixed		Intercept (older 2021)	4.29	0.08	53.74	3848.09	<0.01
Fixed		age, linear	36.37	5.09	7.14	1384.67	<0.01
Fixed		age, quadratic	11.02	4.85	2.27	1417.66	0.02
Fixed		frequency (log10)	0.33	0.03	11.18	3681.08	<0.01
Random	Word	SD (Intercept)	1.43				
Random	Participant	SD (Intercept)	0.46				
Random	Residual	SD (Observation)	1.46				

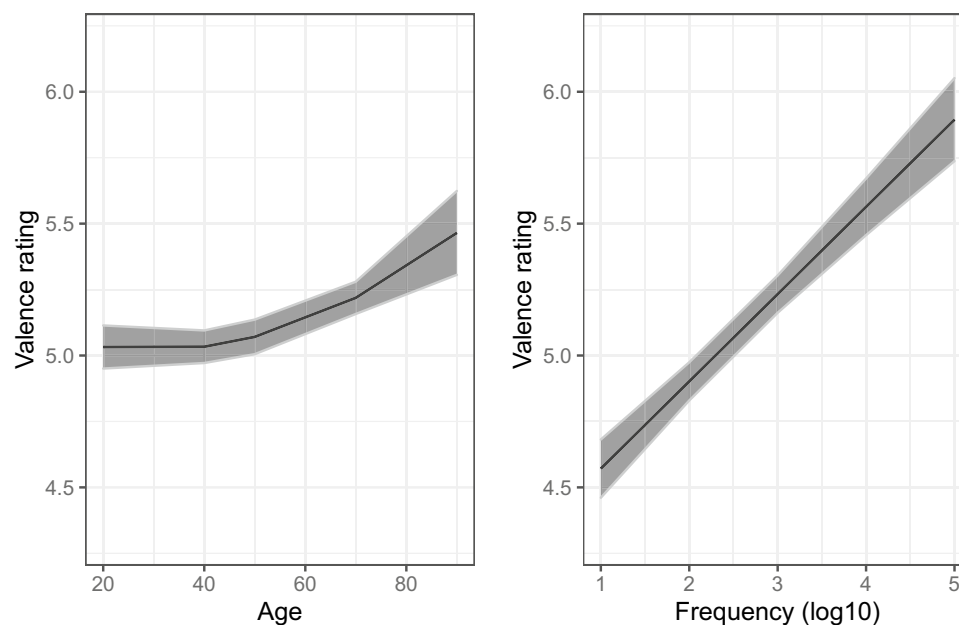


Fig. 1 Estimated effect of frequency and age on valence rating with 95% CIs

Figure 1 left panel reveals a strong nonlinear positive partial effect of age on valence ratings. Specifically, it indicates a plateau around the mid-scale of valence (5 points) in the younger 2021 cohort (<65 y.o) and a steep and continuous increase in valence ratings throughout the age span in the older cohort (65–81 y.o.). The model-estimated valence rating at 81 y.o. is 5.45, i.e., an increase of 5.6% from the estimated rating of 5.03 for the younger cohort. This increase was estimated while controlling for the effect of word frequency and the between-items and -participants variability. The effect of frequency on valence ratings was strong and positive (Figure 1 right panel) in line with a previous report by Kyröläinen and Kuperman (submitted); Warriner and Kuperman (2015). This analysis of the data collected during the pandemic demonstrates unequivocally that positivity bias increases gradually with age, with the onset of the increase found around the age of 60 y.o. This finding converges with the notion of improved emotion regulation in older adults (see the Introduction) and further suggests that the improvement is enhanced with age. The frequency effect confirms that the positivity bias exists at the level of word tokens (i.e., a more frequent use of relatively positive words).

As a next step, we expanded the comparison of the older and younger 2021 cohorts beyond the central tendency. To this end, we made use of effect sizes that estimated cross-cohort differences in valence ratings for all words in the dataset. Specifically, we calculated Cohen's d values for the older 2021 - younger 2021 difference in valence ratings: a positive/negative d value indicated a word that older adults

rated more positively/negatively. We fitted a linear regression model to Cohen's d values as a dependent variable with valence ratings aggregated over the entire 2021 dataset (younger and older cohorts) as the sole predictor. These ratings were centered around the mid-scale of the valence scale (5 points). Figure 2 visualizes the estimated effect.

The estimated slope revealed a reliable positive effect [$\hat{\beta} = 0.04$, $SE = 0.004$, $p < 0.001$]. The amount of explained variance (adjusted R^2) was 2.3%. Figure 2 demonstrates that younger participants were more moderate in their responses towards the extremes of the affective continuum, while older ones showed increasingly more polarized responses to these extremes. The more extreme the rating was, as evaluated by both younger and older participants, the greater the difference was between younger and older participants. Specifically, positive words elicited systematically higher valence ratings from older rather than younger participants, and this discrepancy increased with the average valence of the word. Similarly, negative words elicited lower valence ratings from older rather than younger participants: the more negative the word was, the stronger the discrepancy.

Age-group differences between pre-pandemic and pandemic measurements of valence

The analyses above focused on the effect of age on valence ratings collected in a relatively narrow timeframe of a few months, during the COVID-19 pandemic. Availability of comparable data from (predominantly) younger participants

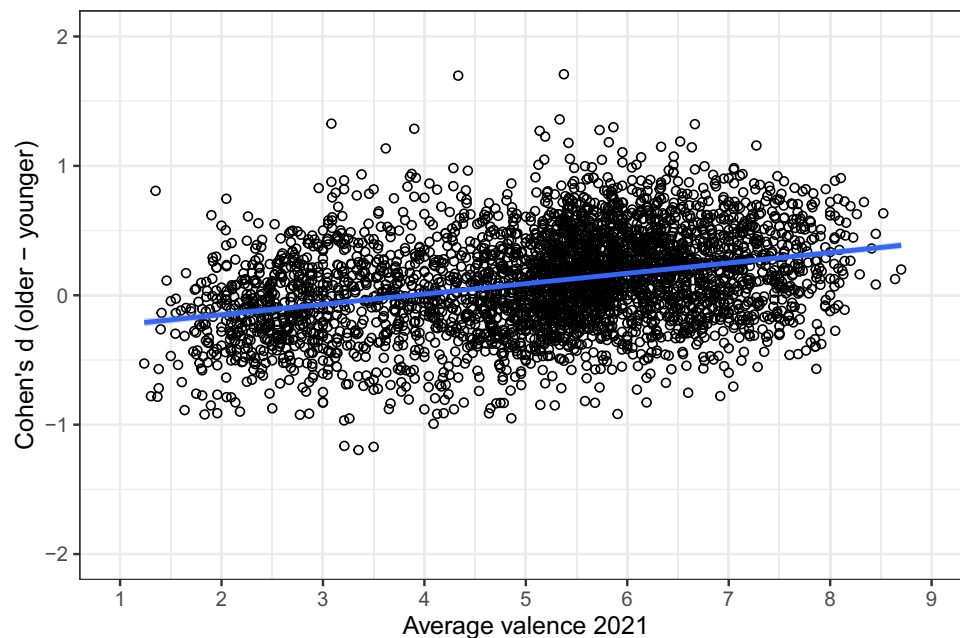


Fig. 2 Cohen's d standardized difference in valence ratings between older and younger 2021 cohorts as a function of average valence ratings of the 2021 dataset, with the estimated partial effect of valence and the 95% confidence interval

in Warriner et al. (2013) further enables us to quantify the change in valence ratings over time. As we discuss below, we attribute the lion's share of this change to the psychological fallout of the pandemic rather than the passage of roughly 8 years.

In this analysis we considered age-related differences at the group level, with three cohorts (older 2021, younger 2021 and younger 2013) serving to quantify effects of age or the pandemic or both. A linear-mixed effects model was fitted to the data where the average valence rating was modeled as a function of age group and words were included as random intercepts. This fitted model was trimmed by removing data points that exceeded the absolute value of the residuals of 2.5 resulting in a removal of 169 data points (1.56% of the data). The overall effect of age group was statistically significant (ANOVA III with Satterthwaite's method, $F(2, 7035.2) = 168.26, p < 0.001$) and the summary information of the trimmed model is reported in Table 3.

The regression model 3 estimated that, on average, older adults generated higher valence ratings (older 2021 $M = 5.20$) than both younger counterpart cohorts (younger 2021 $M = 5.03$, younger 2013 $M = 5.17$): these contrasts were statistically significant at the 1% level. The younger cohort of the present 2021 dataset was also associated with significantly lower valence ratings ($p < 0.01$) that the age-matched counterparts from Warriner et al. (2013). This pattern of results suggests the following hierarchy of positivity (from more optimistic to more pessimistic): older 2021 > younger 2013 > younger 2021. In the General Discussion, we elaborate on the implications of these findings for our goal of identifying the effects of age and the pandemic on valence ratings and, by implication, emotion regulation.

Existing literature puts emphasis on positivity bias (or the Pollyanna principle) – a preference for positive experience, expression, and evaluation – as a universal feature of human communication present across multiple languages and populations (e.g., Dodds et al. (2011, (2015); Warriner and Kuperman (2015)). To contribute to this literature, we tested whether positivity bias holds in our cohorts. A typical operationalization of the test is to determine whether the mean valence rating to a balanced

stimulus set (representative of both negative and positive words) is statistically different from the middle of valence scale, i.e., the rating of 5. Both older 2021 and younger 2013 demonstrated a strong positivity bias (one-sample t-test p -values < 0.001). However, no positivity bias was indicated for the younger 2021 cohort (one-sample t-test $t = 1.55, df = 3599, p = 0.121$). Thus, when tested during the pandemic, younger adults (<65 y.o.) – though not the older adults – failed to replicate the bias robustly documented in the literature on language and emotion.

Data availability

We make the data introduced in this study publicly available at https://osf.io/e6px8/?view_only=0bef4b18a10e4397b3dd04d7fb60559b. The data are published as six source files covering both the averaged (summary_valence_ratings) and the cohort-specific trial-level data (older_adult_trial_data and younger_adult_trial_data) and combined trial-level data (combined_trial_data) as well as the collected demographics data (older_adult_data) and (younger_adult_data). Additionally, we have provided semantic categories associated with the words as part of the data release. The semantic categories are based on Linguistic Inquiry and Word Count (LIWC) curated dictionaries provided by Pennebaker and Francis (1999), which assign words into thematically and semantically coherent groups such as BODY, AFFECT OR ANGER. A given word may be attributed to one or more LIWC categories, for example the word *abdomen* is associated with the semantic categories of BIO and BODY.

The released data files are provided as a text (.txt) and an R object (.Rds) file. The full description of the variables in each of the source files can be found in the README file along with the license information. It is worth pointing out that the reaction times associated with ratings are also provided although they were not analyzed as part of this study. A snippet of the average word-level data is shown in Table 4.

Table 3 Summary information of the linear mixed-effects model fitted to valence ratings with age group as a critical predictor

Effect	Group	Term	Estimate	Std.error	Statistic	Df	P-value
fixed		Intercept (Older 2021)	5.20	0.02	212.17	3957.41	<0.01
Fixed		Younger 2013	-0.03	0.01	-3.73	7034.82	<0.01
Fixed		Younger 2021	-0.16	0.01	-17.40	7035.33	<0.01
Random	word	SD (Intercept)	1.42				
Random	Residual	SD (Observation)	0.38				

Table 4 A snippet of the averaged data showing the five first variables that are included in the source file `summary_valence_ratings`

stimulus_word	younger_2021_rt_mean	valence_younger_2021	older_2021_rt_mean	valence_older_2021	valence_younger_2013
abdomen	1.95	4.52	2.04	4.95	5.43
abdominal	2.04	4.48	2.30	4.11	4.48
able	1.56	7.23	1.93	7.55	6.64
abnormality	2.19	3.75	2.36	3.12	3.05
abolish	1.99	4.00	3.48	4.11	3.84

General discussion

This paper addresses the affective aspect of aging and its manifestation in language perception. It does so by presenting a new dataset with valence (positivity) ratings for 3,600 commonly used English words. The ratings were elicited from the population of North American and British native speakers of English that covers the entire span of adulthood (18 y.o. and older) including an over-sampled cohort of older adults (65 y.o. or older). The new data are enriched with distributional and semantic information, as well as demographic data about the raters. These data bring into focus the overall emotional state of an older vs. younger adults and make possible the investigation of specific semantic areas where aging engenders the strongest affective change relative to younger adults.

An important characteristic of the dataset is its collection period (November 2020–March 2021), which falls within the timeframe of the world-wide COVID-19 pandemic and lockdown. The psychological fallout of these disruptions has had a demonstrably harmful effect on the emotional well-being, increasing the incidence of depression, dysphoria, anxiety, and numerous other symptoms. Thus, the new data are of use in determining the effect of aging on emotional responses to a broad variety of affective stimuli under the taxing conditions of the pandemic. Moreover, comparisons with a highly comparable pre-pandemic dataset Warriner et al. (2013) enable an estimation of the effect that the pandemic had within and across age groups.

We envision a broad use of the reported dataset to tackle a range of issues related to aging, language and affect. In this report, we demonstrate such use by adding a new perspective on one robust finding in the literature. With age, people are argued to maintain a more positive emotional tenor overall and also improve in their strategic ability to allocate more attention to positive rather than negative experiences and memories (see the Introduction). That is, older individuals appear to show superior emotion regulation compared to their younger counterparts. In experiments and mega-studies using linguistic stimuli, this tendency has emerged either in a processing or attentional bias towards positive stimuli or in the preferential use of positive words

in natural written productions (see references above). The present study probed the idea of improved emotion regulation in a different way. Namely, we presented a participant with an equal number of positive and negative word stimuli, including very extreme ones, for direct affective evaluation. This task enabled quantification of group and individual differences both in the central tendency and dispersion of responses to the broad range of emotionally loaded stimuli. Below we present a summary of how our findings from the valence rating task shed light on the analytical goals of this study.

Effect of aging

Our first goal was to pin down the effect of aging on the distribution of valence ratings and, by implication, the efficiency of emotion regulation. The results provide insight into responses elicited from mature adults when a well-regulated system is confronted both with preferred and dispreferred stimuli. A further comparison with responses to the same stimuli from a younger population augmented our understanding of the overall impact of aging on the affective side of language perception.

Our findings confirmed and refined the existing proposal of an age-related advantage in emotion regulation. First, the regression analysis of the present dataset (Figure 2) showed a continuous super-linear increase in valence ratings throughout the age range associated with older adulthood (>60 y.o.). No parallel increase or any other change was observed in the average valence rating in younger adults. This evidence – obtained with a method of affective evaluation that has rarely been used in emotion regulation research – indicates that an improvement in emotion regulation continues throughout the entire lifespan, well into mature adulthood. This notion converges well with a proposal that aging is accompanied by an efficient development and implementation of strategies of maintaining a stable and relatively optimistic emotional tenor (see the Introduction for references).

We also note that while the average affective evaluation is more positive in older than younger adults, the former cohort is also more polarized in their responses. Thus, in

comparison to younger counterparts, older adults elicit more positive/negative evaluations of the stimuli that all cohorts find extremely positive/negative. This might indicate an exaggerated response to the stimuli that threaten the relative emotional homeostasis that older adults tend to sustain. To our knowledge, this greater emotional range in responses of older versus younger participants is a new finding that future research needs to reconcile with the proposal of the age-related advantage in emotion regulation. We believe that a targeted experiment, with an over-representation of extreme-valence and extreme-arousal stimuli, is a promising way to exploring this finding further.

Effect of the psychological impact of the pandemic

Our second goal was to isolate the impact of the COVID-19 pandemic on emotion regulation Killgore et al. (2020); Shanahan et al. (2020) and also to establish whether the proposed superiority in emotion regulation among older adults is found under extraordinarily stressful living conditions of the pandemic. Several of our findings bear relevance to this goal. First, a comparison of two cohorts of younger adults (18–64 y.o. in 2013 vs 2021) revealed a substantial drop in the average valence of ratings ($M_{2013} = 5.16$ vs $M_{2021} = 5.03$), equivalent to 1.6% of the valence scale. As a result, the younger 2021 cohort did not demonstrate the positivity bias presumed to be a universal emotional characteristic of human language Dodds et al. (2011, 2015); Warriner and Kuperman (2015). We ascribe this decrease to the widely reported prevalence of pessimism, depression and other common symptoms of mood disorders reported in the medical practice, research surveys and text-analytical studies during the COVID-19 pandemic (e.g., Boon-Itt and Skunkan (2020); Lwin et al. (2020)).

There is a logical possibility that the decrease in mean valence observed in two testing sessions separated by 8 years is due to the passage of time, independent of the pandemic. If the societies from which participants were recruited (i.e., USA in 2013 and 2021, and Canada and the UK in 2021) can be shown to have grown more pessimistic over time, this would account for our finding without resorting to the pandemic as an explanans. However, several recent studies using historical text data argued for the opposite. At least prior to the COVID-19 pandemic, English-speaking societies have shown a continuous rise in optimism, as gauged from the increasing positive connotation of texts as a function of their recency Hamilton et al. (2016); Iliev et al. (2016), but see also Morin and Acerbi (2017). Thus, we attribute the observed decrease in valence ratings to the impact of the pandemic rather than a stable temporal trend.

Another intriguing finding that sheds light on both goals of the study is that older adults tested during the pandemic (older 2021, 65 y.o. or older) demonstrated a significantly

higher average value of valence than either younger cohort (younger 2013 and 2021). This suggests that even during the pandemic older adults display a more positive emotional tenor than their younger counterparts showed either before or during the pandemic. In the absence of pre-pandemic data from an older group as a comparator, we cannot directly quantify the amount or the direction of change in the emotional well-being of older adults. Yet, given reports of increasing pessimism and anxiety in North American older adults Kyröläinen and Kuperman (submitted), we can safely assume that older adults did not improve their emotional well-being during the pandemic. Thus, we infer that the advantage we observed in the older 2021 cohort over the pre-pandemic younger 2013 cohort would have been amplified were we to use the (unfortunately, unavailable) pre-pandemic data from older adults.

Taking our observations and inferences together, we conclude that older adults possess a greater ability (than younger adults) for emotion regulation. They maintain higher levels of positivity both (i) during regular daily lives. Also, while the times of crises bring increased pessimism and anxiety to individuals across the lifespan, older adults are still more positive than younger adults during these times (ii). Conclusion (i) finds rich support in prior research on aging, see discussion above. Conclusion (ii) is also corroborated by recent reports that – despite the objectively greater strain on their physical and social mobility and greater medical risks – the COVID-19 pandemic has had a more moderate impact on the older rather than younger adults. Specifically, older adults experience less acute feelings of loneliness and a lower incidence of depressive and stress-related symptoms than their younger peers García-Portilla et al. (2021); González-Sanguino et al. (2020); Klaiher et al. (2020); Luchetti et al. (2020). Thus, the valence data offer an independent confirmation of trends observed through recent surveys, while avoiding biases associated with a direct interrogation of an individual's feelings and beliefs. We believe that investigating additional dimensions of affective evaluation of or a sensorimotor response to lexical stimuli in older adults (e.g., arousal, dominance, concreteness) is a fruitful direction of future research.

To conclude, this study of older adults adds to the research effort of creating mega-studies reporting lexical and affective information for populations of specific interest for psychological research, e.g., L2 speakers, children and adolescents Imbault et al. (2020); Liu et al. (2021). We view it as a testbed for generating and testing hypotheses both independently and in comparison to other demographic groups, in order to create a comprehensive picture of how affect develops and finds expression through language over the lifespan. This study also adds to the literature concerned with individual and societal response to catastrophic events, including armed conflicts, terrorist acts,

natural disasters, and pandemics Updegraff et al. (2008); Cheung-Blunden and Blunden (2008); Cohn et al. (2004); Kontoangelos et al. (2020).

Author contribution Contributions of all authors were partly supported by the Social Sciences and Humanities Research Council of Canada Partnered Research Training Grant 895-2016-1008 (Libben, PI). Contribution by VK was further partially supported by the Canada Research Chair award (Tier 2; Kuperman, PI), and the CFI Leaders Opportunity Fund (Kuperman, PI). Furthermore, this work was supported by funding from the Labarge Centre for Mobility in Aging within the McMaster Institute for Research on Aging at McMaster University (Kuperman, PI). Finally, contribution by AJK was partially supported by the Social Sciences and Humanities Research Council of Canada Insight Development Grant, 430-2019-00851, (Kyröläinen, PI) and AGE-WELL Graduate Student and Postdoctoral Award In Technology and Aging.

References

- Bates, D., Mächler, M., Bolker, B., and Walker, S. Fitting linear mixed-effects models using lme4. 1–48.
- Boon-Itt, S., and Skunkan, Y. Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study. *JMIR Public Health and Surveillance* 6, 4 (2020).
- Boyd, R. L., and Schwartz, H. A. Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology* 40, 1 (2021), 21–41.
- Brysbaert, M., and New, B. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. 977–990.
- Brysbaert, M., Stevens, M., Mandera, P., and Keuleers, E. How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. 1116.
- Cacioppo, J. T., Hawkey, L. C., Kalil, A., Hughes, M. E., Waite, L., and Thisted, R. A. Happiness and the invisible threads of social connection: The Chicago Health, Aging, and Social Relations Study. In *The Science of Subjective Well-Being*, M. Eid and R. J. Larsen, Eds. Guilford Press, 2008, p. 195–219.
- Carstensen, L. L., and DeLiema, M. The positivity effect: A negativity bias in youth fades with age. 7–12.
- Carstensen, L. L., Mikels, J. A., and Mather, M. Aging and the intersection of cognition, motivation, and emotion. In *Handbook of the Psychology of Aging*, J. E. Birren, K. W. Schaie, R. P. Ables, M. Gatz, and T. A. Salthouse, Eds., 6th ed. Elsevier, 2006, p. 343–362.
- Charles, S. T., Reynolds, C. A., and Gatz, M. Age-related differences and change in positive and negative affect over 23 years. 136–151.
- Cheung-Blunden, V., and Blunden, B. The emotional construal of war: Anger, fear, and other negative emotions. *Peace and Conflict* 14, 2 (2008), 123–150.
- Cohen, J. A power primer. *Psychological Bulletin* 112, 1 (1992), 155–159.
- Cohn, M. A., Mehl, M. R., and Pennebaker, J. W. Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science* 15, 10 (2004), 687–693.
- Cullen, W., Gulati, G., and Kelly, B. D. Mental health in the COVID-19 pandemic. *QJM: An International Journal of Medicine* 113, 5 (2020), 311–312.
- Davitz, J. R. *The language of emotion*. Academic Press, 1969.
- Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., Mitchell, L., Harris, K. D., Kloumann, I. M., Bagrow, J. P., Megerdooian, K., McMahon, M. T., Tivnan, B. F., and Danforth, C. M. Human language reveals a universal positivity bias. In *Proceedings of the National Academy of Sciences* (2015), vol. 112, National Acad Sciences, p. 2389–2394.
- Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., and Danforth, C. M. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE* 6, 12 (2011), e26752.
- Fai, A. H.-T., and Cornelius, P. L. Approximate *F*-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. 363–378.
- García-Portilla, P., de la Fuente Tomás, L., Bobes-Bascarán, T., Jiménez Treviño, L., Zurrón Madera, P., Suárez Álvarez, M., Menéndez Miranda, I., García Álvarez, L., Sáiz Martínez, P. A., and Bobes, J. Are older adults also at higher psychological risk from COVID-19? *Aging & Mental Health* 25, 7 (2021), 1297–1304.
- Giesbrecht, F. G., and Burns, J. C. Two-stage analysis based on a mixed model: Large-sample asymptotic theory and small-sample simulation results. 477–486.
- González-Sanguino, C., Ausín, B., Castellanos, M. n., Saiz, J., López-Gómez, A., Ugidos, C., and Muñoz, M. Mental health consequences during the initial stage of the 2020 coronavirus pandemic (covid-19) in Spain. *Brain, behavior, and immunity* 87 (2020), 172–176.
- Gross, J. J. The emerging field of emotion regulation: An integrative review. 271–299.
- Hamilton, W. L., Clark, K., Leskovec, J., and Jurafsky, D. Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing* (2016), vol. 2016, NIH Public Access, p. 595.
- Iliev, R., Hoover, J., Dehghani, M., and Axelrod, R. Linguistic positivity in historical texts reflects dynamic environmental and psychological factors. *Proceedings of the National Academy of Sciences* 113, 49 (2016), E7871–E7879.
- Imbault, C., Titone, D., Warriner, A. B., and Kuperman, V. How are words felt in a second language: Norms for 2,628 English words for valence and arousal by L2 speakers. *Bilingualism: Language and Cognition* (2020), 1–12.
- Killgore, W. D. S., Taylor, E. C., Cloonan, S. A., and Dailey, N. S. Psychological resilience during the COVID-19 lockdown. *Psychiatry Research* 291 (2020).
- Klaiber, P., Wen, J. H., DeLongis, A., and Sin, N. L. The Ups and Downs of Daily Life During COVID-19: Age Differences in Affect, Stress, and Positive Events. *The Journals of Gerontology: Series B* 76, 2 (07 2020), e30–e37.
- Kontoangelos, K., Economou, M., and Papageorgiou, C. Mental health effects of COVID-19 pandemic: A review of clinical and psychological traits. *Psychiatry Investigation* 17, 6 (2020), 491.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. 1–26.
- Kyröläinen, A.-J., Keuleers, E., Mandera, P., Brysbaert, M., and Kuperman, V. Affect across adulthood: Evidence from English, Dutch and Spanish. 792–812.
- Kyröläinen, A.-J., and Kuperman, V. Temporal dynamics of emotional well-being and loneliness in older adults during the first year of the COVID-19 pandemic: Insights from the Cognitive and Social Well-Being (CoSoWELL) corpus. *Emotion* (submitted).
- Lindquist, K. A. The role of language in emotion: Existing evidence and future directions. 135–139.
- Liu, P., Lu, Q., Zhang, Z., and Han, B. Age-related differences in affective norms for Chinese words (AANC). 927.

- Luchetti, M., Lee, J. H., Aschwanden, D., Sesker, A., Strickhouser, J. E., Terracciano, A., Sutiu, A. R., et al. The trajectory of loneliness in response to COVID-19.
- Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., and Yang, Y. Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. *JMIR Public Health and Surveillance* 6, 2 (2020).
- Majid, A. Current emotion research in the language sciences. 432–443.
- Mather, M., and Carstensen, L. L. Aging and motivated cognition: The positivity effect in attention and memory. 496–502.
- Morin, O., and Acerbi, A. Birth of the cool: A two-centuries decline in emotional expression in Anglophone fiction. *Cognition and Emotion* 31, 8 (2017), 1663–1675.
- Panksepp, J., and Miller, A. Emotions and the aging brain: Regrets and remedies. In *Handbook of Emotion, Adult Development, and Aging*, S. H. McFadden and C. Magai, Eds. Elsevier, 1996, p. 3–26.
- Pennebaker, J. W., and Francis, M. E. *Linguistic Inquiry and Word Count (LIWC): A computer-based text analysis program*. Erlbaum Publishers, 1999.
- Pfefferbaum, B., and North, C. S. Mental health and the COVID-19 pandemic. *New England Journal of Medicine* 383, 6 (2020), 510–512.
- R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, 2020.
- Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., and Baayen, R. H. The myth of cognitive decline: Non-linear dynamics of lifelong learning. 5–42.
- Ramscar, M., Sun, C. C., Hendrix, P., and Baayen, R. H. The Mismeasurement of Mind: Life-Span Changes in Paired-Associate-Learning Scores Reflect the “Cost” of Learning, Not Cognitive Decline. 1171–1179.
- Reed, A. E., Chan, L., and Mikels, J. A. Meta-analysis of the age-related positivity effect: Age differences in preferences for positive over negative information. 1–15.
- Satterthwaite, F. E. An approximate distribution of estimates of variance components. 110–114.
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., and Ungar, L. H. Personality, gender, and age in the language of social media: The open-vocabulary approach.
- Shanahan, L., Steinhoff, A., Bechtiger, L., Murray, A. L., Nivette, A., Hepp, U., Ribeaud, D., and Eisner, M. Emotional Distress in Young Adults during the COVID-19 Pandemic: Evidence of Risk and Resilience from a Longitudinal Cohort Study. *Psychological Medicine* (2020), 1–10.
- Stawski, R. S., Sliwinski, M. J., Almeida, D. M., and Smyth, J. M. Reported exposure and emotional reactivity to daily stressors: The roles of adult age and global perceived stress. 52–61.
- Updegraff, J. A., Silver, R. C., and Holman, E. A. Searching for and finding meaning in collective trauma: Results from a National Longitudinal Study of the 9/11 Terrorist Attacks. *Journal of Personality and Social Psychology* 95, 3 (2008), 709.
- Urry, H. L., and Gross, J. J. Emotion regulation in older age. 352–357.
- Wahl, H.-W., and Lang, F. R. Aging in context across the adult life course: Integrating physical and social environmental research perspectives. 1–33.
- Warriner, A. B., and Kuperman, V. Affective biases in English are bidimensional. 1147–1167.
- Warriner, A. B., Kuperman, V., and Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. 1191–1207.