SOCIAL PSYCHOLOGY

# Affect labeling in the age of social media

Does tweeting your feelings change how you feel? A study of over a billion tweets shows that we tend to tweet about our feelings after they have escalated. However, such 'affect labeling' tweets — even though they are constrained to 140 characters — lead to rapid reductions in the intensity of our emotions.

Matthew D. Lieberman

In Macbeth, Shakespeare wrote, "Give sorrow words. The grief that does not speak whispers the o'er-fraught heart and bids it break." Although Shakespeare was ahead of his time and wrote for the common man, he could not have possibly anticipated Twitter, which provides a common platform of expression for everyone from high-school students to world leaders and celebrities. Was Shakespeare right that putting feelings into words helps to ameliorate those feelings and is this just as true on social media as it might be in our diaries and therapy sessions? A study by Bollen and collegues[1] in *Nature Human Behaviour* uses an innovative approach to address these questions.

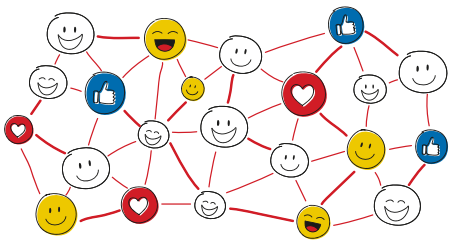
There are at least three possible dynamics between putting feelings into words and the underlying feelings themselves. First, emotional language may reveal feeling states without having any reciprocal impact on them. Second, emotional language may amplify emotional states by bringing additional attention to them. Third, emotional language may serve to regulate emotional states, diminishing their intensity.

Although most people believe putting feelings into words will amplify them[2], several studies of affect labeling suggest that putting feelings into words has the capacity to dampen down those feelings, especially when the labelled emotional state is negative[3]. Successful emotion regulation is associated with a variety of experiential, physiological and neural responses and many of these same patterns have been observed during affect labeling. For instance, affect labeling increases activity in the ventrolateral prefrontal cortex and leads to associated decreases in amygdala activity, similar to what is seen in reappraisal, a typical form of emotion regulation[4–6]. Affect labeling has even been shown to have therapeutic effects in several mental health contexts, including phobias, post-traumatic stress disorder and public speaking[7–9].

Studies of affect labeling have all been conducted in experimental laboratory settings, which limits these studies in various ways. Naturally occurring affect labeling has not been studied, so little is known about the effects of spontaneous affect labeling in the 'real world'. Lab studies have only looked at responses immediately after affect labeling (or occasionally a week later) and thus have not been able to examine the temporal components of emotion before, during and after an affect labeling episode. Finally, lab studies are constrained to relatively small participant samples.

The exciting study by Bollen and colleagues[1] addresses these issues and adds significantly to our understanding of naturally occurring emotions and emotional language by examining spontaneous occurrences of affect labeling on Twitter. The authors examined over a billion tweets to find those involving affect labeling. Such tweets had to include a variant of "I feel…" followed by one of several positive or negative emotion words (for example, 'bad', 'good', 'sad', 'amazing'). The authors then performed sentiment analysis of any other tweets in the six hours before and after any of the 109,943 affect labeling tweets that were identified. This creates a temporal profile of emotion time locked to the affect labeling event. It is important to note that the affect labeling tweet itself was merely used to identify the critical moment in time and was not actually included in the sentiment analyses themselves.

For both positive and negative affect labeling, there was a ramping up period before the labelling event. That is, the valence of tweets increased in their intensity for the half hour to hour before the affect labeling tweet. For both positive and negative affect labeling, the moment of greatest emotional intensity occurred right around the same time as the labelling event. What is of greatest interest theoretically is what happened after the affect labeling event. When a tweet labelled a negative affective state, there was an almost immediate dramatic return to emotional baseline in subsequent tweets. Statistically, this same pattern was present for positive affect labeling as well, but in less dramatic form.



Credit: Karolina Madej/iStock/Getty Images Plus/Getty

This is a remarkable set of findings. These results demonstrate a temporal sequence of emotion that affect labeling studies have all predicted but have never shown. It appears that as emotional intensity increases, the likelihood of spontaneous affect labeling increases. Then, as soon as this labelling has occurred there is a rapid decline in affective intensity. In the case of negative emotion, the decline is precipitous. These results are very consistent with the hypothesis that affect labeling causes the downregulation of both positive and negative emotions[2].

Of course, all studies have limitations and this one is no exception. Sentiments in tweets are not a perfect mirror of experienced emotion. On Twitter, people curate what they say and how they choose to represent themselves in a variety of ways. Experienced emotion may interact with the general tendency to post, distorting the relationship between actual emotion and what is presented online. The authors also excluded many tweets that were not variants of "I feel…" and yet would still have constituted affect labeling (for example, "sadness describes me too much today"), but this was a conservative approach

ensuring that included tweets were definitely instances of affect labeling.

In the end, the authors took a creative approach to studying affect labeling out in the real world and produced some of the strongest and most comprehensive data in support of the role of affect labeling in dampening affective intensity. Apparently, Shakespeare's maxim to "give sorrow words" works just as well online as it did in his time. ❐

Matthew D. Lieberman
*Psychology Department, University of California, Los Angeles, Los Angeles, CA, USA.*
*e-mail:* lieber@ucla.edu

Published online: 17 December 2018
https://doi.org/10.1038/s41562-018-0487-0

## References

1. Fan, R. et al. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-018-0490-5 (2018).
2. Lieberman, M. D., Inagaki, T. K., Tabibnia, G. & Crockett, M. J. *Emotion* **3**, 468–480 (2011).
3. Torre, J. B. & Lieberman, M. D. *Emot. Rev.* **10**, 116–124 (2018).
4. Hariri, A. R., Bookheimer, S. Y. & Mazziotta, J. C. *Neuroreport* **11**, 43–48 (2000).
5. Lieberman, M. D. et al. *Psychol. Sci.* **18**, 421–428 (2007).
6. Burklund, L. J., Creswell, J. D., Irwin, M. & Lieberman, M. D. *Front. Psychol.* **5**, 221 (2014).
7. Kircanski, K., Lieberman, M. D. & Craske, M. G. *Psychol. Sci.* **23**, 1086–1091 (2012).
8. Niles, A. N., Craske, M. G., Lieberman, M. D. & Hur, C. *Behav. Res. Ther.* **68**, 27–35 (2015).
9. Memarian, N., Torre, J. B., Haltom, K. E., Stanton, A. L. & Lieberman, M. D. *Soc. Cogn. Affect. Neurosci.* **12**, 1437–1447 (2017).

## Competing interests

The author declares no competing interests.