



OPEN The online language of work-personal conflict

Gloria Liou^{1,2}, Juhi Mittal³, Neil K. R. Sehgal³, Louis Tay¹, Lyle Ungar³ & Sharath Chandra Guntuku³

With the blurring of boundaries in this digital age, there is increasing concern around work-personal conflict. Assessing and tracking work-personal conflict is critical as it not only affects individual workers but is also a vital measure among broader well-being and economic indices. This inductive study examines the extent to which work-personal conflict corresponds to individuals' language use on social media. We apply an open-vocabulary analysis to the posts of 2810 Facebook users who also completed a survey for an established work-personal conflict scale. It was found that the language-based model can predict personal-to-work conflict ($r=0.23$) and work-to-personal conflict ($r=0.15$) and provide important insights into such conflicts. Specifically, we found that high personal-to-work conflict was associated with netspeak and swearing, while low personal-to-work conflict was associated with language about work and positivity. We found that high work-to-personal conflict was associated with negative emotion and negative tone, while low work-to-personal conflict was associated with positive emotion and language about birthdays.

With the technology and communication tools available today, the boundary between work and life outside of work has become increasingly blurred, exacerbating issues around work-personal conflict. A recent study found that 55% of workers send digital communication (i.e., texts, calls, emails) to colleagues after working hours, and 30% send digital communication on weekends—while expecting a same-day response¹. Work-nonwork interference occurs in the other direction as well. More than 60% of employees visit non-work-related websites at work, with more than half admitting to constantly surfing the Internet on the job². Additionally, 60% of online purchases and 70% of Internet porn traffic occur during the 9-to-5 workday^{2,3}.

While many studies have looked at work-nonwork conflict, the vast majority of these have considered family as the only or primary part of life outside of work^{4–9}. As we continue to examine work-nonwork conflict, researchers need to consider the personal domain (e.g., friends, hobbies, travel). This is an area that has not been given enough attention in the field, especially considering the increasing number of single and/or childless workers^{10–12}, whose focus outside of work is likely on personal pursuits.

In this digital age, many work and personal pursuits are shared online. There is a growing repository of information on the lives of workers on social media, and analyzing social media language data is an inexpensive, non-intrusive, and naturalistic method¹³. This study draws on organic social media data to examine work-personal conflict in 2810 participants who completed a survey for an established work-personal conflict scale⁹ and volunteered to share their Facebook posts for language analysis. Answering the call for more inductive approaches in organizational research¹⁴, we were able to create a language model that can predict work-personal conflict based on social media language and gather detailed insights about work-personal conflict and language use.

Open-vocabulary approaches to language analysis of social media data

In the past, language analysis used in organizational and psychological studies typically relied on closed-vocabulary methods¹⁵, which depended on a predefined list of words and phrases. Open-vocabulary language analysis, on the other hand, does not rely on predetermined information but instead employs computational linguistic techniques to produce a collection of language features by extracting single words, multiword phrases, and *topics*—clusters of semantically related words—as well as their relative usages. Because open-vocabulary methods do not rely on a predefined list of words and phrases, they can account for unexpected words and phrases^{16,17} and unconventional language, such as emoticons, slang, and new or uncommon words^{13,18}.

Open-vocabulary approaches to language analysis are becoming more popular in many disciplines, and the organizational field is no exception. There have been a variety of studies using open-vocabulary language analysis on social media data to investigate workplace constructs, including job satisfaction¹⁹, employee engagement^{20,21}, workplace culture²², reputation²³, and affect^{24,25}. These studies have shown that open-vocabulary social media

¹Purdue University, West Lafayette, USA. ²Rose-Hulman Institute of Technology, Terre Haute, USA. ³University of Pennsylvania, Philadelphia, USA. ✉email: gloriarliou@gmail.com

methods can provide insights that are challenging to assess with survey data or closed-vocabulary methods. For example, Saha et al. investigated the relationship between language use and two facets of job satisfaction—pay and supervision—by analyzing a Twitter dataset of 1.5 million posts¹⁹. The study found that women expressed greater pay satisfaction than men, and racial minorities expressed greater pay and supervision satisfaction than racial majorities, providing evidence for the “job satisfaction paradox”^{26–28}, where disadvantaged groups feel more satisfied than privileged groups. Saha et al. also uncovered linguistic differences between racial groups, finding that racial minorities talked more about basic livelihood while racial majorities talked more about self-actualization. These organic insights about job satisfaction showcase the benefit of using open-vocabulary social media language methods in organizational research.

Work-personal conflict

There is tremendous value in considering how the work and personal domains interact. The personal domain exhibits different attributes than the work and family domains, including higher levels of choice and lower costs for noncompliance, so examining the personal domain can enable a more thorough understanding of workers’ interrole relationships and conflicts^{9,29,30}. The emphasis on choice in the personal domain also makes work-personal conflict an excellent candidate for studying individual-level outcomes (e.g., satisfaction with the need for relatedness). Additionally, there is evidence from previous research that work-personal conflict can help explain the variance in work outcomes (e.g., job performance), outcomes at home (e.g., relationship satisfaction), and health outcomes (e.g., depression, anxiety) that cannot be explained by work-family conflict alone⁹.

Social media data is a sensible option for investigating work-personal conflict. In addition to previous successes with organizational topics^{19–22,31}, social media is largely used for personal purposes, with finding and/or maintaining friends as the primary reason cited for using social media³². Research also shows that some individuals use social media to write about friendships, health, and political views^{33,34}, all of which fall under the personal domain. While both work-to-personal and personal-to-work conflict involve the same two domains, there is evidence from previous work that social media data may not predict them equally well. Specifically, Wilson and Baumann found that personal-to-work conflict was related to satisfaction with the need for relatedness, while work-to-personal conflict was not⁹. Given that social media is used by many to satisfy the need for relatedness^{32,35,36}, there may be differences in whether and how personal-to-work conflict and work-to-personal conflict manifest in social media language.

The present study

The present study utilizes an established work-personal-family scale⁹ that was created for researchers to examine how interrole conflict impacts workers. In this study, we examined whether work-personal conflict can be predicted by language use in Facebook posts. This was accomplished by building and validating a language model that can predict work-personal conflict based on what individuals post on Facebook. Through our analysis, we were also able to identify the topics and topic themes that were prevalent among high conflict and low conflict individuals, providing insights into the impact of work-personal conflict on our daily lives.

Through this study, we aim to fill the following research gaps. First, we investigate whether work-personal conflict can be predicted by online language similar to other individual difference constructs, which is something that has not been studied before. Previous research has shown that individuals’ self-reports of personality ($r = 0.21–0.42$), life satisfaction ($r = 0.14–0.19$), and stress ($r = 0.12–0.21$) can be predicted by social media language^{13,16,37,38}, and we expect that work-personal conflict can be similarly predicted. [We searched for individual difference work constructs such as job satisfaction but could not find studies that analyzed predictive accuracy using individuals’ self-reports and Pearson’s r correlation. Saha et al. used Area Under the Curve and Hickman et al. used Mean Cross-Validation and was not investigating predictive accuracy of self-reports^{19,39}.] At the same time, the strength of this effect may be moderated by the direction of the conflict, with personal-to-work conflict being more predictable than work-to-personal conflict. Second, there are no previous studies that have examined the differences in social media language use between individuals with high and low work-personal conflict.

Our study adds to the important conversation around the conflicts faced by workers today. More broadly, we believe that assessing and tracking work-personal conflict is particularly compelling as similar work constructs have been shown to impact not only individuals’ experiences but also broader sets of well-being and economic indices^{39,40}. Establishing the ability to index and track work-personal conflict through social media data will enable organizational researchers to contribute to related policy and societal discussions.

Method

Participants

We recruited participants using Qualtrics, and participants received a small incentive for their participation. All participants provided their written informed consent and agreed to the anonymous use of their survey responses and Facebook posts for research purposes. This study received approval from the Institutional Review Board of Purdue University, and all methods were performed in accordance with relevant guidelines and regulations.

Our analytic sample ($N = 2810$) was a subset of the participants who completed the Qualtrics survey. All participants in the analytic sample were of legal age (18+), identified as man or woman, currently lived in the United States, were currently employed, and had a Facebook account. We limited our analytic sample to users who wrote at least 500 words across their Facebook status posts in English, removing three participants who only had non-English posts. We also removed the 700,000 non-English posts identified by *langid*, an off-the-shelf language identification tool⁴¹, from the original total of 4 million posts, leaving 3.3 million posts for analysis. Participants

in our study wrote an average of 27,401 words across these posts (median = 11,901; $SD = 52,568$). In our analytic sample, the mean participant age was 43.42 (median = 43; $SD = 12$), and over half (69.62%) were women.

Work-personal conflict measure

All participants completed measures of work-personal-family conflict as defined by Wilson and Baumann's work-personal-family scale⁹. The scale consists of 20 questions and measures four directions of conflict: work interfering with personal (work-to-personal; e.g., "The demands of my work interfere with my personal activities."), personal interfering with work (personal-to-work; e.g., "My personal activities produce stress that makes it difficult to concentrate at work."), family interfering with personal (family-to-personal; e.g., "The amount of time my family takes up makes it difficult to fulfill personal interests."), and personal interfering with family (personal-to-family; e.g., "My personal interests prevent me from completing family responsibilities."). For our study, we focused on two subscales, work-to-personal ($\alpha = 0.89$) and personal-to-work ($\alpha = 0.90$).

We used confirmatory factor analysis (CFA) to test the fit of the measurement model. The model provided a good fit to the data for both work-to-personal and personal-to-work subscales. For work-to-personal conflict: $\chi^2 = 91.90$, $df = 5$, $p < 0.001$; CFI = 0.99, TLI = 0.98, RMSEA = 0.08, SRMR = 0.02. For personal-to-work conflict: $\chi^2 = 29.99$, $df = 5$, $p < 0.001$; CFI = 1.00, TLI = 0.99, RMSEA = 0.04, SRMR = 0.01.

Linguistic analyses

Tokenization

We split Facebook posts into words, punctuation, and emoticons using the *happierfuntokenizing* tokenizer⁴². Words that were used by less than 1% of users were not included in the analysis as a way to remove outliers and to ensure that language markers identified in our analysis generalize to out-of-sample instances^{43,44}.

Open-vocabulary

We used the MALLET implementation of latent Dirichlet allocation (LDA)⁴⁵ to identify latent data-driven word clusters (topics). LDA has been found to offer superior predictive power relative to closed-vocabulary methods like General Inquirer (GI) and Linguistic Inquiry and Word Count (LIWC)⁴⁶. The topics are open-source¹⁶ and are generated on a corpus of about 18 million Facebook posts with alpha adjusted to 0.30 to favor fewer topics per document. We represent each user in terms of their probability of mentioning each of the 2000 topics as $(p(\text{topic}, \text{user}))$, which is derived from their probability of mentioning a word $(p(\text{word}|\text{user}))$ and the probability of the words being in given topics $(p(\text{topic}|\text{word}))$.

Inherently, each topic is realized as a set of words with probabilities. Every individual is thus scored based on their likelihood of mentioning each of the 2000 topics $(p(\text{topic}, \text{user}))$, which is derived from the probability of the individual mentioning a word $(p(\text{word}|\text{user}))$ and the probability of the word appearing in a given topic $(p(\text{topic}|\text{word}))$.

Linguistic inquiry and word count (LIWC)

For robustness, we also compared our results from LDA with the 2022 version of LIWC (LIWC-22), a closed-vocabulary method⁴⁷. LIWC offers similar extraction of linguistic features as MALLET but instead consists of 102 theory-based, manually-curated categories, and is one of the most popular closed-vocabulary methods within psychology^{46,47}. Because our focus is on the open-vocabulary approach, we prioritize discussing and interpreting these results but do present LIWC corresponding findings as well.

Statistical analysis

Each user's language represented as a dimensional vector was used as the input, and the degree of work-to-personal or personal-to-work conflict was used as the output in an ordinary least squares regression. The degree of work-to-personal or personal-to-work conflict was computed as the mean score of the five questions of each subscale. We also added demographic variables (age, gender, education, and personal income) as covariates to control for their influence on users' language. We utilized Benjamini–Hochberg⁴⁸ p-correction to correct for multiple hypothesis testing (false discovery rate) and used $p < 0.05$ to indicate meaningful correlations.

Predictive modeling

We evaluated the feasibility of predicting the work-to-personal conflict and personal-to-work conflict dimensions based on the social media language and demographic conditions. We used topics and demographics as features, which were treated as independent variables in a machine learning algorithm (ridge regression) to predict the dependent variable (i.e., work-to-personal and personal-to-work conflict). We used ridge regression to validate the model in a five-fold cross-validation setting to avoid overfitting^{46,49}. In each case, hyperparameter selection was performed in a cross-validation setting. The prediction performances are reported as Pearson correlation and Mean Absolute Error for the outcomes.

Data availability and transparency

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request. As this collected data was part of a broader collaboration, one part of the data has been used for another journal publication. However, there are no overlaps in the scales used in this manuscript and the other publication. The other publication examines loneliness and depression, whereas we examine work-personal conflict here. We detail this in Appendix D.

Results

Predictive accuracy

We found that personal-to-work conflict ($r = 0.23$) was more predictable from Facebook posts than work-to-personal conflict ($r = 0.15$) when using topics. Combining topics with age, gender, education, and personal income yielded slightly higher predictions for personal-to-work conflict (Table 1).

Linguistic analysis

In addition to predictability, we investigated themes that emerged for predicting work-personal conflict by running correlations between the 2000 topics, work-to-personal subscale score means, and personal-to-work subscale score means, controlling for age, gender, education, and personal income. We defined a theme as seven or more significant ($p < 0.05$) topics that were similar in content. There were three themes for predicting high personal-to-work conflict, two themes for predicting low personal-to-work conflict, and one theme for predicting low work-to-personal conflict. There were no themes for predicting high work-to-personal conflict. We compared these themes to the significant LIWC categories for predicting work-personal conflict. See Appendix B for all significant LIWC categories. We also compared vocabulary sizes between the high and low conflict individuals (Appendix C).

High personal-to-work conflict topics

Netspeak (“u,” “lol,” “wen,” “coz”), swearing (“shit,” “fuck,” “bitch,” “asshole”), and troubled language (“pain,” “losing,” “sanity,” “bothered”) predicted high personal-to-work conflict. For the netspeak theme, we included topics where at least three of the fifteen words in the topic were netspeak; there were 37 netspeak topics. For the swearing theme, we included topics that had at least three swear words; there were 9 such topics. For troubled language, we included topics related to pain and loss; there were 7 such topics. There was one topic (Topic 1573) that qualified as both a netspeak and a swearing topic, and one topic (Topic 1353) that qualified as both a netspeak and a troubled language topic. See Table 2 for illustrative topics for each theme. For LIWC, we found that the Netspeak and Swearing categories were also significant for predicting high personal-to-work conflict. LIWC does not have an analogous category to troubled language.

Notably, there were no low personal-to-work conflict topics with netspeak and swearing, suggesting that netspeak and swearing are reliable indicators of high personal-to-work conflict.

	Personal-to-work mean	Work-to-personal mean
Topics only	0.23	0.15
Age + gender only	0.21	0.15
All demographics	0.20	0.16
Topics + all demographics	0.24	0.14

Table 1. Predictability of personal-to-work and work-to-personal conflict. All demographics includes age, gender, education, and personal income.

Theme	Illustrative topics	Illustrative posts	Effect size (95% CI)
Netspeak	ur, dont, cuz, wen, wat, ppl, kno, urself, wats, urs, isnt, tht, bc, rite, tru	“When ur up ur up n when ur down ur down ur down :/”	0.10 (0.06, 0.13)
	bt, dat, ur, wen, nt, lyf, lyk, luv, dnt, dis, cn, wid, sum, dey, hav	“Lmmfao right nt all bt most”	0.09 (0.06, 0.13)
	ur, urself, u’ll, coz, u’ve, cos, urs, bcoz, wht, givin	“U’ve been warned!”	0.09 (0.05, 0.12)
Swearing and derogatory language	shit, fuck, fucked, fucking, damn, ass, fuckin, bullshit, bitch, shitty, pissed, shits, dude, bitches, fucks	“who’s fucked....im fucked”	0.10 (0.07, 0.14)
	shit, damn, ass, fuk, hell, wtf, fuck, dam, fuckin, bitch, pissed, aint, hell, dumb, piss	“What in the actual fuk”	0.10 (0.06, 0.14)
	shit, ass, bitches, fuck, bitch, niggas, nigga, hoes, aint, yo, hoe, dumb, fake, fuckin, yall	“hoes a be hoes. #morals”	0.09 (0.05, 0.13)
Troubled	mind, lose, losing, lost, control, sight, loose, loosing, sanity, grip, temper, ability, wanting, interest, slowly	“is losing ittt”	0.10 (0.06, 0.13)
	dont, wanna, talk, leave, anymore, cuz, mess, bothered, cus, ughh, =’{	“Dont wanna leave”	0.09 (0.05, 0.12)
	inside, deep, feel, heart, pain, hide, eyes, soul, empty, broken, pride, lies, trapped, skin, lie	“Trapped Inside My Head”	0.08 (0.04, 0.11)

Table 2. Illustrative high personal-to-work conflict topics. Topics were categorized into themes based on qualification criteria. These illustrative topics are significant ($p < 0.05$) after Benjamini–Hochberg p-correction. The complete list of significant topics can be found in Appendix A.

Low personal-to-work conflict topics

Language about work (“work,” “job,” “overtime,” “paycheck”) and positivity (“amazing,” “fantastic,” “excited,” “woohoo”) predicted low personal-to-work conflict. For the work theme, we included topics that contained the word “work” or “job”; there were 13 work topics. The positivity theme included both topics with positive adjectives (“amazing,” “fantastic,” “wonderful”) and words conveying excitement (“excited,” “woohoo,” “yay”). Importantly, we did not count topics that had any negative words even if they were largely positive otherwise. For example, Topic 522, which includes both “excited” and “dreading,” and Topic 793, which includes both “yay” and “bleh,” were not considered positivity topics. There were 14 positivity topics. There were three topics (Topics 420, 1831, and 1862) that qualified as both a work topic and a positivity topic. See Table 3 for illustrative topics for each theme. For LIWC, we found that the Work category was not significant for predicting low work-to-personal conflict. Positive Emotion and Positive Tone, the two categories analogous to our positivity theme, were also not significant for predicting low work-to-personal conflict.

High work-to-personal conflict topics

There were two significant ($p < 0.05$) high work-to-personal conflict topics. The first topic included the following words: “feels,” “weird,” “kinda,” “feel,” “feeling,” “wierd,” “bit,” “felt,” “strange,” “odd,” “hmm,” “suddenly,” “sort,” “awkward,” “dunno.” The second topic included the following words: “kinda,” “sad,” “sorta,” “sucks,” “bummed,” “feelin,” “sucked,” “lame,” “upset,” “depressing,” “pissed,” “tho,” “scary,” “disappointed,” “depressed.” Because we defined a theme as seven or more topics that were similar in content, we did not count this as a theme. However, when conducting LIWC, we did find two categories—Negative Emotion and Negative Tone—that were similar to these two topics and were significant for predicting high work-to-personal conflict.

Low work-to-personal conflict topics

Language about birthdays (“birthday,” “wishes,” “happy,” “celebrate”) emerged as the single theme for predicting low work-to-personal conflict. See Table 4 for illustrative topics for the birthday theme. LIWC does not have an analogous category to birthdays. For LIWC, the Positive Emotion category was significant for predicting low work-to-personal conflict.

Discussion

In this study, we built a language model that can predict work-personal conflict levels and uncovered important insights about work-personal conflict using Facebook status posts. The level of prediction ($r = 0.15$ – 0.23) is similar to past research examining the use of social media language in predicting individual difference constructs, such as personality ($r = 0.21$ – 0.42), life satisfaction ($r = 0.14$ – 0.19), and stress ($r = 0.12$ – 0.21). The ability to predict work-personal conflict is pivotal, as tracking work-personal conflict through social media enables organizations and communities to understand their workers’ collective experiences over time and contribute to broader well-being and economic indices.

Theme	Illustrative topics	Illustrative posts	Effect size (95% CI)
Work	Today, work, tomorrow, lunch, yay, meeting, noon, productive, payday, picking, client	“Meeting then lunch”	–0.09 (–0.05, –0.13)
	Work, hours, working, week, worked, overtime, paycheck, paid, extra, schedule, exhausted, atleast, ot, shifts, volunteer	“is working overtime”	–0.08 (–0.05, –0.12)
Positivity	Amazing, fantastic, absolutely, fabulous, brilliant, gorgeous, towers, excellent, spectacular, simply, amazingly, absolutely, incredible, wonderfully, amazin	“Brilliant”	–0.09 (–0.06, –0.13)
	Amazing, boyfriend, wonderful, absolutely, girlfriend, simply, lucky, incredible, gorgeous, fiance, fantastic, amazingly, amazin, incredibly, wonderfully	“Absolutely amazing”	–0.08 (–0.05, –0.12)
	Night, amazing, awesome, pretty, epic, wow, fantastic, gig, incredible, ended, fireworks, twas, brilliant, spectacular, wicked	“Awesome”	–0.08 (–0.04, –0.11)

Table 3. Illustrative low personal-to-work conflict topics. Topics were categorized into themes based on qualification criteria. These illustrative topics are significant ($p < 0.05$) after Benjamini–Hochberg p-correction. The complete list of significant topics can be found in Appendix A.

Theme	Illustrative topics	Illustrative posts	Effect size (95% CI)
Birthdays	Happy, birthday, wishing, sister, years, wonderful, st, daughter, nephew, brother, son, turns, niece, special, celebrate	“Happy Birthday Brother”	–0.09 (–0.05, –0.12)
	Party, birthday, bday, b-day, surprise, celebrate, st, saturday, celebrating, bash, cake, blast, celebration, graduation, planning	“Party party party(:”	–0.08 (–0.04, –0.12)
	Birthday, happy, wishes, b-day, celebrate, celebrating, birthdays, present, celebration, cake, celebrated, mom’s, belated, dad’s, brother’s	“Celebrating Marge’s B-Day!”	–0.08 (–0.04, –0.12)
	Birthday, wishes, happy, present, wished, cheers, birthdays, celebration, belated	“Thanks for the birthday wishes”	–0.07 (–0.04, –0.11)

Table 4. Illustrative low work-to-personal conflict topics. Topics were categorized into themes based on qualification criteria. These illustrative topics are significant ($p < 0.05$) after Benjamini–Hochberg p-correction. The complete list of significant topics can be found in Appendix A.

In terms of which component was more predictable from social media language, we found that personal-to-work conflict was more predictable than work-to-personal conflict. The higher predictability of personal-to-work conflict was interesting, though not entirely surprising, given that personal-to-work conflict is related to satisfaction with the need for relatedness⁹, a primary reason cited for social media use, while work-to-personal conflict is not. Additionally, Facebook is considered to be a “personal” social media platform where connections and posts shared are more centered around an individual’s personal life compared to more work-focused platforms (e.g., LinkedIn)^{50,51}.

Critically, the use of social media language analysis enabled deeper insights into the phenomenology of work-personal conflict as we could identify key topics (via LDA) and categories (via LIWC) that are discussed online by individuals. We found that high personal-to-work and high work-to-personal conflict were associated with negative language—swearing and troubled language topic themes for the former, and Negative Emotion and Negative Tone LIWC categories for the latter. On the other hand, we found that low personal-to-work and low work-to-personal conflict were associated with positive language—the positivity topic theme for the former, and the Positive Emotion LIWC category for the latter.

The relationship between negative language and high conflict—and between positive language and low conflict—is in line with past research on emotions and conflict. Previous studies have found that negative emotions are associated with work-family conflict, lower job satisfaction, lower marital satisfaction, and higher stress^{52–54}. Positive emotions, on the other hand, are associated with better work relationships, better work outcomes, and lower stress^{55–57}.

Work topics as an indicator of low personal-to-work conflict may be explained by life satisfaction research. Studies show that individuals are generally only able to focus on some (not all) domains, which leads to de-prioritizing the other domains and varying degrees of satisfaction between different domains^{58,59}. It is possible that individuals who post about work invest a lot of energy in that domain, leading to less focus on the personal domain and, subsequently, less conflict from the personal domain to other domains. Although the Work LIWC category was not significant ($p = 0.15$) for predicting low personal-to-work conflict, there was strong evidence from the open-vocabulary approach, which was the focus of our study.

Additionally, birthdays as the single topic theme for low work-to-personal conflict may be explained with similar logic. One explanation for the birthday theme is that birthdays are considered personal and family-oriented and not work-oriented. Given that individuals are generally only able to focus on some domains while de-prioritizing others^{58,59}, it is possible that individuals who post about birthdays do not invest much in the work domain and therefore have less conflict from the work domain to other domains.

The birthday topic theme finding is an example of how using organic data can reveal nuanced individual differences and provide details about people’s daily interactions in ways that most other methodologies cannot, as the relationship between birthday posts and work-personal conflict would be difficult to glean from traditional surveys or lab experiments. The netspeak topic theme finding is another example; the relationship between netspeak and work-personal conflict would be difficult to glean from other methods.

Future directions

Future research could investigate the relationship between social media language use, work-personal conflict, and other measures, such as perceived stress or job satisfaction. Researchers could also compare work-personal conflict in language use across different languages and locales, as this study focused on English posts from the United States. Additionally, the study of work-personal conflict and language use could be expanded to other social media platforms. Given that we are investigating the work domain, platforms like Reddit, Glassdoor, and LinkedIn could be particularly insightful.

Another potential extension of our research could be investigating work-personal conflict and language use following certain events, interventions, or policy implementations. During the COVID-19 pandemic, there were several studies conducted that investigated the effect of mandatory remote work policies on work-life balance^{60,61} and social media sentiment^{22,25,62}. It could be interesting to evaluate the impact of remote work on work-personal conflict levels and social media language in situations where individuals voluntarily choose to work remotely compared to situations in which individuals are mandated to do so.

Limitations

Our study was limited in several ways. Because of the small number of individuals ($n = 11$) who identified outside of the gender binary (i.e., did not identify as a man or a woman), we removed them from our analytic sample as we would not have been able to compute statistically significant gender differences for the non-binary group. We also removed non-English posts and users who did not have at least 500 words across their Facebook status posts in English. This decision was made in order to avoid the confounds of bilingualism. The study was also limited to the subset of the population that has a Facebook account, which may differ from the general population in significant ways (e.g., political beliefs)^{63–65}.

Additionally, Facebook status posts may not tell the complete story of an individual’s work-personal conflict. While posts provide a glimpse into an individual’s activities, moods, and thoughts, it is limited to those that the individual is willing to share publicly or with their broader network^{13,65}. There is research showing that workers are becoming increasingly worried about current and prospective employers reading their social media posts and are consequently sharing less or different content on social media with that possibility in mind^{66–69}. Despite this, it is fascinating that work-personal conflict can still be predicted by language in these posts.

Conclusion

As communication technology continues to advance and more organizations move toward remote work and similar policies, the line between work and personal life may become even blurrier, bringing work-personal conflict issues to the forefront. In this article, we provided evidence that social media language data can be used to predict and track work-personal conflict in a novel and non-intrusive way. Given the connection between work-personal conflict and general well-being, this area of research will remain critical in psychology in the years to come.

Received: 3 August 2023; Accepted: 23 November 2023

Published online: 29 November 2023

References

- Jones, S. (2021, July 15). *If you answer work emails in your leisure hours, here's some data showing why you should stop*. Business Insider. <https://www.businessinsider.com/impact-of-out-of-hour-emails-employee-wellbeing-2021-7>
- Patrick, E. (2018, April 10). *Employee internet management: Now an HR issue*. Society for Human Resource Management. https://www.shrm.org/hr-today/news/hr-magazine/pages/cms_006514.aspx
- Paulsen, R. (2014, November 3). *The art of not working at work*. The Atlantic. <https://www.theatlantic.com/business/archive/2014/11/the-art-of-not-working-at-work/382121/>
- Eby, L. T., Casper, W. J., Lockwood, A., Bordeaux, C. & Brinley, A. Work and family research in IO/OB: Content analysis and review of the literature (1980–2002). *J. Vocat. Behav.* **66**(1), 124–197. <https://doi.org/10.1016/j.jvb.2003.11.003> (2005).
- Ernst Kossek, E. & Ozeki, C. Work-family conflict, policies, and the job-life satisfaction relationship: A review and directions for organizational behavior-human resources research. *J. Appl. Psychol.* **83**(2), 139–149. <https://doi.org/10.1037/0021-9010.83.2.139> (1998).
- Ford, M. T., Heinen, B. A. & Langkamer, K. L. Work and family satisfaction and conflict: A meta-analysis of cross-domain relations. *J. Appl. Psychol.* **92**(1), 57–80. <https://doi.org/10.1037/0021-9010.92.1.57> (2007).
- Greenhaus, J. H. & Beutell, N. J. Sources of conflict between work and family roles. *Acad. Manag. Rev.* **10**(1), 76. <https://doi.org/10.2307/258214> (1985).
- Ilies, R. *et al.* When can employees have a family life? The effects of daily workload and affect on work-family conflict and social behaviors at home. *J. Appl. Psychol.* **92**(5), 1368–1379. <https://doi.org/10.1037/0021-9010.92.5.1368> (2007).
- Wilson, K. S. & Baumann, H. M. Capturing a more complete view of employees' lives outside of work: The introduction and development of new interrole conflict constructs. *Pers. Psychol.* **68**(2), 235–282. <https://doi.org/10.1111/peps.12080> (2015).
- United States Census Bureau. (2020, December 2). *Census bureau releases new estimates on America's families and living arrangements*. Census.Gov. <https://www.census.gov/newsroom/press-releases/2020/estimates-families-living-arrangements.html>
- United States Census Bureau. (2021, August 31). *First-ever census bureau report highlights growing childless older adult population*. Census.Gov. <https://www.census.gov/newsroom/press-releases/2021/childless-older-adult-population.html>
- United States Census Bureau. (2021, September 20). *Unmarried and single Americans week: September 19–25, 2021*. Census.Gov. <https://www.census.gov/newsroom/stories/unmarried-single-americans-week.html>
- Park, G. *et al.* Automatic personality assessment through social media language. *J. Pers. Soc. Psychol.* **108**(6), 934–952. <https://doi.org/10.1037/pspp0000020> (2015).
- Woo, S. E., O'Boyle, E. H. & Spector, P. E. Best practices in developing, conducting, and evaluating inductive research. *Hum. Resour. Manag. Rev.* **27**(2), 255–264. <https://doi.org/10.1016/j.hrmr.2016.08.004> (2017).
- Short, J. C., Broberg, J. C., Cogliser, C. C. & Brigham, K. H. Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organ. Res. Methods* **13**(2), 320–347. <https://doi.org/10.1177/1094428109335949> (2010).
- Schwartz, H. A. *et al.* Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One* **8**(9), e73791. <https://doi.org/10.1371/journal.pone.0073791> (2013).
- Yarkoni, T. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *J. Res. Pers.* **44**(3), 363–373. <https://doi.org/10.1016/j.jrp.2010.04.001> (2010).
- Grimmer, J. & Stewart, B. M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Polit. Anal.* **21**(3), 267–297. <https://doi.org/10.1093/pan/mps028> (2013).
- Saha, K. *et al.* A social media study on demographic differences in perceived job satisfaction. *Proc. ACM Hum. Comput. Interact.* **5**(CSCW1), 1–29. <https://doi.org/10.1145/3449241> (2021).
- Mitra, T., Muller, M., Shami, N. S., Golestani, A. & Masli, M. Spread of employee engagement in a large organizational network: A longitudinal analysis. *Proc. ACM Hum. Comput. Interact.* **1**(CSCW), 1–20. <https://doi.org/10.1145/3134716> (2017).
- Shami, N. S., Muller, M., Pal, A., Masli, M. & Geyer, W. Inferring employee engagement from social media. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3999–4008. <https://doi.org/10.1145/2702123.2702445> (2015).
- Das Swain, V., Saha, K., Reddy, M. D., Rajvanshy, H., Abowd, G. D. & De Choudhury, M. Modeling organizational culture with workplace experiences shared on Glassdoor. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3313831.3376793> (2020).
- Jacovi, M., Guy, I., Kremer-Davidson, S., Porat, S., & Aizenbud-Reshef, N. The perception of others: Inferring reputation from social media in the enterprise. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, 756–766. <https://doi.org/10.1145/2531602.2531667> (2014).
- De Choudhury, M. & Counts, S. Understanding affect in the workplace via social media. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work—CSCW '13*, 303. <https://doi.org/10.1145/2441776.2441812> (2013).
- Min, H., Peng, Y., Shoss, M. & Yang, B. Using machine learning to investigate the public's emotional responses to work from home during the COVID-19 pandemic. *J. Appl. Psychol.* **106**(2), 214–229. <https://doi.org/10.1037/apl0000886> (2021).
- Bartel, A. P. Race differences in job satisfaction: A reappraisal. *J. Hum. Resour.* **16**(2), 294. <https://doi.org/10.2307/145514> (1981).
- Clark, A. E. Job satisfaction and gender: Why are women so happy at work?. *Labour Econ.* **4**(4), 341–372. [https://doi.org/10.1016/S0927-5371\(97\)00010-9](https://doi.org/10.1016/S0927-5371(97)00010-9) (1997).
- Sousa-Poza, A. & Sousa-Poza, A. A. Taking another look at the gender/job-satisfaction paradox. *Kyklos* **53**(2), 135–152. <https://doi.org/10.1111/1467-6435.00114> (2000).
- Mannell, R. C., Kleiber, D. A. & Staempfli, M. Psychology and social psychology and the study of leisure. In *A Handbook of Leisure Studies* (eds Rojek, C. *et al.*) (Palgrave Macmillan, 2006). https://doi.org/10.1057/9780230625181_7.
- Rodell, J. B. Finding meaning through volunteering: Why do employees volunteer and what does it mean for their jobs?. *Acad. Manag. J.* **56**(5), 1274–1294. <https://doi.org/10.5465/amj.2012.0611> (2013).

31. Guy, I., Ronen, I., Zwerdling, N., Zuyev-Grabovitch, I. & Jacovi, M. What is your organization “like”? A study of liking activity in the enterprise. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3025–3037. <https://doi.org/10.1145/2858036.2858540> (2016).
32. Muscanell, N. & Guadagno, R. E. Make new friends or keep the old: Gender and personality differences in social networking use. *Comput. Hum. Behav.* **28**(1), 107–112. <https://doi.org/10.1016/j.chb.2011.08.016> (2012).
33. Acquisti, A. & Gross, R. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *Privacy Enhancing Technologies* (eds Danezis, G. & Golle, P.) (Springer, 2006). https://doi.org/10.1007/11957454_3.
34. Jones, S., Millermaier, S., Goya-Martinez, M. & Schuler, J. Whose space is MySpace? A content analysis of MySpace profiles. *First Monday* <https://doi.org/10.5210/fm.v13i9.2202> (2008).
35. Hoffman, D. L. & Novak, T. Why do people use social media? Empirical findings and a new theoretical framework for social media goal pursuit. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.1989586> (2012).
36. Krishen, A., Berezan, O., Agarwal, S. & Kachroo, P. The generation of virtual needs: Recipes for satisfaction in social media networking. *J. Bus. Res.* **69**(11), 5248–5254. <https://doi.org/10.1016/j.jbusres.2016.04.120> (2016).
37. Guntuku, S. C., Buffone, A., Jaidka, K., Eichstaedt, J. C. & Ungar, L. H. Understanding and measuring psychological stress using social media. *Proc. Int. AAAI Conf. Web Soc. Media* **13**(1), 214–225 (2019).
38. Liu, P., Tov, W., Kosinski, M., Stillwell, D. J. & Qiu, L. Do Facebook status updates reflect subjective well-being?. *Cyberpsychol. Behav. Soc. Netw.* **18**(7), 373–379. <https://doi.org/10.1089/cyber.2015.0022> (2015).
39. Hickman, L., Saha, K., De Choudhury, M. & Tay, L. Automated tracking of components of job satisfaction via text mining of Twitter data. *ML Symposium, SIOP*, 1–11 (2019).
40. Cahill, K., McNamara, T. K., Pitt-Catsoupes, M. & Valcour, M. Linking shifts in the national economy with changes in job satisfaction, employee engagement and work-life balance. *J. Behav. Exp. Econ.* **56**, 40–54. <https://doi.org/10.1016/j.socce.2015.03.002> (2015).
41. Lui, M. & Baldwin, T. langid.py: An off-the-shelf language identification tool. *Proc. ACL* **2012**, 25–30 (2012).
42. Schwartz, H. A., Giorgi, S., Sap, M., Crutchley, P., Ungar, L. & Eichstaedt, J. Dlatk: Differential language analysis toolkit. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 55–60 (2017).
43. Agarwal, S., Guntuku, S. C., Robinson, O. C., Dunn, A. & Ungar, L. H. Examining the phenomenon of quarter-life crisis through artificial intelligence and the language of Twitter. *Front. Psychol.* **11**, 341 (2020).
44. Guntuku, S. C. *et al.* Social media language of healthcare super-utilizers. *NPJ Digit. Med.* **4**(1), 1–6 (2021).
45. Mimno, D. Package ‘mallet’. A wrapper around the Java machine learning tool MALLET (2013).
46. Eichstaedt, J. C. *et al.* Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychol. Methods* **26**(4), 398–427. <https://doi.org/10.1037/met0000349> (2021).
47. Boyd, R. L., Ashokkumar, A., Seraj, S. & Pennebaker, J. W. *The Development and Psychometric Properties of LIWC-22* (University of Texas at Austin, 2022).
48. Ferreira, J. A. & Zwinderman, A. H. On the Benjamini–Hochberg method. *Ann. Stat.* **34**(4), 1827–1849. <https://doi.org/10.1214/09053606000000425> (2006).
49. Friedl, H. & Stampfer, E. Cross-validation. *Encycl. Environ.* **1**, 452–460. <https://doi.org/10.1002/9780470057339.vac062> (2002).
50. Kim, M. & Cha, J. A comparison of Facebook, Twitter, and LinkedIn: Examining motivations and network externalities for the use of social networking sites. *First Monday* <https://doi.org/10.5210/fm.v22i111.8066> (2017).
51. van Dijck, J. ‘You have one identity’: Performing the self on Facebook and LinkedIn. *Media Cult. Soc.* <https://doi.org/10.1177/01634437124686> (2013).
52. Judge, T. A., Ilies, R. & Scott, B. A. Work-family conflict and emotions: Effects at work and at home. *Person. Psychol.* **59**(4), 779–814. <https://doi.org/10.1111/j.1744-6570.2006.00054.x> (2006).
53. Nair, N. Towards understanding the role of emotions in conflict: A review and future directions. *Int. J. Conflict Manag.* **19**(4), 359–381. <https://doi.org/10.1108/10444060810909301> (2008).
54. Stoeva, A. Z., Chiu, R. K. & Greenhaus, J. H. Negative affectivity, role stress, and work-family conflict. *J. Vocat. Behav.* **60**(1), 1–16. <https://doi.org/10.1006/jvbe.2001.1812> (2002).
55. Desivilya, H. S. & Yagil, D. The role of emotions in conflict management: The case of work teams. *Int. J. Conflict Manag.* **16**(1), 55–69. <https://doi.org/10.1108/eb022923> (2005).
56. Ong, A. D. Pathways linking positive emotion and health in later life. *Curr. Dir. Psychol. Sci.* **19**(6), 358–362. <https://doi.org/10.1177/0963721410388805> (2010).
57. Staw, B. M., Sutton, R. I. & Pelled, L. H. Employee positive emotion and favorable outcomes at the workplace. *Organ. Sci.* **5**(1), 51–71. <https://doi.org/10.1287/orsc.5.1.51> (1994).
58. Bowling, A. & Windsor, J. Towards the good life: A population survey of dimensions of quality of life. *J. Happiness Stud.* **2**, 55–82. <https://doi.org/10.1023/A:1011564713657> (2001).
59. Diener, E., Suh, E. M., Smith, H. & Shao, L. National differences in reported subjective well-being: Why do they occur?. *Soc. Indic. Res.* **34**, 7–32. <https://doi.org/10.1007/BF01078966> (1995).
60. Chung, G., Chan, X. W., Lanier, P. & Ju, P. W. Y. Associations between work-family balance, parenting stress, and marital conflicts during COVID-19 pandemic in Singapore. *J. Child Family Stud.* **32**, 132–144. <https://doi.org/10.31219/osf.io/nz9s8> (2023).
61. Palumbo, R. Let me go to the office! An investigation into the side effects of working from home on work-life balance. *Int. J. Public Sector Manag.* **33**(6/7), 771–790. <https://doi.org/10.1108/IJPSM-06-2020-0150> (2020).
62. Zhang, C., Yu, M. C. & Marin, S. Exploring public sentiment on enforced remote work during COVID-19. *J. Appl. Psychol.* **106**(6), 797–810. <https://doi.org/10.1037/apl0000933> (2021).
63. Mellon, J. & Prosser, C. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Res. Politics* **4**(3), 205316801772000. <https://doi.org/10.1177/2053168017720008> (2017).
64. Ribeiro, F. N., Benevenuto, F. & Zaghenni, E. How biased is the population of Facebook users? Comparing the demographics of Facebook users with census data to generate correction factors. *12th ACM Conference on Web Science*, 325–334. <https://doi.org/10.1145/3394231.3397923> (2020).
65. Wilson, R. E., Gosling, S. D. & Graham, L. T. A review of Facebook research in the social sciences. *Perspect. Psychol. Sci.* **7**(3), 203–220. <https://doi.org/10.1177/1745691612442904> (2012).
66. Baccarella, C. V., Wagner, T. F., Kietzmann, J. H. & McCarthy, I. P. Social media? It’s serious! Understanding the dark side of social media. *Eur. Manag. J.* **36**(4), 431–438. <https://doi.org/10.1016/j.emj.2018.07.002> (2018).
67. Ball, K. Workplace surveillance: An overview. *Labor Hist.* **51**(1), 87–106. <https://doi.org/10.1080/00236561003654776> (2010).
68. Duffy, B. E. & Chan, N. K. “You never really know who’s looking”: Imagined surveillance across social media platforms. *New Media Soc.* **21**(1), 119–138. <https://doi.org/10.1177/1461444818791318> (2019).
69. Moore, P. V. The quantified self in precarity: Work, technology and what counts. *Routledge* <https://doi.org/10.4324/9781315561523> (2017).

Author contributions

L.T. and L.U. conceived of the presented idea. J.M. and N.S. performed the computations. G.L. wrote the manuscript with support from J.M., N.S., and S.C.; L.T. and S.C. supervised the project. All authors discussed the results and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-48193-3>.

Correspondence and requests for materials should be addressed to G.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023