

Using Psychologically-Informed Priors for Suicide Prediction in the CLPsych 2021 Shared Task

Avi Gamoran * and Yonatan Kaplan *

Ram Isaac Orr and Almog Simchon † and Michael Gilead †

Ben-Gurion University of the Negev, Israel

{avigam, kaplay, ramor, almogsi}@post.bgu.ac.il

mgilead@bgu.ac.il

Abstract

This paper describes our approach to the CLPsych 2021 Shared Task, in which we aimed to predict suicide attempts based on Twitter feed data. We addressed this challenge by emphasizing reliance on prior domain knowledge. We engineered novel theory-driven features, and integrated prior knowledge with empirical evidence in a principled manner using Bayesian modeling. While this theory-guided approach increases bias and lowers accuracy on the training set, it was successful in preventing over-fitting. The models provided reasonable classification accuracy on unseen test data ($0.68 \leq AUC \leq 0.84$). Our approach may be particularly useful in prediction tasks trained on a relatively small data set.

1 Introduction

Suicide is a troubling public health issue (Haney et al., 2012), with an estimated prevalence of over 800,000 cases per year worldwide (Arensman et al., 2020). Suicide rates have been climbing steadily over the past two decades (Curtin et al., 2016; Naghavi, 2019; Glenn et al., 2020), especially in high-income countries (Arensman et al., 2020; Haney et al., 2012). Research has identified many risk factors linked to suicide (Franklin et al., 2017; Ribeiro et al., 2018), and suicide attempts (Yates et al., 2019; Miranda-Mendizabal et al., 2019). Despite these advances, directing these insights into real-life risk identification and suicide prevention remains challenging (Large et al., 2017b,a). Early identification is crucial, as direct, brief, and acute interventions are helpful in preventing suicide attempts (Doupnik et al., 2020).

For the sake of early detection, there are increasing attempts to try and find warning signs in publicly-available social media data. As part of this effort, the 2021 Computational Linguistics and

Clinical Psychology Workshop (CLPsych), have provided access to de-identified Twitter feeds of individuals who have made suicide attempts (as well as others who have not), with the task of predicting suicide attempts based on tweets up to 30 days (*Subtask 1*) or 182 days (*Subtask 2*) before such attempts.

Machine-learning algorithms and natural language processing ("NLP") methods have proven highly useful on many prediction problems. Current approaches typically rely on inductive algorithms that learn regularities in the data. When data are noisy (as is the case in human behavior), the ability to generalize predictions often depends on the size of the training set. Given the sensitive nature of suicide-related data, labeled data on this matter are scarce. This relative scarcity of training examples (e.g., 114/164 individuals in the current task) presents a difficult prediction problem, and increased risk of model over-fitting.

In light of the unique properties of this problem, we reasoned that an emphasis on domain knowledge (rather than on algorithmic solution) is warranted, and may help reduce over-fitting. Therefore, we adopted the following principles for the prediction task: 1. We used logistic regression rather than potentially more complex models that are often more prone to over-fitting (e.g., DNN, SVM, RF). 2. We engineered and evaluated many theory-driven features, based on our domain expertise in psychology (e.g., Simchon and Gilead, 2018). 3. We integrated prior knowledge and the empirical evidence in a principled manner. Using Bayesian modeling, we incorporated empirical priors from past findings in psychology literature. When we lacked specific priors for a feature of interest, we regularized our parameters using general, domain-level empirical priors (van Zwet and Gelman, 2020), derived from a meta-analysis of replication studies in psychology (Open Science Collaboration et al., 2015).

* These authors contributed equally.

† These authors contributed equally.

2 Methodology

Participants in the Shared Task were given a training set which consisted of 2485 tweets from 114 individuals, 57 having attempted suicide and 57 controls, in the 30-day set, and 15928 tweets from 164 individuals, 82 in each group, in the 182-day set.

2.1 Features

| Feature with Informed Priors | Effect-Size (r) |
|------------------------------|---------------------|
| Adverbs-SD | 0.113 |
| Anger-M | 0.068 |
| Anger-SD | 0.068 |
| Body-SD | 0.07 |
| Female-M | 0.105 |
| Female-SD | 0.105 |
| Focus-On-Present-SD | 0.095 |
| Informal-SD | 0.041 |
| Ingest-SD | 0.021 |
| I-Pronouns-M | 0.046 |
| Negative-Emotion-M | 0.141 |
| Negative-Emotion-SD | 0.141 |
| Pronouns-M | 0.137 |
| Personal-Pronouns-M | 0.015 |
| Sexual-M | 0.073 |
| Sexual-SD | 0.073 |
| Swear-Words-M | 0.055 |
| Swear-Words-SD | 0.055 |
| Verbs-M | 0.101 |
| Work-M | -0.099 |
| They-M | 0.025 |

Table 1: LIWC Features with Informed Priors (Effect sizes from [Eichstaedt et al., 2018](#)). Effect sizes entered the model on the log odds scale. Shown here in Pearson’s r for convenience.

Twitter behavioral aspects: We counted the number of replies to others, and the number of unique fellow users mentioned in replies. The intuition behind these metrics being that they reflect on the social engagement of users. Loneliness and social isolation are robust risk factors for suicide ([Leigh-Hunt et al., 2017](#); [Franklin et al., 2017](#)). The proportion of tweets written late at night (23:00 – 5:00) was measured, as sleep disorders are related to depression and suicidal ideation ([Liu et al., 2020](#)).

LIWC: The Linguistic Inquiry and Word Count ([Pennebaker et al., 2015](#)), is a widely used

dictionary-based program for automatic text analysis. LIWC scales tap into psychological and linguistic features, and provide a good overview into an individual’s psychological makeup ([Chung and Pennebaker, 2018](#)). LIWC has been used in analyzing social media prior to suicide attempts ([Coppersmith et al., 2016](#)), as well as in analysis of suicide notes ([Pestian et al., 2012](#)) and poems of poets who later committed suicide ([Stirman and Pennebaker, 2001](#)). A central finding from LIWC analyses on suicidal populations is an increase in words pertaining to the self, and a decrease in words regarding others. We therefore measured the ratio of self words (‘I’) to group-words (‘We’). Most of the LIWC-derived features were given priors based on previous gold-standard findings in depression prediction, see Table 1 ([Eichstaedt et al., 2018](#)).

The Mind-Perception Dictionary: a dictionary tailored for mind perception which includes a category of agent-related emotions ([Schweitzer and Waytz, 2020](#)). The guiding idea was that individuals at risk of committing suicide may differ in their sense of agency from non-suicidal individuals. This feature was given a weakly-informed prior with center = 0.

Custom Dictionaries: We constructed custom dictionaries based on themes assumed to be linked with mental vulnerability, depression and suicide. The themes included were Social Longing, Fatigue, Self-destructive Behavior, and Unmet Desires and Needs. These features were given weakly-informed priors with center = 0.

2.2 Bayesian Modeling

Due to the large amount of potential predictive features, as a first step, we manually excluded variables which did not differ between suicidal individuals and controls in a univariate statistical analysis. A total of 30 significant variables were retained for the modeling stage (Table 1).

Using the ‘rstanarm’ package, an R wrapper for Stan ([Carpenter et al., 2017](#); [Goodrich et al., 2020](#)), we deployed logistic-regression models with Bayesian MCMC estimation. The Bayesian infrastructure was chosen in order to formally determine custom priors for the various predictive features, based on existing psychological literature, and to regularize parameters based on the distribution of effect sizes in the field.

In order to assess the validity of this approach and its performance relative to inductive “bottom-

up" methods, we chose to submit one psychologically informed model, one "default" weakly-informed Bayesian model, and one regularized regression model.

Our models were: *a*) Informed priors with centers of distributions according to effect sizes found in previous studies (Table 1). In *Subtask 1* the priors were from Cauchy distributions, with centers according to existing effect sizes, and scales set to 2.5 (the 'rstanarm' defaults): $\sim \text{Cauchy}(\mu, 2.5)$. In *Subtask 2* the priors were from Laplace distributions with centers according to effect sizes, and scales of 1.687 as an approximation of a mixture prior, recommended for use in a database of 86 psychological replication studies (van Zwet and Gelman, 2020): $\sim \mathcal{L}(\mu, 1.687)$. For an example of the Bayesian approach see Figure 1. *b*) Weakly-informed priors based on the 'rstanarm' defaults without any formal customizing. *c*) A regularized regression algorithm, using the 'glmnet' (Friedman et al., 2010) and 'caret' (Kuhn, 2020) R packages. In *Subtask 1* the model with optimal accuracy included $\alpha = 0$, ("Ridge" regression), and in *Subtask 2* it included $\alpha = 1$ ("Lasso" regression).

3 Results

3.1 Subtask 1

In *Subtask 1* the goal was to predict which Individuals were likely to attempt suicide based on tweets up to 30 days prior. Model performances on the training set are displayed in Table 2. The first model (M1) was a Bayesian logistic-regression model using psychologically informed priors. We compared 2 types of distributions for the priors (around the custom centers). The first, a Cauchy distribution with scales set at 2.5. The second, a Laplace distribution with scales of 1.687 (see "Bayesian Modeling" above). In the *Subtask 1* training set, the Informed-Priors Cauchy distribution slightly outperformed the Informed-Priors Laplace distribution in a 5-fold cross-validation.

The second model (M2) was a weakly-informed Bayesian logistic-regression model with priors drawn from a Cauchy Distribution with center = 0 and scale = 2.5.

The third model (M3) was logistic-regression model with regularization. We conducted 5-fold cross validation, with 3 repeats for hyper-parameter tuning of the penalty type (α), and the regularization parameter (λ). In the *Subtask 1* training set, the optimal prediction accuracy included the hyper-

| | F1 | F2 | TPR | FPR | AUC |
|-----------------------------|-------|-------|-------|-------|-------|
| Subtask 1 (30 days) | | | | | |
| M1 | 0.466 | 0.452 | 0.447 | 0.423 | 0.543 |
| M2 | 0.480 | 0.474 | 0.476 | 0.436 | 0.546 |
| M3 | 0.589 | 0.580 | 0.573 | 0.374 | 0.599 |
| Subtask 2 (6 months) | | | | | |
| M1 | 0.586 | 0.529 | 0.499 | 0.187 | 0.739 |
| M2 | 0.668 | 0.626 | 0.602 | 0.184 | 0.745 |
| M3 | 0.710 | 0.670 | 0.646 | 0.175 | 0.735 |

Table 2: 5-fold CV Results. M1: Informed priors; M2: Weakly-informed priors; M3: Ridge/Lasso regression.

| | F1 | F2 | TPR | FPR | AUC |
|-----------------------------|-------|-------|-------|-------|-------|
| Subtask 1 (30 days) | | | | | |
| BL | 0.636 | 0.636 | 0.636 | 0.364 | 0.661 |
| M1 | 0.526 | 0.481 | 0.455 | 0.273 | 0.678 |
| M2 | 0.526 | 0.481 | 0.455 | 0.273 | 0.678 |
| M3 | 0.421 | 0.385 | 0.364 | 0.364 | 0.636 |
| Subtask 2 (6 months) | | | | | |
| BL | 0.710 | 0.724 | 0.733 | 0.333 | 0.764 |
| M1 | 0.769 | 0.704 | 0.667 | 0.067 | 0.809 |
| M2 | 0.769 | 0.704 | 0.667 | 0.067 | 0.791 |
| M3 | 0.815 | 0.764 | 0.733 | 0.067 | 0.844 |

Table 3: Official Test Results. BL: Task Baseline; M1: Informed priors; M2: Weakly-informed priors; M3: Ridge/Lasso regression.

parameters $\alpha = 0$ ("Ridge"), and $\lambda = 10$.

3.2 Subtask 2

In *Subtask 2* the goal was to predict which Individuals were likely to attempt suicide from tweets up to 6 months (182 days) prior. M1 was a Bayesian logistic-regression model using psychologically informed priors. Like in *Subtask 1*, We compared 2 types of distributions for the priors: Cauchy and Laplace. In the *Subtask 2* training set, the Informed-Priors Laplace distribution outperformed the Informed-Priors Cauchy.

M2 again included a weakly-informed Bayesian logistic-regression model.

M3 was once more a regularized logistic-regression model. In the *Subtask 2* training set, the optimal prediction accuracy included $\alpha = 1$ ("Lasso"), and $\lambda = 0.1$.

Results on the test set are displayed in Table 3. In both tasks models yielded above-chance predictions, and performed better on the test set than the

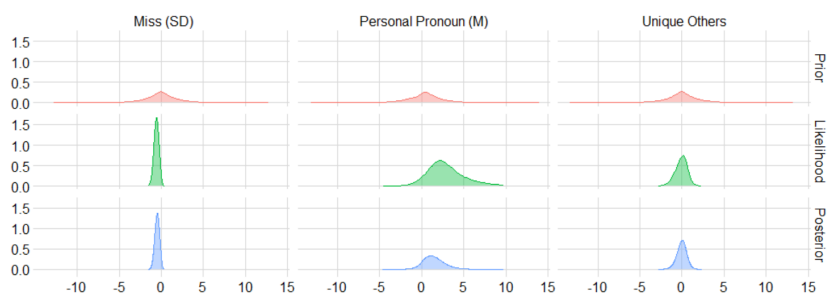


Figure 1: Example of the Bayesian approach using informed (Personal Pronouns) and weakly-informed (Miss, Unique Others) priors and likelihood of the evidence to estimate posterior distributions of three example parameters.

training set. In *Subtask 1*, the models only slightly outperformed the task’s baseline model, but in *Subtask 2*, the models yielded high AUC scores.

4 Discussion

We trained simple classification models, based on psychological features, to determine which individuals may attempt suicide. We used Psychologically-informed and weakly-informed Bayesian models as well as regularized regression models. Our models yielded moderately successful predictions on *Subtask 1*, and considerably better predictions on *Subtask 2* ($0.791 \leq AUC \leq 0.844$, comparable to Cohen’s d of $1.145 - 1.430$). In this task, the informed Bayesian model (M1) was more successful than the weakly-informed (M2). The data-driven regularized regression models (M3) were slightly less accurate in *Subtask 1* than the informed model (M1), and slightly more accurate in *Subtask 2*, perhaps due to the fact that *Subtask 2* included more data than *Subtask 1*.

In addition, in both tasks the Bayesian models (M1, M2) were particularly successful in avoiding False Positive prediction outcomes. Admittedly, in the case of suicide detection, it may be prudent to "err on the side of caution", to avoid missing patients in need of care. However, language-based screening on social media tends to be targeted more for broad risk-detection (Cook et al., 2016). In the case of early risk detection it may also be valid to avoid false alarms in order to reduce unwarranted alarm, especially given the potential for suicidal suggestibility.

Our theory-driven features, as well as the informed Bayesian models, were reliant on domain knowledge to help overcome the problem posed by working with small data sets. Indeed, incorporating knowledge gained from previous research seemed

to have aided in forming a generalized model that did not exhibit over-fitting. Another benefit of this approach lies in model interpretability and in its conduciveness to cumulative scientific discovery. We relied on prior empirical findings, and produced updated empirical priors—in light of the task data—which are simple to interpret and share with others (refer to table 4 for feature importance analysis).

The majority of previous work in suicide prediction was done by using proxies to suicidal behavior such as clinical risk assessment and suicidal ideation, (see Fodeh et al., 2019; Ophir et al., 2020; Coppersmith et al., 2018). Thanks to the CLPsych workshop, and the access to valuable data directly indicative of suicidal behavior, we were able to present similar prediction accuracies on actual suicide attempts. The findings derived from this data show great promise for the use of NLP in suicide prevention.

5 Conclusion

Our current work provides a synthesis between classic scientific and novel data-driven paradigms. Future research is needed to further explore how psychological knowledge and data science methods can be combined to aid in the gradual accumulation of scientific knowledge, and produce actionable predictions that may help save lives.

Ethics Statement

Secure access to the shared task dataset was provided with IRB approval under University of Maryland, College Park protocol 1642625.

Acknowledgements

The authors wish to thank Yhonatan Shemesh, Inon Raz, Mattan S. Ben-Shachar, the CLPsych 2021

| Features | Effect-Size (<i>log – odds</i>) |
|-----------------------------|-----------------------------------|
| Subtask 1 (30 days) | |
| M1 | |
| Negative-Emotion-SD | 2.36 [0.83,4.59] |
| Negative-Emotion-M | -1.68 [-4.05,-0.05] |
| Swear-Words-M | 1.67 [-1.13,6.84] |
| Female-M | 1.06 [0.08,2.64] |
| Want-M | 1.04 [0.29,1.86] |
| M2 | |
| Negative-Emotion-SD | 2.39 [0.88,4.19] |
| Negative-Emotion-M | -1.72 [-3.69,-0.13] |
| Swear-Words-M | 1.53 [-1.24,4.63] |
| Female-M | 1.15 [0.07,2.62] |
| Want-M | 1.04 [0.29,1.88] |
| M3 | |
| They-M | 0.009 |
| I-Pronouns-M | 0.009 |
| Personal-Pronouns-M | 0.009 |
| Want-M | 0.009 |
| Negative-Emotion-SD | 0.008 |
| Subtask 2 (6 months) | |
| M1 | |
| Informal-SD | 2.02 [0.32,4.17] |
| I-Pronouns-M | -1.5 [-2.85,-0.27] |
| Female-M | 1.45 [-0.10,0.4.84] |
| Personal-Pronouns-M | 1.345 [-0.50,3.87] |
| Sexual-M | -1.26 [-2.66,0.09] |
| M2 | |
| Informal-SD | 2.99 [0.13,4.93] |
| Female-M | 2.59 [0.25,5.61] |
| Negative-Emotion-SD | 1.98 [-0.17,4.19] |
| I-Pronouns-M | -1.89 [-3.46,-0.31] |
| Personal-Pronouns-M | 1.87 [-0.80,4.51] |
| M3 | |
| Personal-Pronouns-M | 0.51 |
| Negative-Emotion-SD | 0.11 |

Table 4: Most Important Features based on model coefficient values. Model coefficients are on the log-odds scale. Values in brackets denote 95% posterior uncertainty intervals.

Shared Task organizers, and the anonymous reviewers for their help and insight. The organizers are particularly grateful to the users who donated data to the OurDataHelps project without whom this work would not be possible, to Qntfy for supporting the OurDataHelps project and making the data available, to NORC for creating and administering

the secure infrastructure, and to Amazon for supporting this research with computational resources on AWS.

References

- Ella Arensman, Vanda Scott, Diego De Leo, and Jane Pirkis. 2020. Suicide and suicide prevention from a global perspective. *Crisis*.
- Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: a probabilistic programming language. *Grantee Submission*, 76(1):1–32.
- Cindy K Chung and James W Pennebaker. 2018. What do we know when we liwc a person? text analysis as an assessment tool for traits, personal concerns and life stories. *The Sage handbook of personality and individual differences*, pages 341–360.
- Benjamin L Cook, Ana M Progovac, Pei Chen, Brian Mullin, Sherry Hou, and Enrique Baca-Garcia. 2016. Novel use of natural language processing (nlp) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in madrid. *Computational and mathematical methods in medicine*, 2016.
- Glen Coppersmith, Ryan Leary, Patrick Crutchley, and Alex Fine. 2018. Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, 10:1178222618792860.
- Glen Coppersmith, Kim Ngo, Ryan Leary, and Anthony Wood. 2016. Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the third workshop on computational linguistics and clinical psychology*, pages 106–117.
- Sally C Curtin, Margaret Warner, and Holly Hedegaard. 2016. *Increase in suicide in the United States, 1999–2014*. 2016. US Department of Health and Human Services, Centers for Disease Control and . . .
- Stephanie K Douplik, Brittany Rudd, Timothy Schmutte, Diana Worsley, Cadence F Bowden, Erin McCarthy, Elliott Eggen, Jeffrey A Bridge, and Steven C Marcus. 2020. Association of suicide prevention interventions with subsequent suicide attempts, linkage to follow-up care, and depression symptoms for acute care settings: a systematic review and meta-analysis. *JAMA psychiatry*, 77(10):1021–1030.
- Johannes C Eichstaedt, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preoțiu-Pietro, David A Asch, and H Andrew Schwartz. 2018. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44):11203–11208.

- Samah Fodeh, Taihua Li, Kevin Menczynski, Tedd Burgette, Andrew Harris, Georgeta Ilita, Satyan Rao, Jonathan Gemmell, and Daniela Raicu. 2019. Using machine learning algorithms to detect suicide risk factors on twitter. In *2019 International Conference on Data Mining Workshops (ICDMW)*, pages 941–948. IEEE.
- Joseph C Franklin, Jessica D Ribeiro, Kathryn R Fox, Kate H Bentley, Evan M Kleiman, Xieyining Huang, Katherine M Musacchio, Adam C Jaroszewski, Bernard P Chang, and Matthew K Nock. 2017. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychological bulletin*, 143(2):187.
- Jerome Friedman, Trevor Hastie, and Robert Tibshirani. 2010. [Regularization paths for generalized linear models via coordinate descent](#). *Journal of Statistical Software*, 33(1):1–22.
- Catherine R Glenn, Evan M Kleiman, John Kellerman, Olivia Pollak, Christine B Cha, Erika C Esposito, Andrew C Porter, Peter A Wyman, and Anne E Boatman. 2020. Annual research review: a meta-analytic review of worldwide suicide rates in adolescents. *Journal of child psychology and psychiatry*, 61(3):294–308.
- Ben Goodrich, Jonah Gabry, Imad Ali, and Sam Brilleman. 2020. [rstanarm: Bayesian applied regression modeling via Stan](#). R package version 2.21.1.
- Elizabeth M Haney, Maya E O’Neil, Susan Carson, A Low, K Peterson, LM Denneson, C Oleksiewicz, and D Kansagara. 2012. Suicide risk factors and risk assessment tools: A systematic review.
- Max Kuhn. 2020. [caret: Classification and Regression Training](#). R package version 6.0-86.
- Matthew Large, Cherrie Galletly, Nicholas Myles, Christopher James Ryan, and Hannah Myles. 2017a. Known unknowns and unknown unknowns in suicide risk assessment: evidence from meta-analyses of aleatory and epistemic uncertainty. *BJPsych bulletin*, 41(3):160–163.
- Matthew Michael Large, Christopher James Ryan, Gregory Carter, and Nav Kapur. 2017b. Can we usefully stratify patients according to suicide risk? *Bmj*, 359.
- Nicholas Leigh-Hunt, David Bagguley, Kristin Bash, Victoria Turner, Stephen Turnbull, N Valtorta, and Woody Caan. 2017. An overview of systematic reviews on the public health consequences of social isolation and loneliness. *Public health*, 152:157–171.
- Richard T Liu, Stephanie J Steele, Jessica L Hamilton, Quyen BP Do, Kayla Furbish, Taylor A Burke, Ashley P Martinez, and Nimesha Gerlus. 2020. Sleep and suicide: A systematic review and meta-analysis of longitudinal studies. *Clinical psychology review*, page 101895.
- Andrea Miranda-Mendizabal, Pere Castellví, Oleguer Parés-Badell, Itxaso Alayo, José Almenara, Iciar Alonso, Maria Jesús Blasco, Annabel Cebria, Andrea Gabilondo, Margalida Gili, et al. 2019. Gender differences in suicidal behavior in adolescents and young adults: systematic review and meta-analysis of longitudinal studies. *International journal of public health*, 64(2):265–283.
- Mohsen Naghavi. 2019. Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the global burden of disease study 2016. *bmj*, 364.
- OSF Open Science Collaboration et al. 2015. Estimating the reproducibility of psychological science. *Science*, 349(6251).
- Yaakov Ophir, Refael Tikochinski, Christa SC Asterhan, Itay Sisso, and Roi Reichart. 2020. Deep neural networks detect suicide risk from textual facebook posts. *Scientific reports*, 10(1):1–10.
- James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report.
- John P Pestian, Pawel Matykiewicz, Michelle Linn-Gust, Brett South, Ozlem Uzuner, Jan Wiebe, K Bretonnel Cohen, John Hurdle, and Christopher Brew. 2012. Sentiment analysis of suicide notes: A shared task. *Biomedical informatics insights*, 5:BII–S9042.
- Jessica D Ribeiro, Xieyining Huang, Kathryn R Fox, and Joseph C Franklin. 2018. Depression and hopelessness as risk factors for suicide ideation, attempts and death: meta-analysis of longitudinal studies. *The British Journal of Psychiatry*, 212(5):279–286.
- Shane Schweitzer and Adam Waytz. 2020. Language as a window into mind perception: How mental state language differentiates body and mind, human and nonhuman, and the self from others. *Journal of Experimental Psychology: General*.
- Almog Simchon and Michael Gilead. 2018. [A psychologically informed approach to CLPsych shared task 2018](#). In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 113–118, New Orleans, LA. Association for Computational Linguistics.
- Shannon Wiltsey Stirman and James W Pennebaker. 2001. Word use in the poetry of suicidal and non-suicidal poets. *Psychosomatic medicine*, 63(4):517–522.
- Erik van Zwet and Andrew Gelman. 2020. A proposal for informative default priors scaled by the standard error of estimates. *arXiv preprint arXiv:2011.15037*.
- Kathryn Yates, Ulla Lång, Martin Cederlöf, Fiona Boland, Peter Taylor, Mary Cannon, Fiona McNicholas, Jordan DeVlyder, and Ian Kelleher. 2019.

Association of psychotic experiences with subsequent risk of suicidal ideation, suicide attempts, and suicide deaths: a systematic review and meta-analysis of longitudinal population studies. *JAMA psychiatry*, 76(2):180–189.

(?)