

Reddit Rules! Characterizing an Ecosystem of Governance

Casey Fiesler, Jialun “Aaron” Jiang, Joshua McCann, Kyle Frye, Jed R. Brubaker

Department of Information Science, University of Colorado Boulder
{casey.fiesler, aaron.jiang, joshua.mccann, kyle.frye, jed.brubaker}@colorado.edu

Abstract

The social sharing and news aggregation site Reddit provides a unique example of an ecosystem of community-created rules. Not only do individual subreddits create and enforce their own regulations, but site-wide guidelines and norms may also influence behavior. This paper reports on a mixed-methods study of 100,000 subreddits and their rules. Our findings characterize the types of rules across Reddit, the frequency of rules at scale, and patterns of rules based on subreddit characteristics. We find that rules appear to be context-dependent for individual subreddits but also share common characteristics across the site. Taken together, our findings provide a rich description of this ecosystem of rules, motivating further inquiry into underlying mechanisms for rule formation and enforcement in online communities.

Introduction

Online communities are critical to the social health of our society, and a part of everyday life. These spaces deal with increasingly complicated challenges—from harassment (Pater et al. 2016) to hate speech (Chandrasekharan et al. 2017) to misinformation (Starbird 2017)—resulting in public discourse that focuses on the worst consequences of social platforms. Though design affordances can amplify behavior such as harassment (Vitak et al. 2017), blame for bad behavior often falls on regulatory structures—on Facebook for “allowing” fake news, or on Twitter for not effectively combatting harassment. Kraut and Resnick put forth effective regulation as one of the factors that makes online communities successful (Kraut and Resnick 2014).

However, regulation is complex, as it rarely comes in the form of a single, cohesive structure. Different sources of regulation constrain our behavior (Lessig 2006) and therefore it is important to examine the entire ecosystem of governance at different levels—for example, rules instituted by a platform versus norms that form through behavior. Ostrom (2000) suggests that rules created by a community

itself are most effective. However, implicit norms can be ambiguous or difficult for newcomers to learn (Lampe et al. 2014). Therefore, an effective mechanism for governance may be *community-created yet formalized* rules, based on existing norms but also visible for community members.

The social sharing and news aggregation site Reddit¹ provides one example of an ecosystem of community-created rules alongside site-wide policy. Reddit hosts more than one million smaller communities called *subreddits* with different areas of interest, including news, gaming, TV shows, and even personal subreddits that serve as private chat rooms for just a few people. Users on Reddit (or “Redditors”) can post links to outside content or create their own text posts. The primary mode of interaction is commenting, as well as upvoting and downvoting posts and comments.

Behavior on Reddit is formally governed in three ways: (1) a user agreement² and content policy³ similar to any website’s terms and conditions; (2) a set of formalized site-wide values created by redditors themselves, called Reddiquette,⁴ and (3) rules for individual subreddits, set by moderators and typically found in sidebars, though not always present. Moderators can remove posts or ban users, and users have the ability to report rule violations. *Rules* in this paper refer to these community-created rules unique to individual subreddits, existing alongside side-wide policy.

The purpose of this study is to provide a comprehensive description of these multiple layers of rules. Though our findings provoke specific questions for future work around issues such as rule formation and enforcement, our goal was to characterize the nature of rules without delving into underlying mechanisms. As such, the work was motivated by four overarching research questions:

RQ1: What types of subreddits have rules?

¹ www.reddit.com

² <https://www.reddit.com/help/useragreement/>

³ <https://www.reddit.com/help/contentpolicy/>

⁴ <https://www.reddit.com/wiki/reddiquette>

RQ2: What types of rules do subreddits have, and in what frequencies?

RQ3: Do certain types of subreddits have certain types of rules?

RQ4: How similar are rules across subreddits, and how do they relate to other sources of governance on Reddit?

To answer these questions, we conducted a mixed methods analysis of formal rules on subreddits. Qualitative content analysis of the rules from 1,000 subreddits led to a large-scale computational analysis of 100,000 subreddits. Our findings reveal that rules are largely dependent on the context of an individual subreddit, but share some common characteristics across the site. Rules not only have relationships to each other, but also to site-wide guidelines both imposed by Reddit and articulated by redditors. Taken together, these findings reveal patterns that motivate research questions that could delve into the underlying mechanisms of rules and self-governance.

Background & Related Work

Policy and governance in online communities is an important part of user interactions and community health (Preece, Maloney-Krichmar, and Abras 2003). *Policy* in this context is not just traditional law or public policy, but all processes by which users' interactions are regulated. These processes (though often overlooked) are embedded in the fabric of collaborative environments, and arguably should be considered alongside design and practice when creating and studying systems (Jackson, Gillespie, and Payette 2014; Centivany and Glushko 2016). Policy also plays an important role in translating community values into user interaction (Centivany 2016).

A traditional idea of *rules* for online platforms often includes externally imposed policies such as Terms of Service or privacy policies. These documents are typically legally binding yet rarely read and difficult to understand, despite the fact that members of online communities may react badly if they were aware of them (Fiesler, Lampe, and Bruckman 2016). Moreover, policies can be highly variable across websites (Fiesler, Lampe, and Bruckman 2016; Pater et al. 2016). Formal rules also come from other sources, such as the highly adjudicated community-created rules on Wikipedia (Keegan and Fiesler 2017).

However, regulation is not always visible, with governance occurring at the informal norms level instead. Prior work on governance in online communities suggests the importance of social norms in regulating behavior, yet we also know that the difficulty for newcomers learning norms can lead to high drop-out rates (Lampe et al. 2014). Violations of community standards may not only drive members away, but can undermine the purpose of a community (Seering, Kraut, and Dabbish 2017). In the case of Reddit,

with so many subreddits to choose from, redditors can also be newcomers over and over, but making rules more visible may have a positive effect on newcomer participation (Matias 2016b).

Though there are many sets of norms across Reddit, one unifying aspect is its technical design, including algorithms. It is therefore difficult to disentangle community norms from the way they are shaped by the platform itself (Massanari 2015). Reddit's aggregation and algorithms have the ability to suppress or highlight certain types of content, and there is particular power to visibility, such as a subreddit becoming a "default" on the main page. This is one reason why default subreddits can have a disproportionate impact on Reddit culture overall (Centivany 2016).

Reddit does have administrative oversight, though it has shifted over time; in 2008 subreddits became user created and controlled (Matias 2016a). Following this, moderators came to have a lot of power—and also to do a lot of work, including deciding on their individual subreddits' rules (Matias 2016a). As of 2015, there were more than 150,000 accounts with moderator privileges on Reddit (Matias 2016a). In a 2015 paper, Mills and Fish (2015) contrasted current Reddit to Reddit of 2011, noting that subreddits with more recent activity had formed more comprehensive rules. Additionally, they observed heavier enforcement practices from moderators, especially with respect to deleting content, a departure from earlier ideals of distributed moderation on the site (Mills and Fish 2015). In other words, Reddit's *users* have more control, but this includes moderators. Prior work has also shown that moderation becomes heavier when there is a large influx of newcomers (e.g., when a subreddit becomes more visible) (Kiene, Monroy-Hernández, and Hill 2016).

However, a recent incident brought a reminder that there *are* still site-wide controls over individual subreddits. In June 2015 the company banned a handful of subreddits for violating the site's user agreement (Newell et al. 2016). In their announcement, the Reddit team stated, "Today we are removing five subreddits that break our Reddit rules based on their harassment of individuals."⁵ Following the ban, Chandrasekaran et al. (2017) examined causal effects via changes in participation and hate speech, concluding that this policy enforcement did have a positive effect in reducing hate speech. However, community reactions to the ban, as well as the "Reddit blackout" where users and moderators protested a high-level policy decision, have highlighted tensions in the community between administrators and moderators, as well as between moderators and their community members (Centivany and Glushko 2016; Matias 2016a).

⁵ https://www.reddit.com/r/announcements/comments/39bpam/removing_harassing_subreddits/

Moreover, just as subreddit moderators reacted badly to increased administrative oversight, community members may disagree with moderators taking a strong hand. In any online community, rules can be a double-edged sword; though rules can stop unwanted behavior, they can also deter contribution if community members feel stifled (Kraut and Resnick 2014). In cases where an influx of newcomers lead to stronger moderation practices, community members may feel that the environment becomes constrained and contrived (Kiene, Monroy-Hernández, and Hill 2016)—or, alternately, that strong moderators help maintain a positive perception of community content (Lin et al. 2017).

Tension between site-wide regulations and subreddit autonomy demonstrates the importance of examining an entire ecosystem of governance. Conflicts between multiple sources of rules can result in confusion and disagreement (Fiesler, Feuston, and Bruckman 2015). The current work both characterizes rules from multiple sources and considers how they fit together.

Data

We collected two datasets for the work we present in this paper. The first was our manually coded dataset (hereafter “qual dataset,” described in the next section), which consists of data from 1,000 subreddits. These subreddits were randomly selected from a larger dataset of 10,000 subreddits (in order to ensure a range of popularity ranks, described below), collected in November 2016. We excluded subreddits that were private, had zero posts, or were in a language other than English. The second dataset, comprised of data from 100,000 subreddits, was collected in April 2017 and was used for our large-scale computational analysis (hereafter “quant dataset”).

Both datasets were created using Reddit’s API. Notable to our analysis are (1) “description,” a text field that includes text from the sidebar where subreddits typically publish their rules, and (2) metadata, including number of subscribers, number of moderators, date created, and number of active users at the time of collection.

Additionally, we assigned each subreddit a *popularity rank* based on the order in which the subreddits were returned via the API (also reflected in the “popular” tag on the subreddits listing page).⁶ A proprietary sorting by Reddit, this ranking is based not on purely number of subscribers, but how currently active the subreddit is. For example, /r/the_donald (a subreddit dedicated to Donald Trump) has been consistently in the top 3 subreddits since our data collection in November 2016, but has considerably fewer subscribers than consistently lower ranked subreddits such as /r/funny. Additionally, /r/2016Elections appeared in our

| Classification | Accuracy | Precision | Recall | F1 |
|--------------------------------------|----------|-----------|--------|------|
| Rules/No Rules | 0.84 | 0.86 | 0.85 | 0.85 |
| <i>Rule Type Classifiers</i> | | | | |
| Advertising & Commercialization | 0.74 | 0.71 | 0.72 | 0.80 |
| Consequences/Moderation/ Enforcement | 0.69 | 0.69 | 0.70 | 0.69 |
| Content/Behavior | 0.97 | 0.94 | 0.97 | 0.96 |
| Copyright/Piracy | 0.92 | 0.89 | 0.90 | 0.67 |
| Doxxing/Personal Info | 0.85 | 0.60 | 0.77 | 0.68 |
| Format | 0.71 | 0.68 | 0.68 | 0.63 |
| Harassment | 0.76 | 0.76 | 0.75 | 0.88 |
| Hate Speech | 0.80 | 0.69 | 0.75 | 0.67 |
| Images | 0.74 | 0.80 | 0.79 | 0.68 |
| Links & Outside Content | 0.67 | 0.78 | 0.78 | 0.66 |
| Low-Quality Content | 0.87 | 0.77 | 0.87 | 0.76 |
| NSFW | 0.77 | 0.83 | 0.78 | 0.77 |
| Off-topic | 0.64 | 0.57 | 0.60 | 0.81 |
| Personal Army | 0.93 | 0.89 | 0.94 | 0.72 |
| Personality | 0.59 | 0.49 | 0.51 | 0.54 |
| Politics | 0.96 | 0.87 | 0.93 | 0.92 |
| Prescriptive | 0.88 | 0.82 | 0.90 | 0.50 |
| Reddiquette/Sitewide | 0.79 | 0.74 | 0.74 | 0.90 |
| Reposting | 0.87 | 0.81 | 0.85 | 0.68 |
| Restrictive | 0.85 | 0.75 | 0.87 | 0.81 |
| Spam | 0.77 | 0.67 | 0.72 | 0.62 |
| Spoilers | 0.92 | 0.93 | 0.92 | 0.89 |
| Trolling | 0.90 | 0.75 | 0.87 | 0.80 |
| Voting | 0.87 | 0.67 | 0.82 | 0.74 |

Table 1. Classifier metrics.

November 2016 dataset (the height of election season), but had dropped out of the top 100,000 by the time of our April 2017 data collection, likely due to lack of recent activity.

Creating a Qualitative Dataset

As part of our qual dataset, we constructed codebooks for the manual coding of rule types and subreddit topic. First, we manually collected the first 300 rules across subreddits in descending order of popularity. In total, 18 subreddits appeared in our set of rules (e.g., /r/askreddit, /r/the_donald, /r/soccer). Because of difficulty in extracting rules reliably from the metadata, datasets of rules were constructed manually by a researcher visiting each subreddit in a sample.

Next, three researchers worked together in an inductive, iterative coding process (Charmaz 2006; MacQueen et al. 1998), coding independently for emergent rule types and then coming together to adjudicate differences and iterate on codes. The rule codes became saturated within this 300-rule dataset (finding no significant new properties). Codes were based on both properties and topics of rules—for example, “restrictive” for rules that restrict behavior, and

⁶ <https://www.reddit.com/subreddits>

“spam” for rules about spamming content. The codebook consisted of 24 codes; we followed techniques for group codebook creation, laying out definitions, examples, and counter-examples (MacQueen et al. 1998). Codes were not mutually exclusive (e.g., a single rule might be about harassment *and* doxxing). All rule codes are in Table 1.

The three researchers who created the rule codebook came to a group consensus on a set of codes for the 300-rule dataset, which was then used as training data for an additional researcher. We checked for inter-rater reliability with each researcher against coding done by another, achieving a Cohen’s Kappa score of at least 0.65 (considered higher than “substantial agreement” (Landis and Koch 1977)). Researchers also discussed coding disagreements to ensure that they were due to reasonable subjective judgments and not systematic misunderstandings. We then used the codebook to manually code the larger set of rules in our qual dataset. We collected and coded rules (N=3,789) for our November dataset of 1,000 subreddits. The researchers met weekly during this process, discussing any rules that were challenging to code and occasionally spot-checking.

To flesh out subreddit-level data, while manually collecting rules we noted whether the subreddit had rules or not, and the *topic*. To code topics, we conducted a similar inductive, iterative coding process. We adjudicated as a group, eventually coming to consensus on a set of 28 topics, and discussing as a group subreddits that were difficult to categorize. From most common to least common in our qual dataset, the topics included: Video Games (108 subreddits), Image Sharing (80), Entertainment (78) Personal (75), Technology and Science (68), Celebrity (61), Hobby (59), Local (52), Sports (40), Support (38), Music (37), Art (35), Humor (35), Meta (28), Business and Organizations (25), Animals (20), Work (19), Writing (19), Learning (17), Politics (17), Games (15), News (15), Health (14), Food (11), Parody (11), Culture (10), Fashion (8), and Drugs (6).

The final qual dataset was comprised of 1,000 subreddits, 523 of which included rules, for a total of 3,789 rules. Subreddits that included rules had, on average, 7.11 rules (SD = 6.66). We describe the types of rules and subreddit topics in more detail in the next section.

Rules on Reddit

Our findings describe the types, frequency, patterns, and similarity of rules across subreddits. We performed descriptive statistics on our qual dataset, and used machine learning techniques to make predictions at scale from our quant dataset. On all statistical tests and regressions, we applied the stepwise Holm-Bonferroni method to adjust p-values for multiple comparisons.

Using labels from our qual dataset, we built two kinds of predictive classifiers for the analyses that follow: (1) classifiers based on subreddit metadata features (e.g., number of

subscribers); and (2) classifiers based on subreddit description features (i.e., language of the subreddit). For these predictions, we developed (with the Python machine learning library Scikit-learn) several binary classifiers to identify whether rules are present at all, as well as whether a rule of a certain type is present on a given subreddit. These included logistic regression, random forest, support vector machine (RBF kernel and linear kernel), k-nearest neighbors, and naïve Bayes classifiers. We did not develop multi-class classifiers because rule types are not mutually exclusive (e.g., a rule can be both prescriptive and about harassment).

For classifiers based on description, we used TF-IDF weighted unigrams, bigrams, and trigrams from the subreddit’s description. We evaluated all classifiers using F1 scores with 10-fold cross validation, though we also report accuracy, precision, and recall for transparency. The classifiers gave similar performance across classification tasks. In this paper, we only report the results of logistic regression classifiers as they produce probabilistic predictions and are efficient to implement.

Next, we describe findings organized around our central research questions, presenting the results from our two sets of analyses together for the sake of clarity.

What types of subreddits have rules?

Though many subreddits have rules, this is not true for all. The presence of rules is not a site-wide requirement but a decision made at the discretion of the individual subreddit’s moderators. As previously noted, during our manual coding process we found that 52% of subreddits in our qual dataset (523 out of 1,000) had rules, and the rest did not.

Rules and Subreddit Characteristics

Based on available metadata (e.g., number of subscribers), we first conducted a logistic regression between these numeric characteristics and the presence or absence of rules across subreddits in our 1,000 subreddit dataset.

This regression revealed that the only strongly informative characteristic for presence or absence of rules is subreddit popularity rank (as described above, this is based on a metric for overall activity level). The more popular the subreddit according to this ranking (a lower number), the more likely it is to have rules ($\beta = -0.0002$, $p < 0.001$).

Because there was only one informative characteristic from metadata, a classifier based on metadata does not perform well. Classifiers for predicting rules or no rules using available metadata (number of subscribers, date created, etc.) performed at best 60% accuracy, barely better than chance. In other words, numerical characteristics of a subreddit are not good predictors of whether it has rules.

We then considered whether the language of the subreddit was a better predictor, and found that it was. Our logistic regression classifier for rules or no rules performed with 0.85 F1 score (see Table 1). Using this classifier on our quant dataset of 100,000 subreddits, 23,752 had rules

(about a quarter). Note that this is a considerably lower proportion than in our qual dataset, where 52% of the subreddits had rules. The likely reason for this discrepancy is that the average popularity ranking for the qual dataset is much higher than for the quant dataset, as it encompassed the top 10,000 by ranking, contrasted to the top 100,000.

According to redditmetrics.com, there were over 1 million subreddits as of April 2017. However, many of these are abandoned or were never active. In other words, the top 100,000 (and certainly all one million) represent many subreddits that are inactive and unlikely to have rules. We can therefore speculate that of the subreddits that most community members are currently visiting, somewhere around a quarter to a half have rules. The importance of popularity ranking also emphasizes a potential limitation of our analysis, which is that popularity rank could have an impact on the type of rules present as well, and our sampling method was dependent on popularity rank. However, we feel that our sampling represents subreddits with active users.

We also examined *topic* with respect to whether subreddits are more likely to have rules. Within our qual dataset, the highest percentage of subreddits with rules are Writing (79%) and Art (78%). Writing and Art are categories that include a large number of user-submitted artistic content and (as discussed later) are likely to have strict rules around issues such as ownership and formatting. For most topics, at least 30% of subreddits have rules, with one outlier—only 6% of subreddits labeled Personal. This categorization was given to subreddits that appear to be spaces for small numbers of personal friends. Personal subreddits in our qual dataset also have a much lower average popularity ranking than any of the other topics (7,629 out of 10,000) and by far the lowest average number of subscribers ($\mu = 20.04$). 75 out of 1,000 subreddits in our qual dataset were labeled as Personal. Since they tend to be ranked so low with respect to popularity, we speculate that there are even more subreddits like this on Reddit overall, perhaps accounting for many of the subreddits that do not have rules.

In sum, the biggest factor in determining whether a subreddit is likely to have rules is popularity ranking. This supports prior work regarding subreddits that get an influx of new activity; they often end up having to add more moderators as regulation becomes more important (Kiene, Monroy-Hernández, and Hill 2016).

Number of Rules

In addition to the presence or absence of rules, we also considered number of rules present in subreddits that have rules. We conducted a negative binomial regression (since our data is overdispersed and does not include zeroes) between subreddit metadata characteristics and number of rules. With $p < 0.001$, subreddit characteristics indicative of number of rules are number of moderators ($\beta = 1.89$) and how old the subreddit is ($\beta = -0.76$). Here, we see that newer subreddits have more rules on average than older subred-

aits, which could potentially suggest a shift towards a norm of more regulation with more moderators.

What types of rules do subreddits have?

Our analysis thus far revealed that characteristics of subreddits were not very predictive of whether or not a subreddit has rules—but for subreddits that do have rules, what about the types of rules that they have?

Rules Types and Frequencies

In addition to our rules/no rules classifier based on subreddit description, we built classifiers to predict the presence or absence of each type of rule. For these classifiers, the average F1 score was 0.75, ranging from 0.50 to 0.96. The low F1 score of some rule types may be attributed to the small amount of and imbalanced training data of those types (the F1 score and precision/recall/accuracy of each classifier is described in Table 1).

Therefore, based on initial qualitative analysis to identify and categorize rule types and subsequent analysis at scale using predictive machine learning, we are able to describe the frequency of different types of rules across Reddit. Table 2 shows frequency at three levels: total number of rules in our qual dataset of 1,000 subreddits (including rules that may appear more than once in a single subreddit), number of subreddits that have at least one instance of that rule in that dataset, and number of subreddits that are likely (based on classifiers) to have at least one instance of that rule from the quant dataset of 100,000 subreddits.

It should be noted that our categories do not represent the entirety of the wide range of rules across Reddit. Some rules are highly specific and likely do not represent overarching patterns. For example, */r/otterable* contains the rule “Your post must contain otters.” This rule is categorized as “prescriptive” (because it tells someone what to do rather than what not to do) and “off-topic” (because it specifies topic constraints for posts). However, there is not an “otter” category. Therefore, there is the potential for information or categories that we did not capture, but our frequencies do capture the rule types in our codebook.

Restrictive and Prescriptive Rules

One type of rule we captured were the higher-level categories of restrictive (“don’t do this”) versus prescriptive (“do this”) rules, as they represent different strategies for rule enforcement that could reflect values within Reddit or subcommunities. In other words, are moderators creating and enforcing rules that explain what the community *is not* or what the community *is*? Criminology literature suggests that there may be advantages to strategies of “here is what you do next time” rather than “you have been bad” (Braithwaite 1989). The majority of subreddits in our dataset have a mixture of both restrictive and prescriptive rules, and the presence of these rules at all are about in equal number (see Table 3). We also wanted to know whether certain types of rules tend to be framed as restric-

| Rule Type | Manually Coded | | | Large-Scale | | |
|-------------------------------------|----------------|--------|------------|-------------|---------|--------|
| | Rules | % | Subreddits | % | Dataset | % |
| Advertising & Commercialization | 208 | 5.49% | 134 | 25.62% | 1793 | 7.55% |
| Consequences/Moderation/Enforcement | 788 | 20.80% | 295 | 56.41% | 21485 | 90.46% |
| Content/Behavior | 2726 | 71.95% | 505 | 96.56% | 23737 | 99.94% |
| Copyright | 53 | 1.40% | 51 | 9.75% | 233 | 0.98% |
| Doxxing/Personal Info | 109 | 2.88% | 82 | 15.68% | 128 | 0.54% |
| Format | 529 | 13.96% | 237 | 45.32% | 9231 | 38.86% |
| Harassment | 157 | 4.14% | 127 | 24.28% | 637 | 2.68% |
| Hate Speech | 122 | 3.22% | 105 | 20.08% | 378 | 1.59% |
| Images | 273 | 7.21% | 157 | 30.02% | 1844 | 7.76% |
| Links & Outside Content | 368 | 9.71% | 195 | 37.28% | 5757 | 24.24% |
| Low-Quality Content | 90 | 2.38% | 70 | 13.38% | 200 | 0.84% |
| NSFW | 175 | 4.62% | 141 | 26.96% | 969 | 4.08% |
| Off-topic | 274 | 7.23% | 205 | 39.20% | 3923 | 16.52% |
| Personal Army | 41 | 1.08% | 34 | 6.50% | 45 | 0.19% |
| Personality | 244 | 6.44% | 210 | 40.15% | 7218 | 30.39% |
| Politics | 33 | 0.87% | 23 | 4.40% | 24 | 0.10% |
| Prescriptive | 1491 | 39.35% | 458 | 87.57% | 23698 | 99.77% |
| Reddiquette | 142 | 3.75% | 113 | 21.61% | 1412 | 5.94% |
| Reposting | 88 | 2.32% | 76 | 14.53% | 390 | 1.64% |
| Restrictive | 1777 | 46.90% | 442 | 84.51% | 23610 | 99.40% |
| Spam | 146 | 3.85% | 123 | 23.52% | 401 | 1.69% |
| Spoilers | 66 | 1.74% | 42 | 8.03% | 173 | 0.73% |
| Trolling | 59 | 1.56% | 52 | 9.94% | 73 | 0.31% |
| Voting | 89 | 2.35% | 60 | 11.47% | 71 | 0.30% |

Table 2. Frequency of rule types within 1) manually coded rules (N=3,789), 2) presence within manually coded subreddits that had rules (N=523), and 3) presence in subreddits with rules from the large-scale dataset (N=23,752).

tive or prescriptive (e.g., the difference between the common rules “don’t be a dick” versus “be civil”). Using the presence or absence of each rule code for a single rule as categorical variables, we conducted chi-squared tests to determine the relationship between each rule type and the restrictive and prescriptive categories (see Table 3). With $p < 0.05$, the majority of rule types are more likely to be framed as restrictive: don’t harass people, don’t spam, don’t use hate speech, etc. Those that are significantly likely to be framed as prescriptive are formatting (providing formatting requirements rather than restrictions), Reddiquette (telling one to follow Reddiquette or other rules, rather than to not break other rules), and personality. For this last, this means that we see more “be civil” than “don’t be a dick”; though subreddits may specifically prohibit bad behavior, they still model good behavior, and are more likely to model personality for a person who is welcome rather than unwelcome.

Common Rules

Table 2 shows the frequency of types of rules in both our qual dataset and predicted at scale. Notably, 74% of all of the rules in the qual dataset fall into the content/behavior category. Because interaction on Reddit takes place via posts and comments, it is often difficult to differentiate

| Type | Restrictive vs. Prescriptive | Chi-Squared |
|-------------------------------------|------------------------------|-------------|
| Advertising & Commercialization | Restrictive | 53.85*** |
| Consequences/Moderation/Enforcement | Restrictive | 28.63*** |
| Content/Behavior | Restrictive | 128.84*** |
| Doxxing/Personal Info | Restrictive | 53.48*** |
| Format | Prescriptive | 226.71*** |
| Harassment | Restrictive | 102.73*** |
| Hate Speech | Restrictive | 72.43*** |
| Links & Outside Content | Restrictive | 11.13** |
| Low-Quality Content | Restrictive | 40.00*** |
| Personal Army | Restrictive | 23.44*** |
| Personality | Prescriptive | 47.12*** |
| Politics | Restrictive | 6.82* |
| Reddiquette/Sitewide | Prescriptive | 28.85*** |
| Spam | Restrictive | 60.52*** |
| Trolling | Restrictive | 37.44*** |

* = $p < .05$, ** = $p < .01$, *** = $p < .001$

Table 3. Rule types and whether they are more likely to be restrictive vs. prescriptive. Only significant results included.

rules about content from behavior, but the majority of rules on Reddit apply to what content users can post and/or how they can behave.

Rules about consequences include information about how rules are enforced or what the consequences are for breaking the rules. The most frequent consequences mentioned are the post being removed or the user being banned. However, only 21% of the rules in our dataset specifically mention consequences. Of the subreddits in our qual dataset that have rules, only about half (295 out of 523) have any consequences noted at all. Common types of rules that note consequences include personality, links and outside content, and off-topic. Taken together, these rules essentially provide information about what kind of behavior and content is welcome (or not welcome) in a particular community. As Mills (2017) noted in his study of subreddits during the 2016 U.S. presidential election, rules about what kind of content is welcome or not can have a huge influence on the culture of the community. Formalized rules therefore work towards clarifying what behavior is appropriate for each of these small communities.

Do certain types of subreddits have certain types of rules?

Beyond the presence or absence of rules altogether, we also examined whether features of subreddits might be predictive of certain types of rules. For example, are subreddits of a certain size or about a certain topic likely to have rules about harassment, or copyright, or politics?

Rule Types & Subreddit Characteristics

Similar to predicting the presence of rules, meta characteristics are not very predictive of specific types of rules. We conducted binary logistic regressions between each rule type and meta characteristics for the 523 subreddits from our qual dataset with rules. We found no significant relationships between rule types and number of subscribers, moderators, active users, or date of creation. Eleven rule types were significantly more likely ($p < 0.05$) to be found in subreddits with a lower number for a popularity ranking (i.e., more popular). These included rules that suggest strong regulation (restrictive rules and consequence rules, both $\beta = -0.0001$), rules that specify what kind of content is welcome (links & outside content, $\beta = -0.0001$; format, $\beta = -0.0001$; low quality content, $\beta = -0.0003$; and reposting, $\beta = -0.0002$), and rules about bad behavior (hate speech, $\beta = -0.0001$; doxxing, $\beta = -0.0002$; advertising, $\beta = -0.0002$). This finding could suggest that these are important factors in regulating communities with heavy activity.

In sum, there are few relationships between meta-characteristics of subreddits and rule types. This suggests that other, more context-specific features of subreddits are determinants of what kinds of rules are created.

Topics and Rules

For each topic, we calculated the percentage of subreddits of that topic that have each type of rule, and though our coded dataset is too small to make significant claims about the relationship between topic and rules, there are a few interesting observations. For example, art subreddits are most likely to have rules about formatting (85% of Art subreddits have a formatting rule), copyright (56%), and NSFW content (70%). Every drug-related subreddit has a rule restricting advertising, and are also most likely to have a rule pointing to external rules—typically, legality. Entertainment subreddits are most likely to have a rule about spoilers (44%). Politics subreddits are most likely to have rules about hate speech (67%), trolling (56%), and personal armies (78%). These patterns suggest that the subject matter is more likely to be an important contextual factor for the kinds of rules that are a present than structural features.

How similar are rules across Reddit?

Considering the entire ecosystem of rules on Reddit poses questions not just about rules' relationships to their individual subreddits, but also how similar or different they might be across the site. Are there copies of rules, and how do they reference each other or site-wide rules?

Copies of Rules

First, we wondered how often subreddits are simply copying rules verbatim from each other. It would be reasonable to think that smaller, newer subreddits bootstrap by directly borrowing rules from more established subreddits. However, within our qual dataset, only 111 out of 3,789 rules (3%) appeared verbatim on multiple subreddits, and only 95 subreddits (18%) shared one or more rules with each other.

We also examined this question at scale. In the course of our qualitative analysis we observed that most rules are represented in numbered lists or bullet points. Therefore, for our quant dataset of 100,000 subreddits, we extracted text from HTML list elements that directly follow the word “rules” in their descriptions. Using this method (assuming that these are likely rules), we found that only 2489 subreddits (2.5%) have direct copies of one or more rules with other subreddits, and only 3667 rules (5.3%) out of 69237 appeared multiple times. Though this method may have both false positives and false negatives, it seems that verbatim copying of rules is relatively rare.

Other Sources of Rules

Reddit Policies. Like most websites, Reddit has a set of official policies. The main page notes that use of the site constitutes acceptance of the user agreement and privacy policy. The user agreement sets out similar terms as a traditional TOS, and though it contains a number of legal provisions, is written more clearly than most similar documents (Fiesler, Lampe, and Bruckman 2016). The agreement also states that users should follow the content policy, and nods to subreddit rules: “Subreddits may create their own rules

and enforce them as they see fit, providing they do not violate the terms of this agreement.”

The content policy is a separate document that lays out the type of content that is acceptable or not on the site. For example, it prohibits content that (1) is illegal; (2) encourages or incites violence; (3) threatens, harasses, or bullies or encourages others to do so; (4) is personal and confidential information; (5) impersonates someone in a misleading or deceptive manner; or (6) is spam. Most of these restrictions reflect what we see in rules across the site.

Subreddits rarely directly refer to these policies. Only 164 out of the 100,000 subreddits in our quant dataset mention “terms of service” or “content policy” in their description. Discounting mentions of “terms of service” that reference rules external to Reddit (e.g., “Please don’t break Frontier’s [a video game company] terms of service”), a typical example is “Follow the Reddit content policy, including the provisions on unwelcome and prohibited behavior” (*/r/DnD*). These typically appear along with a set of other rules. Subreddits are not frequently using a single pointer to the content policy to replace having rules, but rather to supplement them by filling in gaps.

Reddiquette. In addition to policies imposed by Reddit, the site is also governed by site-wide “Reddiquette” that includes guidelines such as “report any spam you find” and “please don’t ask people to troll others on Reddit.” It is intended as an “informal expression of values” written by redditors rather than site operators. Reddiquette is fairly comprehensive, with 56 guidelines, falling into “please do” (i.e., prescriptive rules) and “please don’t” (i.e., restrictive rules). There is even a rule that points back to subreddit governance: “Read the rules of a community before making a submission. These are usually found in the sidebar.”

The Reddiquette guidelines track closely to the types of rules we notice elsewhere on Reddit. There is a Reddiquette rule for every one of our topic categories with the exception of content-specific ones: off-topic, low-quality content, politics, copyright (though this may be covered under “don’t engage in illegal activity”), and spoilers. Because Reddiquette has not always existed and was written by redditors themselves, it is likely that it expresses rules and values that already existed across subreddits, but may also have in turn influenced later subreddits.

One of our rule types refers to references to Reddiquette or other external rules. In our qual dataset, only 113 out of 523 subreddits with rules had a rule like this - most commonly, simply stating a rule of following Reddiquette (“Remember your Reddiquette”) or pointing to Reddiquette for clarification of a certain rule (“Reddiquette is to be observed, especially the rule pertaining to grammar”). We also examined this at scale, and found that out of our 100,000 subreddits in the quant dataset, 3,674 (less than 4%) mention “Reddiquette” somewhere in the description.

We also examined whether specific rules from Reddiquette are addressed on individual subreddits. We found only 151 subreddits with exact string matches to Reddiquette rules, and none include every Reddiquette rule, with the majority only mentioning one. The most common by far is “remember the human,” the first rule stated in Reddiquette, and typically appears as shorthand for civility.

Overall, these findings suggest that though for the most part subreddits are creating their own rules, they are using Reddiquette as a common frame of reference and/or to fill in the gaps in rules they may not explicitly state. They may also serve as a kind of “tasting menu” of rules that moderators choose from when they are considering the kinds of rules that might be appropriate for their particular community. In general, however, Reddiquette rules track well to the kinds of rules that exist elsewhere on Reddit and the most dominant values.

It is also much more common for subreddits to refer to Reddiquette than official policy, suggesting again that the rules closest to the community itself are the most visible, prioritizing an individual subreddit over Reddiquette over Reddit policy. However, these all refer to each other at various points, creating an ecosystem of related values.

Discussion

Our findings present a description of the rules ecosystem across Reddit. Though many of the *whys* and *hows* of rules remain for future work (e.g., why certain rules, how they are created and enforced), this exploration of the *whats* provides high-level insight that can motivate further inquiry.

First, we see patterns with respect to whether or not subreddits have rules, as well as how they frame those rules when they do. More “popular” subreddits by Reddit’s ranking appear to have a more structured rule system (more likely to have rules, more rules, and more often framed as restrictions). We know from prior work that an influx of newcomers may necessitate heavier moderation (Kiene, Monroy-Hernández, and Hill 2016), so this may be an effect that tapers off with time, or that ebbs and flows with subreddit activity level. This effect could also be more specific to a moment in time—the most active subreddits may be reacting to policy changes on Reddit that led to the deletion of subreddits that broke Reddit’s content policies (Newell et al. 2016; Chandrasekharan et al. 2017). In other words, the presence of rules in more popular/active subreddits could be a way of nodding to the official rules in order to avoid site-level consequences.

Our findings about the types and relative frequencies of rules across Reddit also provide insight into the site’s overall culture and values. One way to build community is to define what it is (and/or what it is not). The most common rules across subreddits are those that define and model the type of content and behavior that belongs (and does not be-

long). The fact that Reddiquette tracks so well to the overall rules common across the site also suggests that the Redditors who created it did a good job of modeling the value system that they already had or that they hoped to encourage. It is possible that, even if not strongly enforced, formalized rules make norms more visible and therefore help the community be more accessible to newcomers. This is an elegant solution to a known problem within online communities (Lampe et al. 2014; Kraut and Resnick 2014), by providing a clear articulation of group boundaries.

Overall, this work through a number of low-level findings also suggests one overarching one: Though Reddit as a site has a core set of shared values that are reflected in many of the million small communities there, production of the rules on individual subreddits appear to be context dependent. Our findings reveal that objective subreddit characteristics such as size are not predictive of the kinds of rules that appear, or even whether there *are* rules. Instead, we see clearer patterns of rules when it comes to topic—for example, “art” subreddits having rules about copyright.

Similarly, we showed that rules are rarely duplicated verbatim across the site. Policies at different levels (subreddits, Reddiquette, content policy) have relationships to each other in terms of similarities and values articulated, but at the level of subreddits, the ways that they pull from and respond to site-wide rules varies widely.

It is also difficult to infer directionality when it comes to the relationship between subreddit rules and site-wide policies. Redditors themselves created (and still modify) Reddiquette, suggesting that norms and rules from subreddits influence this articulation of values. However, Reddiquette used as a “tasting menu” could also mean that it directly influences decisions about rules. Moreover, Reddit’s content policy could be influencing rules, particularly now that it is clear that these particular rules can be harshly enforced. These all leave interesting open questions about directionality in this rule ecosystem as well as about creation mechanisms for rules. Though the data for this study did not include temporal information, a direction for future work is examining changes in rules over time.

Another important next step for this work would be to delve into the *hows* and *whys* of rule formation by talking to subreddit moderators, in the tradition of prior work that has examined moderation practices on Reddit (Massanari 2015; Matias 2016a; Centivany and Glushko 2016; Centivany 2016; Mills 2017). Our findings support the idea that moderators of individual subreddits have a great deal of influence over the culture of the small communities they lead. Ostrom (2000) suggests that in the context of collective action, when people organize to enforce their own rules, these rules are more successful than those externally imposed upon them. Inquiry into the effectiveness of this model could provide insight as both researchers and platform designers

tackle problems of negative behavior online and how best to regulate it.

Though Reddit as a whole is a single sociotechnical system, our findings suggest that policy formation at the subreddit level is a largely social process. Therefore, Reddit provides a unique opportunity to explore these processes in that social context, while controlling for other regulatory factors such as technical design and external policy (Lessig 2006). Further work to explore how moderators make decisions about rules (including how they are influenced by the rest of the rule ecosystem and how they decide which informal norms to formalize), as well as how Redditors think about which rules to follow, could be valuable in ongoing discussions about governance in online communities.

Moreover, though Reddit itself is a valuable site of study, there are lessons to be learned for other online communities that have their own ecosystems of governance. For example, the wiki community Wikia, though consisting of many small individual wikis, exists in part to support these individual communities that want to work together.⁷ Foote et al. (2017) found that community creators, though unique in their motivations and goals, consider regulation to be an important strategy for community building. It would be useful to explore how Reddit’s model of contextual rules for individual communities supplemented by articulation of site-wide values might translate to other platforms. Our mixed-methods approach to examining patterns of rules across different smaller communities could also be appropriated for analysis of other online spaces.

Conclusion

In our study of the rules ecosystem across Reddit, we have demonstrated that there is a complex relationship between rules at different levels, as well as between subreddits. We also found few patterns based on meta characteristics, suggesting that community-created rules are largely context-dependent. This work contributes a rich description of the nature of rules across Reddit, both qualitatively and quantitatively. With the increasing acknowledgment of social computing scholars that policy is an important piece of understanding collaborative spaces (Jackson, Gillespie, and Payette 2014), we hope that this work provides a grounding for future inquiry that considers governance from the perspective of platforms, moderators, and users.

Acknowledgments

Many thanks to reviewers and colleagues whose comments made this a stronger and more impactful paper.

⁷ http://www.wikia.com/Community_Creation_Policy

References

- Braithwaite, J. 1989. *Crime, Shame, and Reintegration*. New York, NY: Cambridge University Press.
- Centivany, A. 2016. Values, Ethics and Participatory Policymaking in Online Communities. In *Proceedings of the Association for Information Science and Technology 53* (1): 1–10.
- Centivany, A., and Glushko, B. 2016. "Popcorn Tastes Good": Participatory Policymaking and Reddit's "AMAgeddon." In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 1126–37.
- Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J., and Gilbert, E. 2017. You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech. *Proc. ACM Hum.-Comput. Interact.: CSCW* 1 (2).
- Charmaz, K. 2006. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. London, UK: SAGE Publications.
- Fiesler, C., Feuston, J., and Bruckman, A.C. 2015. Understanding Copyright in Online Creative Communities. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*, 116-129.
- Fiesler, C., Lampe, C., and Bruckman, A.S. 2016. Reality and Perception of Copyright Terms of Service for Online Content Creation. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*, 1450-1461.
- Foote, J., Gergle, D., and Shaw, A. 2017. Starting Online Communities: Motivations and Goals of Wiki Founders. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 6376–80.
- Jackson, S.J., Gillespie, T., and Payette, S. 2014. The Policy Knot: Re-Integrating Policy, Practice and Design in CSCW Studies of Social Computing. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*, 588–602.
- Keegan, B.C., and Fiesler, C. 2017. The Evolution and Consequences of Peer Producing Wikipedia's Rules. In *Proceedings of the AAAI International Conference on Web and Social Media (ICWSM)*, 112-121.
- Kiene, C., Monroy-Hernández, A., and Hill, B.M. 2016. Surviving an "Eternal September." In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 1152–56.
- Kraut, R.E., and Resnick, P. 2014. *Building Successful Online Communities: Evidence-Based Social Design*. Boston, MA: MIT Press.
- Lampe, C., Zube, P., Lee, J., Park, C.H., and Johnston, E. 2014. Crowdsourcing Civility: A Natural Experiment Examining the Effects of Distributed Moderation in Online Forums. *Government Information Quarterly* 31 (2). Elsevier Inc.: 317–26.
- Landis, J.R., and Koch, G.G. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33 (1): 159–74.
- Lessig, L. 2006. *Code: And Other Laws of Cyberspace, Version 2.0*. New York: Basic Books.
- Lin, Z., Salehi, N., Yao, B., Chen, Y., and Bernstein, M.S. 2017. Better When It Was Smaller? Community Content and Behavior After Massive Growth. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*, 132-141.
- MacQueen, K.M., McLellan, E., Kay, K., and Milstein, B. 1998. Codebook Development for Team-Based Qualitative Analysis. *Field Methods* 10 (2): 31–36.
- Massanari, A. 2015. #Gamergate and The Fapping: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures. *New Media & Society*, 1–18.
- Matias, J.N. 2016a. Going Dark: Social Factors in Collective Action Against Platform Operators in the Reddit Blackout. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 1138–51.
- Matias, J.N. 2016b. Posting Rules in Online Discussions Prevents Problems & Increases Participation. *CivilServant*. http://civilservant.io/r_science_sticky_coments_1.html
- Mills, R., and Fish, A. 2015. A Computational Study of How and Why Reddit.Com Was an Effective Platform in the Campaign Against SOPA. In *Social Computing and Social Media*, 1-13.
- Mills, R.A. 2017. Pop-up Political Advocacy Communities on Reddit.Com: SandersForPresident and The Donald. *AI & Society* 33(1), 39-54.
- Newell, E., Jurgens, D., Saleem, H.M., Vala, H., Sassine, J., Armstrong, C., and Ruths, D. 2016. User Migration in Online Social Networks: A Case Study on Reddit During a Period of Community Unrest. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*, 279-288.
- Ostrom, E. 2000. Collective Action and the Evolution of Social Norms. *Journal of Economic Perspectives* 14 (3): 137–58.
- Pater, J.A., Kim, M., Mynatt, E.D., and Fiesler, C. 2016. Governing Online Harassment: Characterizing Policies Across Social Media Platforms. In *Proceedings of the ACM Conference on Supporting Group Work (GROUP)*, 369-374.
- Preece, J., Maloney-Krichmar, D., and Abras, C. 2003. History of Emergence of Online Communities. *Encyclopedia of Community* 3, 1032-1027
- Seering, J., Kraut, R., and Dabbish, L. 2017. Shaping Pro and Anti-Social Behavior on Twitch Through Moderation and Example-Setting. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*.111–125.
- Starbird, K. 2017. Examining the Alternative Media Ecosystem Through the Production of Alternative Narratives of Mass Shooting Events on Twitter. In *Proceedings of the AAAI International Conference on Web and Social Media (ICWSM)*.
- Vitak, J., Chadha, K., Steiner, L., and Ashktorab, Z. 2017. Identifying Women's Experiences With and Strategies for Mitigating Negative Effects of Online Harassment. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*, 1231–1245.