Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup

Sarah Seraj[a,1] , Kate G. Blackburn[a] , and James W. Pennebaker[a]

[a]Department of Psychology, The University of Texas at Austin, Austin, TX 78712

Using archived social media data, the language signatures of people going through breakups were mapped. Text analyses were conducted on 1,027,541 posts from 6,803 Reddit users who had posted about their breakups. The posts include users' Reddit history in the 2 y surrounding their breakups across the various domains of their life, not just posts pertaining to their relationship. Language markers of an impending breakup were evident 3 mo before the event, peaking on the week of the breakup and returning to baseline 6 mo later. Signs included an increase in I-words, we-words, and cognitive processing words (characteristic of depression, collective focus, and the meaning-making process, respectively) and drops in analytic thinking (indicating more personal and informal language). The patterns held even when people were posting to groups unrelated to breakups and other relationship topics. People who posted about their breakup for longer time periods were less well-adjusted a year after their breakup compared to short-term posters. The language patterns seen for breakups replicated for users going through divorce (n = 5,144; 1,109,867 posts) or other types of upheavals (n = 51,357; 11,081,882 posts). The cognitive underpinnings of emotional upheavals are discussed using language as a lens.

breakups | divorce | romantic relationships | language | LIWC

Breakups of intimate relationships rarely happen out of the blue. While one partner may have been caught off guard, the other may have been planning the breakup for weeks as they gradually distanced themselves from their partner. There are several overlapping models that track the unfolding of breakups (1–4). Some suggest that people transition from an internal decision-making stage, which involves reflection about the relationship, to a dyadic stage, where people attempt to communicate their concerns to their partner which may or may not result in dissolution, to a social network stage, where people construct narratives about their breakup to share with their community. The models provide frameworks for understanding the end of relationships but fall short of identifying the precise psychological and social processes that exist at certain moments during breakups.

It is difficult to study the unraveling of real-world romantic relationships in the broader social contexts in which they occur. Studies have typically relied on retrospective reports rather than ongoing narratives in the months before, during, and after the breakup (5–9). Even among studies that track people's relationships over several months or years including before the breakup, the sample sizes are generally small and rely on self-reports. New computational methods have recently been used to conduct large-scale meta-analyses of relationships (10). However, we still do not know if and how long there are warning signs before breakups or how long the disruptive effects last on people's broader social lives.

In an age where people's social lives are intertwined with their online presence, new methods of studying breakups and other personal upheavals have emerged. By studying social media posts, researchers have already discovered language patterns related to people's emotional and psychological states, such as the onset of depression (11–13), suicide ideation (14, 15),

posttraumatic stress disorder diagnosis (16), and attention deficit hyperactivity disorder symptoms (17). Through the analysis of people's language within social media platforms, we can finally track people's evolving social and psychological processes as they go through breakups.

Cognitive Processing and Analytic Thinking. Although many relationship dissolution models emphasize the cognitive work inherent in breakups, none have been able to track the cognitive processes in real time. Researchers face several daunting challenges in studying changes in cognitive processes, including how to identify and measure the precise cognitive dynamics of interest. Recent work has identified two general language-based thinking patterns—analytic thinking and cognitive processing.

Analytic thinking involves formal, logical, and hierarchical thinking that people draw on to understand and explain complex problems (similar to Kahneman's (18) System 2 thinking). Analytic thinking is typically dispassionate and "cold" where the person lays out a problem in a reasoned way. In the context of breakups, analytic thinking is relevant during the period that people are analyzing their situation and making the decision to leave a relationship.

Various methods of capturing analytic thinking through language have been proposed, including Mehrabian and Wiener's (19) evaluations of writing samples along a dimension from high verbal nonimmediacy to high verbal immediacy. A more objective system of analytic language was discovered by Biber (20) through the factor analysis of parts of speech in works of fiction. The current research expands considerably on Biber's early work

and uses a top-down validated word counting method to identify analytic thinking.

Several studies have explored the language markers of analytic thinking and their links to psychological processes. Factor analysis of function word categories (pronouns, prepositions, articles, auxiliary verbs, negations, conjunctions, and nonreferential adverbs) on college admissions essays of about 25,000 students yielded a single factor where articles and prepositions were positively loaded and the remaining categories were negatively loaded (21). The resulting factor, analytic thinking, was positively correlated with 4-y grade point average, College Board scores, and markers of socioeconomic status. Lower analytic scores are found in speeches and communications of leaders who write and speak more informally (22) and authors who write about more personal and emotional topics compared with more factual ones (23).

The second thinking mode, called cognitive processing or working through, is the thinking that occurs when people are trying to understand problems for which they have limited knowledge. Markers of working through can be measured by a group of words, which include insight words (e.g., understand, meaning), causal words (e.g., because, result), and self-discrepancy or modal words (e.g., would, should). Multiple studies have found that cognitive processing words are connected to the ways people process traumatic events (24–26). Researchers hypothesize that cognitive words are used at greater rates to describe negative events because people are in the middle of figuring out why the event occurred and to ultimately produce a coherent narrative (27). For example, when people talk about breaking up with a romantic partner, they use more cognitive words to describe negative details of the event (5).

Self vs. Collective Focus. When breakups occur, people look inward to understand why it happened, which can sometimes lead to rumination and emotional distress (28, 29). One way to track inward focus and mental health is by counting first-person singular pronouns, or I-words. High rates of I-words are associated with modest increases in reports of depression, suicidal ideation, emotional upheavals, negative emotionality, and psychological distress (12, 30–33), indicating that the use of I-words captures the internal focus and preoccupation within individuals. Within relationship research, the use of I-words when talking about a past relationship is associated with being less adjusted after a breakup (9, 34, 35). Looking at I-word use before, during, and after a breakup may provide a means of tracking people's self-focus and adjustment challenges during the entire breakup process.

During a breakup, people's thoughts may also wander to their former partner and their role in the event. First-person plural pronouns, or we-words, have revealed information about people's relationship commitment (36), intent to continue the relationship (37), and problem-solving behaviors (38). The increased use of we-words between couples during conflict resolution (39) and marital discussions (40) highlight the interdependent nature of successful romantic relationships. People in romantic relationships may be signaling their relationship focus by using we-words. It follows that the use of we-words in ongoing committed relationships is linked to a wide range of positive relational outcomes. However, what about when the relationship is in freefall? Some research has found that greater use of we-words when sharing a breakup story is predictive of poorer adjustment (34). One explanation is that people have not yet dissociated themselves from their relationship identity and the use of we-words is to reference that shared identity. No studies have looked at how pronouns use changes over time when people are going through a breakup.

Given how central romantic relationships tend to be in people's lives, we expect that people's language may change prior to a major relationship breakup. Below we highlight some of the ways social networking sites can be used to study romantic relationships and present a methodological framework to unravel core elements of the breakup process.

Tracking Social and Cognitive Processes with Social Media. One way to study breakup processes in an organic, bottom-up approach is to use the language left behind on social networking sites. Social media data allows us to follow relationships as they unfold in real time. For example, analysis of Twitter data demonstrates that for mothers who go on to develop postpartum depression, the emotional expression and linguistic style of their posts change before their baby is even born (41). Additionally, the sheer number of observations online gives the ability to map processes for different types of events, such as people's mental health disclosures (42), social support systems (43), and even diurnal changes in mood (44). In the current study, social media data allows us to explore the breakup process as it unfolds in the real world and understand how people think through and deal with it. It also makes it possible to see how people talk about other areas of their life when they are going through a breakup, even when they are not directly talking about their relationship.

As people experience the end of the relationship, they will naturally be pushed to think about the failed relationship and what went wrong. They must consider their own role as well as their partner's. We suspect cognitive and analytic processes, the act of making sense of the event and rationally evaluating the relationship, are taking place at the same time. Given that words have been found to reflect elements of the breakup process, there may be a natural evolution to language before, during, and after a breakup. Consequently, the following set of hypotheses are posed:
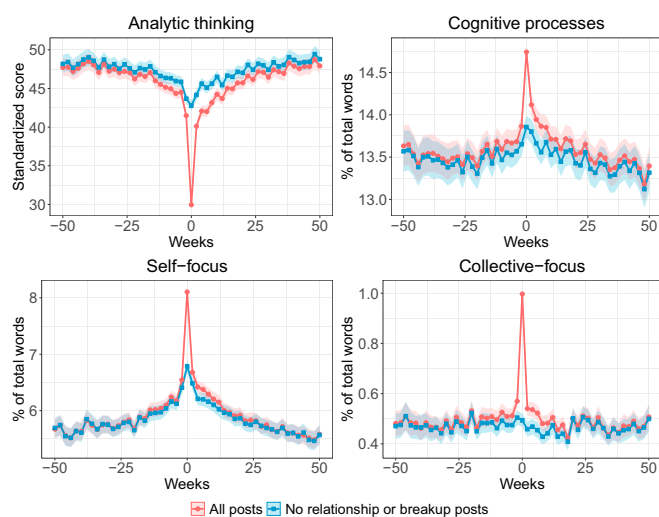
1) Cognitive processing words in people's language will increase as a breakup occurs, gradually returning to baseline levels in the months after the breakup. As people come to terms with the end of their relationship, people's use of cognitive processing words should increase, particularly when talking about their relationship. However, the rate of use will decrease as people form a coherent narrative about their breakup for themselves, their family, and friends.

2) People's language will become less analytical and more informal, personal, and dynamic as a breakup occurs. We expect that people's language will decrease on the analytic thinking dimension as the breakup unfolds, because they are more likely to be looking inward and orienting toward people, which will lead to more informal, dynamic, and personal language (mathematically the opposite of the analytic thinking dimension).

3) People's language will become more self-focused before, during, and after a breakup. The end of a relationship is a stressful life experience and can lead to inward focus and psychological distress. Since a relationship can go south long before a breakup even occurs, we expect that greater self-focused language will be observed before, during, and after the actual breakup.

4) People will use more we-words when a breakup first occurs but gradually use less as more time passes after the breakup. During the meaning-making process in a breakup, people may make more references to their shared identity with their partner, leading to increased use of we-words. As they gradually separate themselves from their partner, we-word usage should drop back to baseline levels. However, continued extensive use of we-words when talking about their relationship may be indicative of poorer adjustment following a breakup.

Results

To get a sense of what people talked about in the r/BreakUps subreddit, 1,000 first posts were randomly selected and hand coded for topical content (*SI Appendix*). Most people posting for the first time on r/BreakUps provided a rich, detailed account of the breakup (83%). People told the story of their breakup, recalling the moments that led to the breakup and the aftermath that followed. Below the quantitative analysis highlights how people's language evolved over time during the breakup process.

The linguistic patterns of r/BreakUps users were examined for up to a year before and after they publicly posted about their breakup in Reddit. The dataset contains each user's entire Reddit history in various domains of their life, not just posts pertaining to their relationship. Fig. 1 depicts graphical trends for language categories corresponding to analytic thinking, cognitive processes, self focus, and collective focus. Each data point is a 2-wk average score of Reddit posts across all the users in the dataset, with each user weighted equally at each time point regardless of the frequency of posts. Time 0 represents the time at which each user first publicly posted about their breakup in r/BreakUps and is used as a rough proxy for when the breakup occurred. The red line contains the user's entire posting history across all their subreddits. The blue line excludes any posts that were from r/BreakUps and other subreddits related to relationship issues (e.g., r/relationships, r/ExNoContact, r/Divorce, r/Marriage). In other words, the blue line represents people's language when they are not directly talking about their relationship during the breakup process, while the red line represents their entire Reddit history including their r/BreakUps posts.

Baseline Patterns Before Breakup. A prebreakup baseline period starting from a year before breakup to 4 mo before breakup was identified as a comparison time point for each hypothesis. In this period, there were no significant changes between time blocks for any of the language categories. For any statistical analyses, all posts in the baseline period were averaged by user, such that

each user had only one datapoint in the baseline period, regardless of their posting frequency (see *SI Appendix* for details on analytic strategy).

Analytic Thinking. To quantify language changes, a paired-sample t test was computed for each 2-wk period against the baseline. Results revealed decreases in analytic thinking became significant from baseline levels a month prior to the breakup disclosure [$t(2,499) = 2.11$, $P = 0.0351$, $d = 0.042$, see *SI Appendix*]. While the mean analytic thinking level did not reach baseline level until 6 mo after breakup disclosure, it was no longer significantly different from baseline after 3.5 mo [$t(2,048) = 0.84$, $P = 0.398$, $d = 0.018$]. The sharpest drop in analytic thinking was at the time of disclosure of the breakup [$t(4,707) = 52.07$, $P < 0.00001$, $d = 0.758$]. As is apparent in Fig. 1, people's language became the most personal and informal when they had just gone through the breakup. This is true even when they were not directly talking about their relationship (blue line). The results are consistent with hypothesis 2.

Cognitive Processing, Self Focus, and Collective Focus. While analytic thinking dropped, there were increases in cognitive processing, I-words, and we-words. Changes in these categories did not reach significance until 2 wk before breakup disclosure (see *SI Appendix* for statistics, only Cohen's d reported from this point forward for brevity). Like analytic thinking, these peaked at the time of the breakup with Cohen's d of 0.270, 0.580, and 0.407, respectively. The increase lasted the longest for I-words, losing significance from baseline after 10 wk. Interestingly, the patterns held even when excluding posts from subreddits relevant to breakups and relationships (90,047 posts or 8.9% of the dataset), albeit with the patterns more muted. An impending breakup causes people's language to change in dramatic ways, even when people are not directly talking about their relationship.

Overall Trends. There were observable changes in language patterns starting 3 mo before a breakup disclosure and lasting up to 6 mo after, with the entire process lasting 9 mo. The changes were statistically significant starting a month before the disclosure of a breakup and 3.5 mo after.

The language changes observed in the project are consistent with previous studies mentioned above (32–34, 45). The increase in I-words has been found to be indicative of depressive thoughts and inward focus during the breakup, while the drops in analytic thinking are consistent with less structured but more personal language. The increase in cognitive processing words suggests that people were working through their upheaval. Increase in we-words is indicative of references to shared identity as a couple, as people generate narratives about their breakup. The increase in we-words is almost completely attenuated when posts from relationship-relevant subreddits are removed. We-words are primarily used to talk about the relationship itself and so there is no increase when posting about topics unrelated to the relationship. Interestingly, cognitive processing and we-words go back to baseline within a month after breakup disclosure, while analytic thinking and self-focus takes 14 and 10 wk, respectively. It seems that even if people have cognitively worked through their breakup and stopped making references to their coupled identity, people continue to be self-immersed and have more personal language. There were some additional linguistic variables related to social and cognitive processes that may be of interest to the reader. Due to space limitations, they are provided in *SI Appendix*.

Post Hoc Questions. The results of the initial project raised two additional questions. The first concerned the potential value of writing about a breakup. That is, would writing several times help people to cope with the experience better or worse than writing

Fig. 1. Change in language patterns of Reddit users before and after a breakup. Week 0 is the point at which each user publicly disclosed their breakup on r/BreakUps. Each point is the average score across all users ($n = 6,803$) at that time block (2-wk intervals). The red line contains r/BreakUps users' entire Reddit post history in the displayed time range (1,027,541 posts). The blue line refers to all nonrelationship posts (i.e., with r/BreakUps and other relationship-relevant subreddit posts removed [936,594 posts]). The shaded areas are 95% CIs for each datapoint. Self focus is measured using I-words and collective focus using we-words.

Seraj et al.
Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup

PNAS | 3 of 7
https://doi.org/10.1073/pnas.2017154118

fewer times? The second question was whether there were any parallels in psychological effects between people going through breakups and those going through other types of emotional upheavals.

The Effect of Writing Repeatedly about a Breakup. Writing about an emotional upheaval, such as with expressive writing, is associated with modest improvements in physical and mental health (see ref. 46 for a review). Hundreds of studies have found that writing about emotional experiences for as little as 15 min a day for 1–4 d results in benefits for people dealing with a variety of problems. Other studies suggest that excessive writing about upheavals is akin to rumination and may have negative long-term effects (47). Most expressive writing studies, however, have been laboratory-based experiments.

The r/BreakUps sample could be considered a self-administered emotional writing intervention where participants could write as many times as they liked. Consequently, of the 6,803 participants, 84% wrote 1–4 d and 16% wrote 5 or more days on r/BreakUps. The language patterns of the two groups of writers were compared for shifts in cognitive and social dynamics in the months before and after their first writing submission (Fig. 2).

Both the short-term (1–4 d of posts) and long-term (5 d or more) r/BreakUps users had similar baseline values for analytic thinking, cognitive processes, I-words, and we-words before the breakup. They also mimicked each other in these linguistic patterns in the months leading up to the breakup. However, the linguistic patterns of short-term and long-term users were statistically different from each other in the first 2 wk after their breakup (*SI Appendix*). Moreover, long-term r/BreakUps users took a substantially longer amount of time to return to their baseline values following the breakup compared to short-term users. For example, while it took 2 mo after the breakup for analytic thinking of short-term users to be statistically indistinguishable from baseline, it took 6 mo for long-term users. Without considering statistical significance, analytic thinking was still diminished for long-term users a year after their breakup. The patterns are similar for the other linguistic categories. On average, it took about twice as long for long-term users to get back to baseline levels compared to short-term users. Secondary linguistic analyses were conducted to examine the topics people

talk about when they linger in r/BreakUps for longer periods of time. Those who remained active on r/BreakUps for more than 1 mo talked more about hopes of reconciliation with their partner (for details, see *SI Appendix*).

The linguistic patterns suggest that people who post about their breakup for longer time periods are less well-adjusted a year after their breakup compared to short-term posters. It is possible that long-term posters had more difficult breakups, which led them to post more, seek more social support, and be less adjusted a year after the breakup. However, the causal mechanism may be reversed. Posting about their breakup for longer periods may make it harder for people to move on from the breakup.

Comparing Breakups with Posting about Other Types of Upheavals or a Control Topic. Are the effects found in the study unique to breakups or do they extend to disclosing and writing about other events such as a divorce or even unburdening about any significant upheaval? More broadly, is it possible that merely joining a new subreddit that is not particularly emotional could provoke comparable language changes?

To assess the relative language shifts of r/BreakUps, we compared the language of people who made posts to r/Divorce and r/offmychest using the same general participant screening as the original study. Like before, the users' entire Reddit history, including posts in nonrelated subreddits, were extracted. The subreddit r/Divorce is a support forum for people going through divorce, while r/offmychest is for trauma, interpersonal conflicts, and other emotional issues. A user's first post in r/offmychest and r/Divorce is typically about the details of their upheaval. Hence, the time of their first post was used as a rough proxy for when their upheaval occurred, like what was done for r/BreakUps users. Finally, we also extracted user data from r/cooking as a neutral control subreddit, using their first submission in r/cooking as "time zero." Exclusion criteria and data cleaning procedures were applied to these datasets that were analogous to those applied for r/BreakUps. The size and demographic information of the resulting datasets are summarized in Table 1.

Fig. 3 depicts the linguistic trends of users from r/BreakUps, r/Divorce, r/offmychest, and r/cooking (control) a year before and after they first posted in these communities. The trends for r/BreakUps, r/Divorce, r/offmychest users mimic each other with some differences. While the baseline rate of analytic thinking, cognitive processes, I-words and we-words are overlapping for users of r/BreakUps and r/offmychest, r/Divorce users have different baselines. They are higher in analytic thinking and we-words, and lower in I-words and cognitive processes. r/Divorce users represent an older demographic, with a median age of 36 compared to 22 for r/BreakUps users (based on hand-coding of a subset of users, see *SI Appendix*). Previous research has shown that older adults are higher on analytic thinking and we-words and lower on self focus (48).

Despite the difference in baseline values, the magnitude of the linguistic change around the time of the upheaval was quite similar for r/Divorce and r/BreakUps users but different in r/offmychest for cognitive processes and collective focus. Because r/offmychest users post about various types of upheavals, their posts around the time of upheaval do not make as many references to other people or have narratives centered around a partner. Moreover, depending on the type of upheaval, people may not go through a meaning-making phase, leading to lower use of cognitive processing words compared to people going through a breakup or divorce.

Discussion

Breaking up is a complicated social and cognitive process that can last many months. The results suggest a natural evolution in the language people use before, during, and after a breakup.
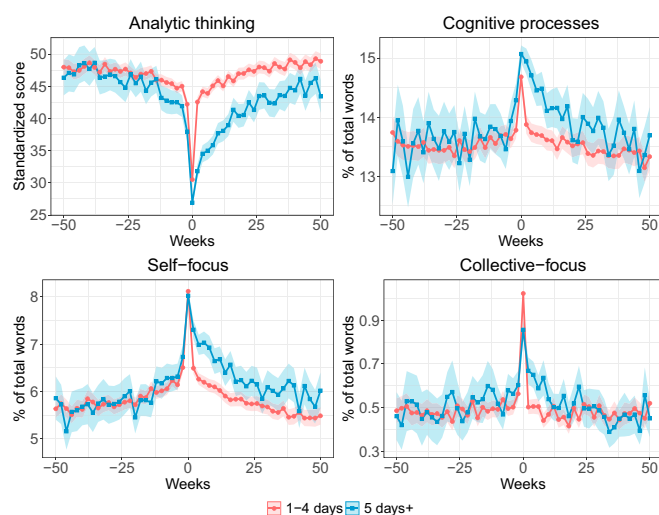
Fig. 2. The effect of writing repeatedly about a breakup. The language patterns of people who post about their breakup for 1–4 d (red line, 805,758 posts, 5,631 users) in r/BreakUps are compared to those who write for 5 d or more (blue line, 221,783 posts, 1,052 users) across all their subreddits. The shaded areas are 95% CIs for each datapoint.

Table 1. Summary of Reddit datasets

Focal subreddit (FS)	r/BreakUps	r/Divorce	r/offmychest	r/cooking
Users	6,803	5,144	51,357	21,058
No. of posts across Reddit in year before and after first FS post	1,027,541	1,109,867	11,081,882	9,169,684
Median no. of posts per user	49	57	63	165
% posts in FS	3.8	4.5	1.5	0.8
Median word count per post	50	51	49	49
Gender of users, % male	69	70	60	—
Median age	23	36	22	—
% of users dumped (% mutual)	65 (12)	53 (9)	N/A	N/A
Median relationship length, y	2	10	N/A	N/A

Entire Reddit posting history across all subreddits was obtained for users who had submissions in the subreddits r/BreakUps, r/Divorce, r/offmychest, and r/cooking. User inclusion criteria 1) ≥5 total posts, 2) posts in subreddits outside of FS, 3) posts for ≥1 mo before/after "time zero," 4) first submission in FS has word count ≥25, and 5) does not comment in FS before their first submission. Post inclusion criteria 1) Post ≥ 25 words, 2) Post ≤ 1 y before or after "time zero." Demographic information (gender, % dumped, relationship length) determined from hand-coding submissions in FS for a random 200-user subsample from each dataset with at least two independent coders. For r/Divorce, median relationship length pertains to the length of the marriage. Median no. of posts per user refers to no. of posts per user in the whole dataset after inclusion criteria have been applied. Focal subreddit (FS), subreddit that users were originally obtained from. Time 0, time of first submission in FS.

Before the breakup, we see people's natural thinking patterns on display, but the breakup disrupts this cognitive equilibrium. In fact, even before the actual breakup, analytic thinking drops as people talk about their relationship in a personal and informal manner. One explanation may be that people can sense the end of the relationship. This prebreakup phase reveals a disruption to people's normal thinking patterns starting almost 3 mo before the breakup.

A second cognitive process is activated when the breakup occurs. As people make decisions about their new lives, their language spikes in the use of cognitive processing words. Finally, as the story becomes more developed and organized, analytic thinking increases again. The fluctuation in analytic and cognitive processing words reveals two dynamic cognitive mechanisms that unfold over the course of a breakup. Indeed, in many ways, these two cognitive processes may be tied to the way we encode experiences and map them into memories. We can see people's thought process through their word use before a breakup even takes place. Additionally, tracking what happens during the moment of the breakup gives access to the ways people are trying to explain to themselves and others why the breakup occurred.

As seen in the post hoc analyses, those who write about their breakups more frequently are slower to return to their pre-breakup language patterns. One explanation is simply that people who need to write continually may have experienced more disruptive or traumatic breakups. Alternatively, writing about the same events repeatedly may be a form of rumination whereby people are reliving the same distressing events over and over. In fact, expressive writing studies have found that people who write about emotional upheavals in similar ways on multiple occasions often do not show as many health benefits compared to those who update their narrative over time (47). By repeatedly recalling the same experience over several months, those who continue to relive the same painful memories might benefit from an alternative coping strategy, such as seeking clinical intervention.

Using language analysis tools, social scientists can now track social shifts in human connections in near-real time. Within hours of people revealing their broken hearts, it is possible to detect how they are communicating with other parts of their social networks about their hobbies, jobs, or religion. Communities such as Reddit provide a laboratory for researchers to measure how different coping strategies can potentially work. One contribution of the current research is that it points to the power of analyzing social media data to understand the unfolding dynamics of interpersonal processes.

Materials and Methods

The current project relied on Reddit, one of the most frequently visited websites in the world with over 430 million unique visitors each month. People gather in one or more of over 180,000 online communities called subreddits to connect with others about their hobbies, interests, and concerns (49). Unlike most other social media sites, Reddit does not require users to provide any identifying information in their profile. Users post using a username or handle, which is typically not tied to their real names.

What makes Reddit unique is that all posts and comments can be read by anyone. Using Reddit's application programming interface, researchers can obtain the entire Reddit history of individuals or of any subreddit. In short, Reddit allows researchers to track the full written record of each contributing Reddit user since the beginning of Reddit in 2005.

The current project relied on the r/BreakUps subreddit which, as of May 22, 2019, had ~87,500 subscribers with about 150 submissions and 370 comments per day (Subreddit Stats, 2019). A similar subreddit called r/BreakUp was not used given that it had only about 6,000 subscribers. The r/BreakUps dataset initially consisted of over 6 million posts from 22,000 users. Posts were defined as either submissions (a new entry in a particular subreddit) or comments (conversation pertaining to a particular submission). We treated a
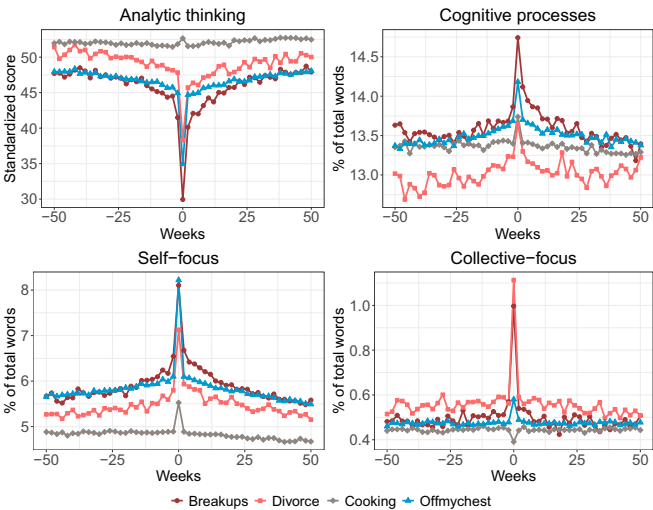

Fig. 3. Comparison of breakups to divorce and other emotional upheavals. The language patterns of people who post about their breakup (dark red line) are compared to those who post about their divorce (light red line) and other upheavals (blue line) across all their subreddits. In each case, Week 0 is when users first posted about their upheaval. Posts from users in the cooking subreddit is included as a control (gray line), with week 0 representing the time they first posted in r/cooking.

Seraj et al.
Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup

PNAS | 5 of 7
https://doi.org/10.1073/pnas.2017154118

user's first submission on r/BreakUps as a rough proxy for when their breakup occurred ("time zero").

The goal of the study was to track the language of people undergoing a breakup in the year before to the year after their breakup. People's language was analyzed both within the r/BreakUps subreddit as well as in all the other subreddits they contributed to during this 2-y window. In collecting the final dataset, users were required to have at least five total posts on Reddit, at least one post outside of r/BreakUps, and posts at least 1 mo before and after their breakup. Since a user's first r/BreakUps submission was used as a proxy for when the breakup occurred, any users who commented on other people's breakup posts first before creating their own submission (~11% of the users) were excluded. This was to remove potential lurkers (users who read posts but contributed little) as well as people who posted about their breakups after a substantial amount of time had passed after the event. The final sample included the text from 6,803 users who had gone through breakups. Text analysis was performed on all user posts of at least 25 words within a year before or after their breakup post (n = 1,027,541 posts). For a more detailed account of the dataset and related Reddit demographics, see *SI Appendix*.

To test whether the effects observed for r/BreakUps users were unique to people going through breakups or could be extrapolated to other types of emotional upheavals, three additional datasets were also obtained with users from r/Divorce, r/offmychest, r/cooking. The datasets are described in more detail in the *Post Hoc Questions*.

Text Analysis Procedure. Text samples were primarily analyzed using the text analysis program Linguistic Inquiry and Word Count (LIWC) (50). LIWC analyzes the emotional, cognitive, and structural components of language samples by calculating the percentage of words that correspond to different language dimensions. The goal was to look at changes in people's cognitive and social lives up to a year before and after a breakup. The analysis focuses on the following LIWC categories that correspond to these dimensions: analytic thinking, cognitive processes, self focus (I-words), and collective focus (we-words).

Analytic thinking is a standardized measure that goes from 0 to 100. It is a factor analytically derived dimension based on the use of function words.

High scores in analytic thinking have higher rates of articles and prepositions, indicating references to objects, concepts, and linkages between them (21, 22). It is also associated with more formal, impersonal language. However, low scores in analytic thinking have higher rates of pronouns, adverbs, auxiliary verbs, negations, and conjunctions, indicative of more dynamic and personal language.

While analytic thinking is a standardized measure, the remaining dimensions indicate the percentage of words in a text corresponding to that category. Cognitive processes include words related to thinking through and understanding events (e.g., think, because, perhaps). Self-focus refers to first-person singular pronouns (e.g., I, myself, me, mine) while collective focus refers to first-person plural pronouns (e.g., we, us, ours). Examples of excerpts high in each of these dimensions are provided in *SI Appendix*. Only four dimensions from LIWC are included here in Fig. 1, but the results from additional dimensions of interest can be found in *SI Appendix*.

1. L. A. Baxter, Trajectories of relationship dissolution. *J. Soc. Pers. Relat.* **1**, 29–48 (1984).
2. L. Lee, Sequences in separation: A framework for investigating endings of the personal (romantic) relationship. *J. Soc. Pers. Relat.* **1**, 49–73 (1984).
3. S. S. Rollie, S. W. Duck, "Divorce and dissolution of romantic relationships: Stage models and their limitations" in *Handbook of Divorce and Relationship Dissolution*, M. A. Fine, J. H. Harvey, Eds. (Lawrence Erlbaum Associates, Inc., 2006), pp. 223–240.
4. M. L. Knapp, A. L. Vangelisti, "Relationship stages: A communication perspective" in *Interpersonal Communication and Human Relationships, 5th Edition*, M. L. Knapp, A. L. Vangelisti, Eds. (Allyn & Bacon, 2005), pp. 3936–3949.
5. A. Boals, K. Klein, Word use in emotional narratives about failed romantic relationships and subsequent mental health. *J. Lang. Soc. Psychol.* **24**, 252–268 (2005).
6. D. Davis, P. R. Shaver, M. L. Vernon, Physical, emotional, and behavioral reactions to breaking up: The roles of gender, age, emotional involvement, and attachment style. *Pers. Soc. Psychol. Bull.* **29**, 871–884 (2003).
7. D. M. Wegner, D. B. Gold, Fanning old flames: Emotional and cognitive effects of suppressing thoughts of a past relationship. *J. Pers. Soc. Psychol.* **68**, 782–792 (1995).
8. J. Mearns, Coping with a breakup: Negative mood regulation expectancies and depression following the end of a romantic relationship. *J. Pers. Soc. Psychol.* **60**, 327–334 (1991).
9. J. L. Borelli, D. A. Sbarra, Trauma history and linguistic self-focus moderate the course of psychological adjustment to divorce. *J. Soc. Clin. Psychol.* **30**, 667–698 (2011).
10. S. Joel *et al.*, Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 19061–19071 (2020).
11. A. Yates, A. Cohan, N. Goharian, "Depression and self-harm risk assessment in online forums" in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (Association for Computational Linguistics, Stroudsburg, PA, 2017), pp. 2968–2978.
12. M. De Choudhury, M. Gamon, S. Counts, E. Horvitz, "Predicting depression via social media" in *Proceedings of the International AAAI Conference on Weblogs and Social Media* (AAAI Press, 2013).
13. J. C. Eichstaedt *et al.*, Facebook language predicts depression in medical records. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11203–11208 (2018).
14. B. Desmet, V. Hoste, Online suicide prevention through optimised text classification. *Inf. Sci. (Ny)* **439–440**, 61–78 (2018).
15. M. De Choudhury, E. Kiciman, "The language of social support in social media and its effect on suicidal ideation risk" in *Proceedings of the 11th International AAAI Conference on Weblogs Social Media* (AAAI Press, 2017), pp. 32–41.
16. G. A. Coppersmith, C. T. Harman, M. H. Dredze, "Measuring post traumatic stress disorder in twitter" in *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (AAAI Press, 2014).
17. S. C. Guntuku, J. R. Ramsay, R. M. Merchant, L. H. Ungar, Language of ADHD in adults on social media. *J. Atten. Disord.* **23**, 1475–1485 (2019).
18. D. Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011).
19. A. Mehrabian, M. Wiener, Non-immediacy between communicator and object of communication in a verbal message: Application to the inference of attitudes. *J. Consult. Psychol.* **30**, 420–425 (1966).
20. D. Biber, A typology of English texts. *Linguistics* **27**, 3–44 (1989).
21. J. W. Pennebaker, C. K. Chung, J. Frazee, G. M. Lavergne, D. I. Beaver, When small words foretell academic success: The case of college admissions essays. *PLoS One* **9**, e115844 (2014).
22. K. N. Jordan, J. Sterling, J. W. Pennebaker, R. L. Boyd, Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 3476–3481 (2019).
23. R. L. Boyd, K. G. Blackburn, J. W. Pennebaker, The narrative arc: Revealing core narrative structures through text analysis. *Sci. Adv.* **6**, eaba2196 (2020).
24. C. K. Chung, J. W. Pennebaker, "The psychological functions of function words" in *Social Communication*, K. Fiedler, Ed. (Psychology Press, 2007), pp. 343–359.
25. M. A. Cohn, M. R. Mehl, J. W. Pennebaker, Linguistic markers of psychological change surrounding September 11, 2001. *Psychol. Sci.* **15**, 687–693 (2004).
26. B. Kleim, A. B. Horn, R. Kraehenmann, M. R. Mehl, A. Ehlers, Early linguistic markers of trauma-specific processing predict post-trauma adjustment. *Front. Psychiatry* **9**, 645 (2018).
27. J. W. Pennebaker, M. R. Mehl, K. G. Niederhoffer, Psychological aspects of natural language. Use: Our words, our selves. *Annu. Rev. Psychol.* **54**, 547–577 (2003).
28. E. B. Slotter, W. L. Gardner, E. J. Finkel, Who am I without you? The influence of romantic breakup on the self-concept. *Pers. Soc. Psychol. Bull.* **36**, 147–160 (2010).
29. A. Mirsu-Paun, J. A. Oliver, How much does love really hurt? A meta-analysis of the association between romantic relationship quality, breakups and mental health outcomes in adolescents and young adults. *J. Relationsh. Res.* **8**, 1–12 (2017).
30. W. Weintraub, *Verbal Behavior: Adaptation and Psychopathology* (Springer, 1981).
31. D. M. Markowitz, J. T. Hancock, The 27 club: Music lyrics reflect psychological distress. *Commun. Rep.* **30**, 1–13 (2017).
32. M. Lyons, N. D. Aksayli, G. Brewer, Mental distress and language use: Linguistic analysis of discussion forum posts. *Comput. Human Behav.* **87**, 207–211 (2018).
33. A. M. Tackman *et al.*, Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *J. Pers. Soc. Psychol.* **116**, 817–834 (2019).
34. K. Blackburn, N. Brody, L. LeFebvre, The I's, we's, and she/He's of breakups: Public and private pronoun usage in relationship dissolution accounts. *J. Lang. Soc. Psychol.* **33**, 202–213 (2014).

35. L. A. Lee, D. A. Sbarra, A. E. Mason, R. W. Law, Attachment anxiety, verbal immediacy, and blood pressure: Results from a laboratory analog study following marital separation. *Pers. Relatsh.* **18**, 285–301 (2011).

36. C. R. Agnew, P. A. M. Van Lange, C. E. Rusbult, C. A. Langston, Cognitive interdependence: Commitment and the mental representation of close relationships. *J. Pers. Soc. Psychol.* **74**, 939–954 (1998).

37. R. B. Slatcher, S. Vazire, J. W. Pennebaker, Am "I" more important than "we"? Couples' word use in instant messages. *Pers. Relatsh.* **15**, 407–424 (2008).

38. J. N. Biesen, D. E. Schooler, D. A. Smith, What a difference a pronoun makes: I/We versus you/me and worried couples' perceptions of their interaction quality. *J. Lang. Soc. Psychol.* **35**, 180–205 (2016).

39. R. A. Simmons, P. C. Gordon, D. L. Chambless, Pronouns in marital interaction. *Psychol. Sci.* **16**, 932–936 (2005).

40. A. L. Sillars, W. Shellen, A. McIntosh, M. Pomegrenate, Relational characteristics of language: Elaboration and differentiation in marital conversations. *West. J. Commun.* **61**, 403–422 (1997).

41. M. De Choudhury, S. Counts, E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media." *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM, New York, 2013), pp. 3267–3276.

42. M. De Choudhury, S. De, "Mental health discourse on reddit: Self-disclosure, social support, and anonymity". *Eighth International AAAI Conference on Weblogs Social Media* (AAAI Press, 2014).

43. A. R. D'Agostino *et al.*, Social networking online to recover from opioid use disorder: A study of community interactions. *Drug Alcohol Depend.* **181**, 5–10 (2017).

44. K. Ethayarajh, F. Rudzicz, The effect of photoperiod on the mood of Reddit users. *Cyberpsychol. Behav. Soc. Netw.* **20**, 238–245 (2017).

45. A. Boals, J. B. Banks, L. M. Hathaway, D. Schuettler, Coping with stressful events: Use of cognitive words in stressful narratives and the meaning-making process. *J. Soc. Clin. Psychol.* **30**, 378–403 (2011).

46. J. Frattaroli, Experimental disclosure and its moderators: A meta-analysis. *Psychol. Bull.* **132**, 823–865 (2006).

47. J. W. Pennebaker, J. M. Smyth, *Opening Up by Writing It Down* (The Guilford Press, 3rd Ed., 2016).

48. J. W. Pennebaker, L. D. Stone, Words of wisdom: Language use over the life span. *J. Pers. Soc. Psychol.* **85**, 291–301 (2003).

49. N. Djordjevic, 109 ridiculous reddit statistics & facts to know in 2020. WebsiteBuilder.org (2020). https://websitebuilder.org/blog/reddit-statistics/. Accessed 17 January 2021.

50. J. W. Pennebaker, R. L. Boyd, K. Jordan, K. Blackburn, The development and psychometric properties of LIWC2015. (2015). http://liwc.wpengine.com/wp-content/uploads/2015/11/LIWC2015_LanguageManual.pdf. Accessed 17 January 2021.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Seraj et al.
Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup

PNAS | 7 of 7
https://doi.org/10.1073/pnas.2017154118