Retention and Relapse in Gambling Self-Help Communities on Reddit

Niklas Hopfgartner*[0000−0002−4672−7956], Thorsten Ruprechter*[0000−0002−7552−6256], and Denis Helic[0000−0003−0725−7450]

Graz University of Technology, Graz, Austria
{n.hopfgartner,ruprechter,dhelic}@tugraz.at

Abstract. Problem gambling endangers the mental health and financial stability of those affected. Individuals who are trying to overcome this problematic behavior often seek support in self-help groups, whether offline or online. Although online gambling self-help groups have become increasingly prominent in recent years, research is still needed to fully understand the social interactions and effectiveness of this medium. In this paper, we analyze behavior of problem gamblers in two gambling self-help communities on Reddit. In particular, following social impact theory, we quantify the strength, frequency, and promptness of users' social interactions and measure their effects on problem gamblers' retention in the community, as well as on their relapse to gambling. Firstly, we find that the magnitude of the community response, and a higher response from users who currently sustain prolonged gambling-free periods are positively associated with retention in the community. Secondly, our relapse analysis indicates that users who engage more strongly in discussions following their own submissions have a lower risk of gambling relapse. We believe that our findings provide useful insights on how to improve support for problem gamblers in online self-help communities.

Keywords: Social impact · Problem gambling · Online self-help groups.

1 Introduction

Online and land-based gambling are wide spread forms of recreation and safe for most individuals [6,22]. Nevertheless, a small but significant number of individuals suffer from severe negative consequences of gambling, including financial problems or psychological distress induced by problematic gambling behavior [10,11]. Therefore, preventing and curbing problematic gambling behavior has become a major concern for clinicians, researchers, and policy-makers leading, among others, to the development of so-called responsible gambling tools. Such tools enable gamblers to control and limit their gambling behavior through money and time spending limits, mandatory cooling-off periods, or voluntary self-exclusions [1,2,14,21]. However, money and time limits can be changed after a certain period, and gamblers might circumvent platform-specific self-exclusions

* Both authors contributed equally.

by registering on other gambling platforms. For this reason, problem gamblers usually need additional support during the critical phase of their rehab to help them prevent gambling relapses.

One widely available and cost-effective support measure are self-help groups (e.g., "Gamblers Anonymous"), which provide supportive environments for problem gamblers to receive help and advice for personal problems [27,30]. Besides traditional offline self-help groups, recent studies have also shown that brief motivational treatments via telephone [12,16] as well as online forums and online self-help groups are effective in supporting problem gamblers [13].

Online self-help groups also exist on Reddit, a community-driven discussion platform [29] and one of the most popular websites on the Internet[1]. Reddit aggregates a large amount of so-called *subreddits* (i.e., communities and discussion forums), in which registered users can submit text or media posts (*submissions*) for others to comment on [5]. Due to its anonymity, Reddit allows users to discuss negative feelings as well as personal experiences more openly with like-minded users. However, to successfully provide peer-to-peer support through such an online platform, users need to actively engage with each other [28]. Similar as in offline self-help groups, a person who seeks support (*support seeker*) needs to interact with a person who is willing to provide support (*supporter*).

Research questions. In this paper, we focus on self-help subreddits related to problem gambling and gambling addiction. Given Reddit's unique characteristics, we address the following research questions in this paper:

- **RQ1:** What factors affect support seeker retention in online self-help groups?
- **RQ2:** What factors affect the risk of relapse for support seekers?

Approach. We address these research questions using a theory-driven approach and derive a set of indicators based on social impact theory [18] to quantify the interaction between support seekers and supporters on Reddit. Specifically, we operationalize (i) supporter strength by the percentage of positive comments and the percentage of comments from players currently in rehab that support seekers' submissions receive, (ii) the social immediacy by the percentage of comments they receive shortly after they post their submissions, and (iii) the interaction magnitude as the number of supporters commenting on their submissions. Furthermore, we apply those indicators to estimate the risk of relapse for individual support seekers using a well-known Cox proportional hazards model.

Results. We find that the number of comments on the support seeker's first submission as well as the percentage of comments by supporters who themselves are currently in rehab have a substantial positive effect on the seeker's retention. For example, each additional comment to the seeker's first submission increases the odds of making another submission by more than 10%. Similarly, a one percent increase of comments made by supporters currently in rehab is associated with almost one percent increase in the odds of seekers making another submission. Further, our results on relapse indicate that user's self-motivation together with positive response from supporters are associated with reduced relapse hazard.

[1] https://www.similarweb.com/top-websites/

Contributions. As opposed to the study of emotions and sentiment in online gambling communities [17], our work is, to the best of our knowledge, the first to investigate *social interactions* in self-help groups for gambling problems on Reddit. Contextualizing our work within the framework of social impact theory enables us to focus on a specific set of variables to study social interactions in online self-help groups and, in particular, in online gambling self-help groups. We believe that our findings provide actionable insights on how to steer online communities to better support problem gamblers in their efforts to curb their problematic gambling behavior, and hence improve their quality of life. For example, problem gambling subreddits may introduce bots or guidelines that could—based on our results—encourage support seekers to keep reporting their process and supporters to comment on submissions of support seekers. Lastly, we make the Python and R scripts used for this work available online[2].

2 Materials and Methods

2.1 Data

Reddit communities. Reddit covers a wide variety of topics across hundreds of thousands of active subreddits, where users can consume and contribute content based on their interests [31]. For example, a user interested in video games and American Football might engage with the subreddits */r/gaming* and */r/nfl* (note that the */r/* is a naming convention for communities on Reddit—similar to @ preceding account names on Twitter). Besides these examples of casual discussion topics, some users also use Reddit to get help in challenging times, such as emotional support during the COVID-19 crisis [4,19], community aid for drug addiction, rehab and withdrawal [3,20], or advice for problem gambling [17].
Problem gambling subreddits. For problem gambling, there are several subreddits that offer support for affected individuals. In this work, we focus on */r/problemgambling* and */r/GamblingAddiction*, as those are the most prolific subreddits covering this topic. These subreddits act as online self-help groups where support seekers disclose their negative gambling experiences, ask other users for support, or keep a diary about their journey to becoming gambling-free in diary-like submissions. Such diary-like submissions usually include a brief title indicating how many days support seekers have been gambling-free (e.g., "Day 0", "Day 13", "Week 3"), and a text describing their general feelings and state of mind. Figure 1 gives an example of such a submission on */r/problemgambling*. In this submission, the support seeker (blue) admits to a relapse, while additionally expressing negative feelings about it. Afterwards, a supporter (gold) adds a comment to the submission, encouraging the support seeker to keep going and that this setback "does not stop the momentum" they "have created in this fight". The supporter also has a so-called *flair* next to their name, showing that the supporter is currently gambling-free for 13 days. Finally, the support seeker greatly appreciates the comment by responding "your words mean more than you

[2] https://gitlab.tugraz.at/3B57777CAD304A73/reddit-gambling

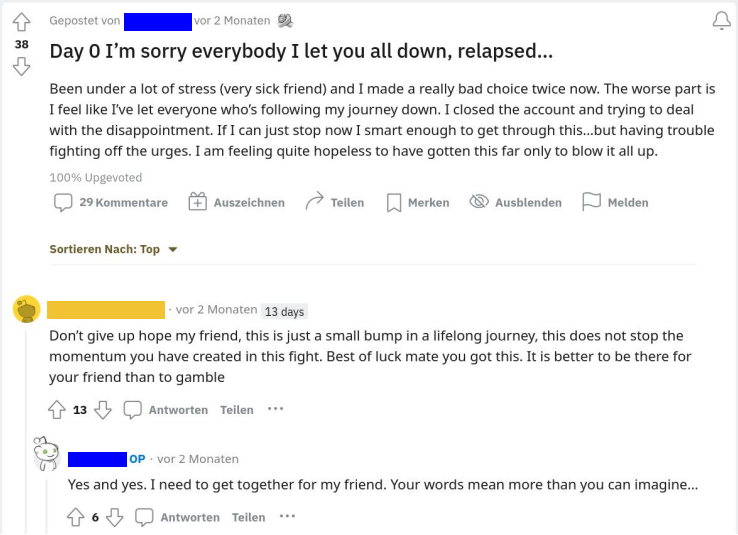


Fig. 1: **Diary-like submission on */r/problemgambling*.** A submission on */r/problemgambling*, in which a support seeker (blue) elaborates that they relapsed and have to restart their gambling-free journey, as indicated by "Day 0" in the title. A supporter (gold)—also themselves a problem gambler currently in recovery, as indicated by their gambling-free *flair* ("13 days")—consoles the support seeker via a comment, which the support seeker appreciates by responding positively to that comment.

can imagine" to the supporter. Overall, this exchange illustrates the dynamics of these problem gambling subreddits, and exemplifies the characteristics of such diary-like submissions that mimic traditional offline self-help groups.

Data collection & preprocessing. We use the Pushshift API [5] to retrieve all submissions and comments for */r/problemgambling* and */r/GamblingAddiction* from January 2014 up until December 2020. Afterwards, we extract all submissions resembling diary-like entries that mention temporal progress (e.g., "Day 0", "Week 3", "Month 5") through a regular expression and ignore submissions by accounts that have been deleted since then. For the extracted submissions, we collect the author's username, contribution timestamp, and title. Furthermore, for all comments to these submissions we collect the user name and flair, timestamp, as well as the text of the comment.

From the collected data, we then compute the following features for each submission: (i) the number of comments (both by the submission seeker as well as supporters), (ii) the percentage of supporters having a flair, (iii) the percentage of comments within one hour of the submission, and (iv) the percentage of positive comments. We calculate sentiment scores (between -1 and 1) of raw comment texts using VADER, and utilize a threshold of greater than 0.05 (less than -0.05) to label comments as positive (negative) [15].

In this work, we only consider submissions by support seekers for which the day they report in their first submission equals "Day 0" or "Day 1". We treat both 0 and 1 as valid first submissions, because that is how support seekers typically document their first day in rehab on the problem gambling subreddits. With this, we ensure our dataset does not include left-censored data (i.e., support seekers starting to report days after the start of their rehab) [9]. Overall, our pre-processed dataset contains 958 support seekers with an average (median) of 5.53 (1) submissions, 5 298 total submissions with an average (median) of 2.36 (1) comments, and 12 506 comments in total.

2.2 Social Impact Indicators on Reddit

In this work, we propose a framework to assess the social impact of the Reddit community on problem gamblers. Social impact (I) is driven by three forces [18]:

$$I \sim S \times i \times N, \tag{1}$$

where S is the strength or power of individuals on the target, i the immediacy or proximity of individuals to the target, and N the number of individuals. We quantify the social impact of the Reddit community by analyzing commenting behavior on the first submission of support seekers signaling the start of their rehab. We operationalize the social impact I by measuring whether support seekers make a second submission in the corresponding problem gambling subreddit (RQ1), and whether they relapse after a certain period of time (RQ2). Furthermore, we quantify the three forces S, i, and N by: the percentage of supporters with a flair and the percentage of positive comments by supporters (S), the percentage of comments within the first hour (i), and the total number of comments on (N) the first submission.

2.3 Retention Analysis

Retention dataset. To assess the effects of the Reddit community on support seeker retention in online self-help groups, we analyze their first submissions and determine whether they contributed at least another submission within a 30-day observation period. As our measures of social impact include percentage calculations based on the number of received comments (e.g., the percentage of comments by supporters within the first hour), we only analyze submissions with at least one comment. In addition, we remove the top 5% of submissions with the highest number of comments to reduce the impact of outliers. Therefore, we consider only posts with less than 10 comments. We refer to this as the *retention dataset*, which we utilize to predict whether a support seeker will create another submission solely based on social impact features of their first submission. The retention dataset contains the first submission of 782 support seekers, with an average (median) of 3.6 (3) comments per submission.

Approach. We fit a logistic regression to measure the social impact of the Reddit community on the retention of support seekers. A binary variable indicating

whether a support seeker makes at least one other submission within 30 days after their first submission forms the dependent variable, and our measures for social impact form the independent variables. Specifically, we fit a logistic regression modeling the probability of the binary retention variable r for user i as $Pr(r_i = 1) = logit^{-1}(\boldsymbol{X}_i\boldsymbol{\beta})$, where $\boldsymbol{X}_i$ is the feature (predictor) vector of user i and $\boldsymbol{\beta}$ is the vector of coefficients. The logistic model for r is then:

$$r \sim n + f + p + h, \tag{2}$$

where n is the number of comments to the support seeker's first submission, f the percentage of comments for the first submission coming from supporters with the gambling-free flair, p the percentage of positive comments for the first submission, and h the percentage of comments posted in one hour after the first submission of a given support seeker.

The coefficients $\boldsymbol{\beta}$ can be interpreted as follows. In logistic regression, logarithm of the odds for retention of user i equals the linear predictor, i.e.,

$$log\left(\frac{Pr(r_i = 1)}{Pr(r_i = 0)}\right) = \boldsymbol{X}_i\boldsymbol{\beta}, \tag{3}$$

and hence increasing the j-th predictor (i.e., X_{ij}) by 1 has the effect of adding the corresponding β_j coefficient to both sizes. Exponentiating both sides multiplies the odds with a factor of e^{β_j}, which is interpreted as the multiplicative factor for the odds corresponding to a unit difference in the j-th predictor.

2.4 Relapse Analysis

Relapse dataset. To analyze the relapse risk, we begin by aggregating first submissions of relapsed support seekers. For this, we only consider support seekers that contributed at least one other submission within 60 days after their first submission. We then detect whether they relapse or not by examining their self-reported days since the start of their rehab. Specifically, we label an individual as relapsed if they first create one or multiple consecutive diary entries documenting their gambling-free journey (e.g., "Day 0", "Day 5", etc.), before then self-reporting a relapse by creating a post that is again titled "Day 0" (or "Day 1"). If a support seeker relapses, we also record the number of days since their first submission. We refer to this dataset as the *relapse dataset*, as we use it to investigate the factors leading to relapse during gambling addiction rehab. The relapse dataset contains 338 first submissions by users that posted at least twice, with an average (median) of 5.33 (4) comments per submission.

Right-Censored Datasets. Note that through our preprocessing, the retention dataset as well as the relapse dataset are right-censored. A dataset is considered right-censored when records about the investigated event—in our case, the self-reported relapse in the relapse dataset or the second submission in the participation dataset—exist only for parts of the support seekers within the respective observation period [9]. For example, considering the relapse dataset, right-censoring may occur due to three reasons. First, support seekers might not

have relapsed by the time the observation period ended and thus no information about this event might exist. Secondly, they could have experienced a relapse during the observation period but might have not reported it. Lastly, a support seeker may have experienced a different event which did not allow them to complete the full observation period (e.g., in the most extreme case, death). Overall, out of our 338 support seekers in the relapse dataset 117 experienced a relapse event during the 60-day observation period.

Approach. The analysis whether (or how long) an individual "survives" an observation period without encountering an event has been termed *survival analysis* in past research [7,8,20,26]. Survival analysis is especially convenient for right-censored data, such as our relapse dataset. In the following, we briefly touch on the aspects and metrics most relevant to survival analysis as performed in this work: The Kaplan–Meier survival estimate as well as the Cox proportional hazards model.

Kaplan–Meier survival estimate. To estimate the probability of not relapsing, we use a so-called survival estimate. In particular, the probability that individuals survive to a given time in an observation period can be estimated using the Kaplan–Meier (KM) survival estimate [9]. The KM estimate computes the cumulative survival probability S for each point in time as following:

$$S(t_j) = S(t_{j-1})(1 - \frac{d_j}{n_j}) \qquad S(0) = 1, \quad t_0 = 0 \tag{4}$$

Equation 4 recursively computes a step survival function S over time, where value of S changes at each event trigger t_j. Events typically refer to "deaths" (i.e., support seeker's relapse). Computation of S in the KM estimate takes into account the number of "alive" individuals (n_j) at t_j as well as the number of events or deaths d_j happening at t_j. In this way, both censored and uncensored individuals contribute to estimating S by providing information for the time they were not exposed to an event (i.e., in our case, relapse). Overall, to estimate the probability of survival at any point during the observation period, the KM estimate considers, for each individual, the last reported day and whether the event was observed up to that day.

Cox proportional hazards model. To estimate the effect of social impact features on Redditor's relapse into gambling, we utilize a semi-parametric Cox proportional hazards model ("Cox regression"). A Cox regression is a regression model that estimates the probability of an event occurring given a set of features x_i via a hazard function h [7]. This hazard h is derived from the nonparametrically estimated survival probability, i.e., from the KM estimator [9]:

$$h(t) = -\frac{d}{dt}[logS(t)] \tag{5}$$

Using this hazard function, we can mathematically express a Cox regression:

$$h(t) = h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p) \tag{6}$$

In Equation 6, the hazard function $h(t)$ describes the probability that an observed individual (i.e., a recovering gambling addict) observes the event (i.e.,

a relapse) at a given point in time t. Correspondingly, the estimated baseline hazard $h_0(t)$ equals the hazard if all features (x_i) are 0, similar to the intercept in a regular multivariate regression setup but varying with time t. All features x_i are accompanied by a respective coefficient β_i, which describes the effect of the feature x_i on the hazard. Coefficients are usually interpreted in their exponential form e^{β_i}, which is termed "hazard ratio". A hazard ratio (HR) greater than 1 signals that as the value of a variable increases, the hazard also increases and thus the estimated length for survival decreases (i.e., the time to relapse shortens). Consequently, a HR smaller than 1 corresponds to a decrease in hazard and therefore an increase in survival time when the associated variable increases, while a HR of exactly 1 means that the variable has no effect on the hazard. Altogether, a Cox regression can be understood as a multivariate linear regression with the logarithm of the hazard function as the dependent variable and the features x_i as the independent variables, while HR measures the effect sizes of the corresponding x_i.

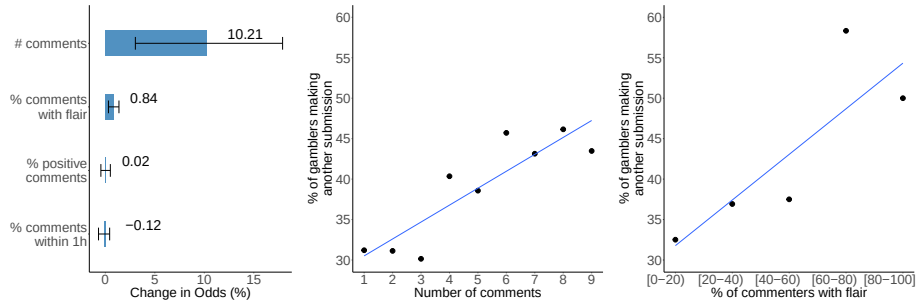
Using this survival analysis framework, we fit multiple Cox regressions on support seekers' first submissions in the relapse dataset, utilizing the same social impact features as we did in the retention analysis: the percentage of comments by supporters with flair, percentage of comments by supporters with positive sentiment, percentage of comments within one hour after the submission, and total number of comments.

3 Results and Discussion

3.1 Retention

In Figure 2a we show the exponentiated coefficients reduced by 1, i.e., we show $100(e^{\beta_j} - 1)$, which we interpret as percentage change in the odds corresponding to the j-th predictor. For each coefficient we also show 95% confidence intervals. We can then interpret coefficients with confidence intervals not crossing zero either from above or from below as having a significant positive or negative effect, respectively, on support seeker retention. In our data, we observe that the number of comments a support seeker receives on their first submission, as well as the percentage of comments from supporters with gambling-free flair, both have a significant positive effect on the odds of support seekers making at least one further submission. Numerically, this means that each additional comment increases the odds of a support seeker making a second submission by 10.21%. Similarly, we observe that when the percentage of comments from supporters with flair increases by one (ten) percents, the odds of making a second submission increase by approximately 0.84% (8.4%). Both the percentage of positive supporter comments and the percentage of comments within the first hour are non-significantly associated with the outcome variable as their confidence intervals intersect with zero. The model achieves a Nagelkerke R^2 of 0.033 [24].

In Figures 2b and 2c we also show the relationship between the percentage of support seekers making another submission and the number of comments, as well as the percentage of supporters commenting with a flair. In both figures,



(a) Change in odds in percent with 95% confidence intervals for the predictors of the logistic regression.

(b) Percentage of support seekers making another submission vs. the number of comments.

(c) Percentage of support seekers making another submission vs. the percentage of supporters with a flair.

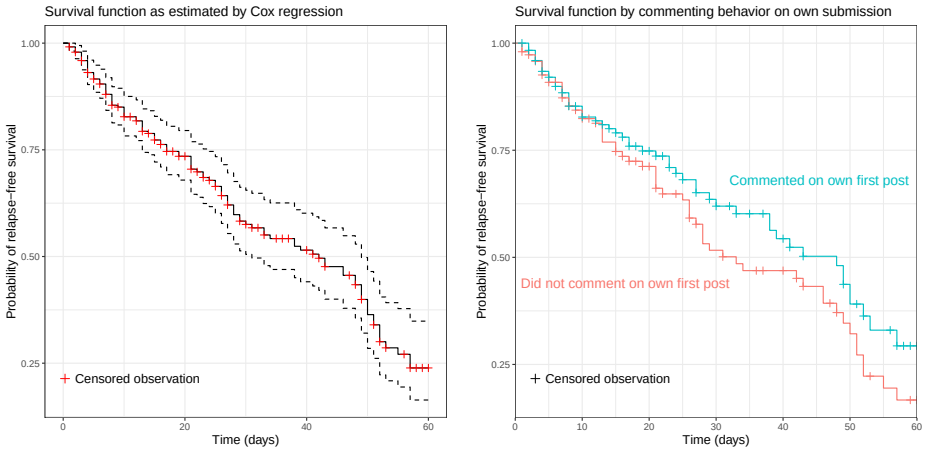
Fig. 2: **Retention analysis using logistic regression.** Figure (2a) shows the change in odds of the independent variables, where a one-unit change in the predictor variable translates into a corresponding percentage change in the odds of making another submission. Both the number of comments, as well as the percentage of supporters with flair have a significant positive effect on the odds of making another submission. Figures (2b) and (2c) highlight the positive correlation between the two significant predictor variables and the percentage of gamblers making another submission.

we observe that the more comments support seekers receive, and the higher the percentage of supporters commenting with a flair, the higher the percentage of support seekers making another submission. Altogether, our analysis indicates that two of our conceived social impact features (i.e., number of comments and presence of supporters with flair) have significant effects on support seeker retention, suggesting the relevance of such features.

3.2 Relapse

First, we compute the *KM* estimate for the cumulative survival function, which we plot in Figure 3a. We observe a monotonically decreasing survival probability, with slightly higher relapse probabilities in the first 10 days. The relapse probability has another peak around day 50. Using this estimate, we proceed to fitting several Cox regressions. To that end, we perform multiple univariate Cox regressions using each social impact feature individually before then fitting a multivariate regression using all features.

Although the univariate models generally signal a reduction of the hazard by the social impact features, none of the employed features affect survival significantly ($p > 0.1$, cf. first column in Table 1). Most notably, the *HR*s for percentage of positive supporter comments and the percentage of supporter comments with flairs (i.e., also currently recovering gamblers) indicate that interacting with individuals who are going through the same recovery process as well as receiving



(a) *KM* survival function and 95% confidence intervals for relapse dataset. We observe slightly higher risk for relapse in the first 10 days and around day 50.

(b) Separate *KM* estimates for users commenting/not commenting on their own post. Users engaging with their own posts have a higher chance of survival.

Fig. 3: **Relapse analysis using *KM* survival function.** In Figure (3a), the plot for the KM survival curve computed for the relapse dataset depicts a regularly decreasing survival function, with slightly higher risk for relapse in the first 10 days as well as high risks of relapse around day 50. In Figure (3b), the KM estimate is shown for two groups of individuals: support seekers who commented on their own first submissions (cyan curve) and those who did not (red curve). The survival curve hints towards support seekers who show self-motivation to engage with the community being more successful in surviving until the end of the observation period (logrank test, $p = 0.1$).

positive encouragement prolong survival times for support seekers (HR of 0.6817 and 0.8329, resp.). Next, we fit a Cox regression combining all of our social impact features. This multivariate model yields similar insignificant HRs ($p > 0.3$) for all features and does not numerically differ considerably from the univariate models (cf. second column in Table 1). Altogether, our investigation of the influence of social impact on support seeker survivability does not yield significant results, although a trend in reduction of the hazard is observable for some of the social impact indicators.

Effect of self-motivation on survival. Due to the non-significant effect of social impact features on survival, we now extend our analysis to also include self-motivation. For this, we categorize all support seekers in the relapse dataset (338 users) into two groups signaling whether they themselves commented on their first submission or not (189 and 149 users, resp.). We consider this to be an indicator of (self-)motivation to actively engage with others in an online self-help group. The KM survival curves for the corresponding two groups (Fig. 3b) hint

Table 1: **Cox regression results.** The hazard ratio (HR) computed from the coefficients (β_i) as well as their 95% confidence interval (CI) and p-values (Wald test) for both univariate and multivariate analysis of social impact features.

	Univariate analysis			Multivariate analysis		
Covariate	**HR[**$\exp(\beta_i)$**]**	**95% CI**	**p-value**	**HR[**$\exp(\beta_i)$**]**	**95% CI**	**p-value**
#comments	0.971	(0.923–1.018)	0.219	0.992	(0.939–1.048)	0.777
% positive comments	0.682	(0.401–1.159)	0.157	0.736	(0.407–1.331)	0.310
% comments with flair	0.833	(0.411–1.687)	0.612	1.040	(0.491–2.205)	0.918
% comments within 1h	0.959	(0.733–1.124)	0.159	0.970	(0.907–1.037)	0.369

toward support seekers that are motivated to comment on their own submission having slightly higher chance of survival (logrank test, $p = 0.1$). Moreover, we investigate the difference in effects between the two groups for our social impact features using Cox regression. While we find no significantly different effects for the number of comments (HR= 1.126, $p = 0.125$), percentage of comments with flair (HR= 0.947, $p = 0.22$), or percentage of comments within one hour after the submission (HR= 0.072, $p = 0.51$), we observe a significant difference for the percentage of positive comments (HR= 0.286, $p < 0.05$). This suggests that support-seekers that exhibit self-motivation in commenting on their own posts benefit considerably more from receiving positive encouragement than those that do not. Overall, future work attempting relapse analysis in online support groups may also consider such characteristics of self-motivation as relevant factors.

3.3 Limitations

There are several limitations of our analysis. Firstly, Reddit only covers a portion of the population affected by problematic gambling behavior. Most Reddit users are younger than 40, male, and from the United States (75.2%, 62.8%, 50% of all users, respectively[3]). Thus, we can assume that other demographics are underrepresented in our dataset. Secondly, our investigated sample may be even more biased as we strictly focus on problem gamblers that self-select to actually report on their problems online. Therefore, gamblers without the motivation to actively engage in online discussions of their problems with others are also underrepresented in our sample. Thirdly, although we took care about controlling for several factors such as left- and right-censored data or self-motivation in engaging with the community, we still mainly reported on correlations and predictive effects (as opposed to causal effects) of social impact indicators on retention and relapse of gamblers. We leave the investigation of causality as a promising avenue for future research. On a separate note, Latane [18] mentions mulitple limitations of social impact theory. For example, the initial theory does not consider people as active seekers of social impact. In addition, the model does not allow for dynamic components. This is also prevalent in our analysis,

[3] From: statista.com/statistics/1125159/reddit-us-app-users-age, statista.com/statistics/1255182/distribution-of-users-on-reddit-worldwide-gender, statista.com/statistics/325144/reddit-global-active-user-distribution

as we mostly consider features of user's first posts. Future work could incorporate information from user's later posts to gain further insights about the social impact Reddit's online self-help groups have on problem gamblers.

4 Conclusions

In this work, we propose a framework on how to measure the social impact of the Reddit community on problem gamblers who seek support online. The framework is based on social impact theory and aims to operationalize its factors to analyze social interactions between users in self-help gambling communities on Reddit.

Our results suggest that a higher number of comments from supporters, as well as a higher percentage of supporter comments with gambling-free flair increase the retention of support seekers in online self-help groups on Reddit. We additionally analyze the effect of social impact on relapses of recovering problem gamblers using survival analysis, but obtain mostly statistically non-significant results. The only statistically significant exception is related to support seekers who engage in the discussion with others, e.g., by commenting on their own first submission. For those users, the percentage of positive comments from others has a positive effect reducing the probability of their relapse and indicating that a positive encouragement for motivated support seekers is beneficial.

There are multiple possibilities for extending the work presented in this paper. Firstly, user activity in subreddits other than */r/problemgambling* and */r/GamblingAddiction* could be taken into account. It might be possible that behavior of recovering addicts as indicated by their Reddit activity is a predictor of their survival probability. For example, if users find an alternative outlet or substitution for their addictive behavior (e.g., video games or sports), they might have higher chances of successfully recovering from their gambling addiction. Secondly, besides parsimonious sentiment measurements it might prove feasible to include other indicators of emotion, such as valence, arousal, and dominance [23,25,32]. Furthermore, incorporating word embeddings or other similar text features into such analysis might prove fruitful as well [20].

Overall, our results may help in providing additional support to potential problem gamblers in their strive to change their gambling behavior. For example, senior users and users with a gambling-free flair should be encouraged—either through subreddit rules or by bots—to support newcomers (support seekers) when they post their very first submission. Moreover, we believe that our approach to analyzing social interactions in gambling self-help groups on Reddit through the lenses of social impact theory can be useful for future studies of user interactions in the context of online self-help communities.

Acknowledgements. We thank David Garcia for his helpful feedback and discussion on this topic. The authors also thank the anonymous reviewers for their comments, which improved this work.

References

1. Auer, M., Hopfgartner, N., Griffiths, M.D.: An empirical study of the effect of voluntary limit-setting on gamblers' loyalty using behavioural tracking data. International Journal of Mental Health and Addiction pp. 1–12 (2019)
2. Auer, M., Hopfgartner, N., Griffiths, M.D.: The effects of voluntary deposit limit-setting on long-term online gambling expenditure. Cyberpsychology, Behavior, and Social Networking **23**(2), 113–118 (2020)
3. Balsamo, D., Bajardi, P., Panisson, A.: Firsthand opiates abuse on social media: monitoring geospatial patterns of interest through a digital cohort. In: The World Wide Web Conference. pp. 2572–2579 (2019)
4. Basile, V., Cauteruccio, F., Terracina, G.: How dramatic events can affect emotionality in social posting: The impact of covid-19 on reddit. Future Internet **13**(2), 29 (2021)
5. Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., Blackburn, J.: The pushshift reddit dataset. In: Proceedings of the international AAAI conference on web and social media. vol. 14, pp. 830–839 (2020)
6. Bernhard, B.J., Dickens, D.R., Shapiro, P.D.: Gambling alone? A study of solitary and social gambling in america. UNLV Gaming Research & Review Journal **11**(2), 1 (2007)
7. Bradburn, M.J., Clark, T.G., Love, S.B., Altman, D.G.: Survival Analysis Part III: Multivariate data analysis - Choosing a model and assessing its adequacy and fit. British Journal of Cancer **89**(4), 605–611 (2003). https://doi.org/10.1038/sj.bjc.6601120
8. Clark, T.G., Altman, D.G.: Developing a prognostic model in the presence of missing data: an ovarian cancer case study. Journal of Clinical Epidemiology **56**(1), 28–37 (2003)
9. Clark, T.G., Bradburn, M.J., Love, S.B., Altman, D.G.: Survival analysis part i: basic concepts and first analyses. British journal of cancer **89**(2), 232–238 (2003)
10. Díaz, A., Pérez, L.: Online gambling-related harm: findings from the study on the prevalence, behavior and characteristics of gamblers in spain. Journal of Gambling Studies **37**(2), 599–607 (2021)
11. Dowling, N.A., Suomi, A., Jackson, A.C., Lavis, T.: Problem gambling family impacts: development of the problem gambling family impact scale. Journal of gambling studies **32**(3), 935–955 (2016)
12. Erevik, E.K., Pallesen, S., Mohn, M., Aspeland, T., Vedaa, Ø., Torsheim, T.: The norwegian remote intervention programme for problem gambling: Short-and long-term outcomes. Nordic Studies on Alcohol and Drugs **37**(4), 365–383 (2020)
13. Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., Stern, A.: Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. Bmj **328**(7449), 1166 (2004)
14. Hopfgartner, N., Auer, M., Santos, T., Helic, D., Griffiths, M.D.: The effect of mandatory play breaks on subsequent gambling behavior among norwegian online sports betting, slots and bingo players: A large-scale real world study. Journal of Gambling Studies pp. 1–16 (2021)
15. Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the international AAAI conference on web and social media. vol. 8, pp. 216–225 (2014)
16. Jonsson, J., Hodgins, D.C., Munck, I., Carlbring, P.: Reaching out to big losers leads to sustained reductions in gambling over 1 year: A randomized controlled trial of brief motivational contact. Addiction **115**(8), 1522–1531 (2020)

17. Kaakinen, M., Oksanen, A., Sirola, A., Savolainen, I., Garcia, D.: Emotions in online gambling communities: A multilevel sentiment analysis. In: International Conference on Human-Computer Interaction. pp. 542–550. Springer (2020)
18. Latané, B.: The psychology of social impact. American psychologist **36**(4), 343 (1981)
19. Low, D.M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., Ghosh, S.S.: Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. Journal of medical Internet research **22**(10), e22635 (2020)
20. Lu, J., Sridhar, S., Pandey, R., Hasan, M.A., Mohler, G.: Investigate transitions into drug addiction through text mining of reddit data. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2367–2375 (2019)
21. Luquiens, A., Dugravot, A., Panjo, H., Benyamina, A., Gaïffas, S., Bacry, E.: Self-exclusion among online poker gamblers: effects on expenditure in time and money as compared to matched controls. International Journal of Environmental Research and Public Health **16**(22), 4399 (2019)
22. McNeilly, D.P., Burke, W.J.: Gambling as a social activity of older adults. The International Journal of Aging and Human Development **52**(1), 19–28 (2001)
23. Mohammad, S.M.: Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers) **1**, 174–184 (2018). https://doi.org/10.18653/v1/p18-1017
24. Nagelkerke, N.J., et al.: A note on a general definition of the coefficient of determination. Biometrika **78**(3), 691–692 (1991)
25. Pellert, M., Schweighofer, S., Garcia, D.: The individual dynamics of affective expression on social media. EPJ Data Science **9**(1), 1–14 (2020), http://dx.doi.org/10.1140/epjds/s13688-019-0219-3
26. Raban, D.R., Moldovan, M., Jones, Q.: An empirical study of critical mass and online community survival. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work. pp. 71–80 (2010)
27. Schuler, A., Ferentzy, P., Turner, N.E., Skinner, W., McIsaac, K.E., Ziegler, C.P., Matheson, F.I.: Gamblers anonymous as a recovery pathway: A scoping review. Journal of gambling studies **32**(4), 1261–1278 (2016)
28. Sharma, A., Choudhury, M., Althoff, T., Sharma, A.: Engagement patterns of peer-to-peer interactions on mental health platforms. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 14, pp. 614–625 (2020)
29. Singer, P., Flöck, F., Meinhart, C., Zeitfogel, E., Strohmaier, M.: Evolution of reddit: from the front page of the internet to a self-referential community? In: Proceedings of the 23rd international conference on world wide web. pp. 517–522 (2014)
30. Syvertsen, A., Erevik, E.K., Mentzoni, R.A., Pallesen, S.: Gambling addiction norway–experiences among members of a norwegian self-help group for problem gambling. International Gambling Studies **20**(2), 246–261 (2020)
31. Tan, C.: Tracing community genealogy: how new communities emerge from the old. In: Twelfth International AAAI Conference on Web and Social Media (2018)
32. Warriner, A.B., Kuperman, V., Brysbaert, M.: Norms of valence, arousal, and dominance for 13,915 english lemmas. Behavior research methods **45**(4), 1191–1207 (2013)