

Community Archetypes: An Empirical Framework for Guiding Research Methodologies to Reflect User Experiences of Sense of Virtual Community

GALE H. PRINSTER*, University of Colorado Boulder, USA

C. ESTELLE SMITH*, Colorado School of Mines, USA

CHENHAO TAN, University of Chicago, USA

BRIAN C. KEEGAN, University of Colorado Boulder, USA

Humans need a sense of community (SOC), and social media platforms afford opportunities to address this need by providing users with a sense of *virtual* community (SOVC). This paper explores SOVC on Reddit and is motivated by two goals: (1) providing researchers with an excellent resource for methodological decisions in studies of Reddit communities; and (2) creating the foundation for a new class of research methods and community support tools that reflect users' experiences of SOVC. To ensure that methods are respectfully and ethically designed in service and accountability to impacted communities, our work takes a qualitative, community-centered approach by engaging with two key stakeholder groups. First, we interviewed 21 researchers to understand how they study "community" on Reddit. Second, we surveyed 12 subreddits to gain insight into user experiences of SOVC. Results show that some research methods can broadly reflect users' SOVC regardless of the topic or type of subreddit. However, user responses also evidenced the existence of five distinct *Community Archetypes*: Topical Q&A, Learning & Perspective Broadening, Social Support, Content Generation, and Affiliation with an Entity. We offer the Community Archetypes framework to support future work in designing methods that align more closely with user experiences of SOVC and to create community support tools that can meaningfully nourish the human need for SOC/SOVC in our modern world.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: Sense of community, sense of virtual community, online communities, community support, Reddit, survey, interview, data science, users, researchers, community archetypes

ACM Reference Format:

Gale H. Prinster, C. Estelle Smith, Chenhao Tan, and Brian C. Keegan. 2024. Community Archetypes: An Empirical Framework for Guiding Research Methodologies to Reflect User Experiences of Sense of Virtual Community. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 33 (April 2024), 33 pages. <https://doi.org/10.1145/3637310>

1 INTRODUCTION

For decades, concerning trends point toward the degradation of in-person communities and a growing epidemic of loneliness [115]. Concurrently, participation in online communities has exploded, with rapidly evolving moderation affordances [10] and styles [101], rules and norms [28,

*Both authors contributed equally to this research.

Authors' addresses: Gale H. Prinster, abpr9862@colorado.edu, University of Colorado Boulder, Boulder, Colorado, USA, 80304; C. Estelle Smith, estellesmith@mines.edu, Colorado School of Mines, , Golden, CO, USA, 80401; Chenhao Tan, University of Chicago, Chicago, Illinois, USA, 80304, chenhao@uchicago.edu; Brian C. Keegan, University of Colorado Boulder, Boulder, Colorado, USA, 80304, brian.keegan@colorado.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2024 Copyright held by the owner/author(s).

2573-0142/2024/4-ART33

<https://doi.org/10.1145/3637310>

40, 43], and user groups who may or may not be connected offline [106, 121]. For example, people might not talk to neighbors like they used to, but they might also go on Nextdoor or their city’s subreddit to observe neighbors’ discussions. Academics might opt to work from home more frequently and miss out on water cooler chat, but check Twitter twice an hour to catch the latest hot takes. People suffering from stigmatized illness may never speak a word of it at the office, later going home to an intimate Zoom chat hosted by a private Facebook group for patients.

Examples like these make it clear that how modern humans participate in communities—and consequently, how we experience a Sense of Community (SOC) or Sense of Virtual Community (SOVC)—have shifted over time across different geographical, cultural, and sociotechnical contexts. Studying how people experience SOC/SOVC is meaningful because it is a deep and unavoidable human need. The access to or lack of community has profound psychological and material impacts on people’s lives. For example, SOC impacts life satisfaction [90], well-being [30], perceived safety [84], problem solving [9], and social or political participation [81]. Considerations like these underpin the importance of improving our scientific understanding of: how we develop SOC/SOVC; how we can best understand, measure, and support the formation of healthy communities; and how SOC/SOVC can either nourish our needs or cause damage, depending on the circumstances.

This paper investigates SOVC on Reddit, a platform that serves millions of users and communities worldwide and has attracted growing research attention [91].¹ Our work is motivated by two overarching goals: (1) to provide future researchers with an excellent resource for thinking through major methodological decisions in studies of Reddit communities; and (2) to provide the foundation for a new class of data science techniques and data-driven community support tools. Regarding the first goal, studies that aim to foster healthy online communities should ensure that community spaces and members are properly identified, and that researchers’ selected methods align with community members’ experiences and values. Although much prior work refers to each subreddit as its own community, recent work suggests that individual subreddits are not the best “units” of community. Most users report greater SOVC across *multiple* subreddits [105], and each subreddit in these *sets* may meet different needs [54, 112]. If researchers assume that each subreddit is individually experienced as a community, they risk inaccurate conclusions, or they may miss key aspects of the larger picture of user experience. Providing better ways to identify community spaces will enable researchers to better support the human need for SOC.

Drawing upon prior work in organizational psychology and HCI, our second goal aims to advance our ability to measure and predict SOVC. Inspired by recent work on Twitch [59], we are motivated by the promising opportunity to design and evaluate human-centered data science techniques. However, academic constructions of SOC/SOVC do not always align with the lived experiences of community members [89, 92, 93]. Consequently, *“the problems raised by the use of the notion of community are increased if the gap between academic and lay meaning is not taken into account.”* [74] To reduce this gap, we took an empirical, qualitative, and community-centered approach to developing a guiding framework for new methods. We engaged with two stakeholder groups—Reddit researchers *and* Reddit users—to ask:

RQ1: How do researchers conceptualize and operationalize virtual community on Reddit?

RQ2: How do Reddit users experience a sense of virtual community in their use of the platform? And how does this vary across communities?

RQ3: How can we operationalize important community-related concepts to better align researchers’ methods with users’ experiences?

¹This paper assumes familiarity with Reddit-specific terminologies, however supplemental material section 1 includes detailed descriptions of all terms, such as “subreddit,” “thread,” “Original Poster” (OP), etc. for reference.

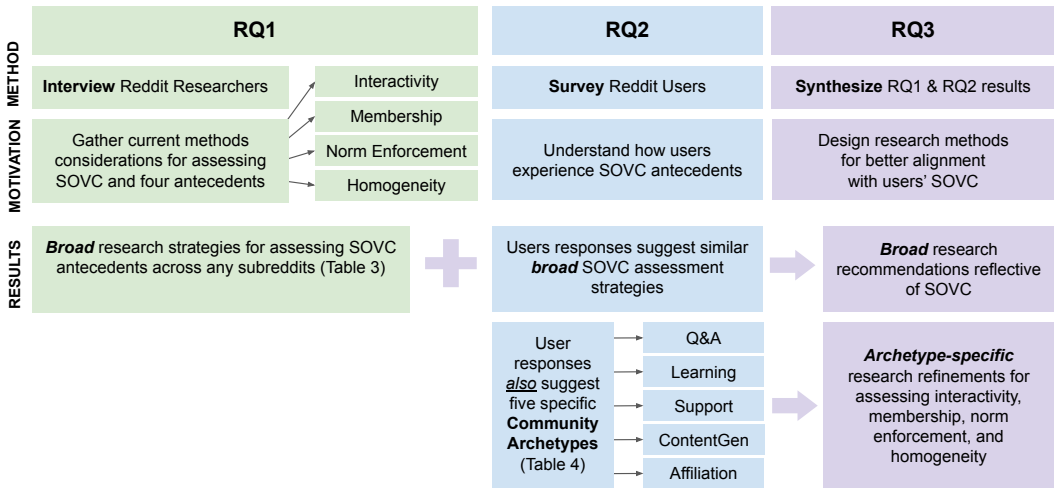


Fig. 1. Summary of Research Approach.

Figure 1 overviews our research approach. We conducted interviews with 21 researchers to address RQ1 and found that they typically view subreddits as topical affinity groups; not all subreddits can or should be described as communities. Rather, certain types of interactivity, membership boundaries, homogeneity, and norm enforcement may distinguish non-community affinity groups from communities (Table 3). For RQ2, we surveyed 12 subreddits. User responses confirmed that their SOVC formation is broadly tied to how researchers tend to study communities. However, users expect nuanced forms of community activity in different types of subreddits; their experiences of SOVC arise primarily when these specialized activity forms are happening well and frequently. Thus our data suggest the existence of five underlying **Community Archetypes** distinguished by particular user roles and content patterns (Table 4). Our discussion synthesizes these two sets of results to address RQ3. Some methods for measuring community structures and SOC/SOVC can be used broadly, no matter the type of subreddit, whereas others must be designed according to which archetype(s) the subreddit(s) embody. We provide examples to illustrate these concepts within existing subreddits and studies. Finally, we offer a road map for future work to apply the Community Archetypes framework and to create community support tools that can meaningfully nourish the human need for SOC/SOVC in our modern world.

2 RELATED LITERATURE

2.1 The history of sense of community research

Organizational and community psychology. Long before the emergence of the Internet, researchers in organizational and community psychology devised varied ways of assessing the structural elements that define geographical communities, and for measuring sense of community (SOC), examining factors like common fate [25], supportive climate [37], and the amount of time one expects to stay [45, 48]. As research advanced, definitions became less attached to geographical centers and more inclusive of *relational* communities denoted by facets like interpersonal connection, interests, and hobbies [8, 48, 70, 76]. One prominent theory of SOC by McMillan and Chavis (1986) proposes four elements: membership, influence, fulfillment of needs, and shared emotional connection [76]. This theory can be applied to geographical *or* relational communities and remains popular today, with occasional uses in HCI (e.g., [63]).

Sociology and communication studies. The fields of sociology and communication studies have also embraced a relational conception of communities. For example, Benedict Anderson examined national identity as a community builder. Although all citizens do not know each other personally, they may nonetheless view themselves as members in the “imagined community” of their nation [7]. This concept of imagined communities provides a helpful lens for online community scholars to understand how users can build communities online without either geographical closeness or personal connections. Sociologists in the 90s have also examined the weak and strong ties that users made both off- and online and found that ties still existed online, despite users not having any geographical proximity [117]. Communications scholars have also operated with the idea that communication is a major indicator and strengthener of communities both off- and online since at least the 90s [96, 103]. Social media may simply be the next step in humans’ millennia-long history in mediated, text-based communities rather than a fundamentally “new” innovation in the course of human history [110].

Measuring SOC/SOVC. While decades of scholarship across different academic disciplines grappled with what a community is (as does our present work in this paper), people seem to reliably know when they are a part of communities and when they are not [6, 70, 98], allowing researchers to use their cognitive evaluations to measure SOC using psychometric instrumentation [53]. While early scales were developed for offline communities, researchers in the early 2000s began to investigate how online communities may differ from offline and to create new psychometric measures for SOVC [17, 65].² One critical issue with these preliminary scales is that they conflated the cognitive evaluation of SOVC with a variety of psychological factors that cause SOVC to develop. Therefore, a more modern approach seeks to distinguish *antecedents* to SOVC from the *experience of SOVC* itself, as well as from specific *outcomes* of SOVC. For example, similarity, interactivity, membership boundaries, common goals, and history of interactions can be distinguished as individual psychological constructs which all *precede* SOVC. Similarly, identification with the group, commitment to the group, group satisfaction, and centrality are all distinct, measurable psychological outcomes, when SOVC exists [16]. Because of the increasing prevalence of automated bots as social actors in online communities [100, 102], recent work also introduced bot governance as a new SOVC antecedent [105].

Analogizing aspects of behavioral trace data as SOVC antecedents. Our work’s theoretical basis benefits from the recent move in organizational psychology to distinguish antecedents from SOVC and its outcomes. In particular, SOVC is an internal and highly variable user experience; it cannot be directly measured via behavioral trace data. However, we aim to understand how users’ internal experiences of *antecedents* to SOVC may relate to empirical measures that *can* be calculated from behavioral trace data. If we can develop data science-based analogs of known SOVC antecedents, we may be able to effectively predict SOVC. As described in section 3.2.3, this work uses four of the most well-studied SOVC antecedents to structure our interviews with researchers: interactivity, homogeneity, norm enforcement, and membership boundaries.

2.2 Bridging from organizational psychology to HCI

HCI and social computing have a long and rich history of studying online communities. However, this history has largely developed adjacent to the literature in organizational psychology. Studies of platforms like Facebook [121], YouTube [97], Twitter [18, 55, 95], Twitch [59], Discord [56],

²For one example, people in online communities are not embodied but are represented by online profiles; consequently issues of purposefully creating an identity, and making identifications of others, are possibly more important to SOVC, relative to in-person communities, where much more information is naturally communicated about identity by physical appearance and proximity [17].

Instagram [18, 113], TikTok [19], and Wikipedia [14, 83] have explored whether and how users experience community and sense of community on the platform. However, other papers use “community” without specifically justifying it. Prior work points to the importance of triangulating qualitative and quantitative data to develop deeper understandings of online communities [59, 97]. As we aim to develop more precise measures and techniques, incorporating concepts across organizational psychology, HCI, and users’ subjective perspectives will improve our ability to delineate and study what users experience as community and how to build community support tools—e.g., moderation policies and toolkits [99, 119], community-in-the-loop algorithms [108] and governance bots [105]. Therefore, we will next summarize four main ways that the HCI literature has previously delineated online communities: platform affordances; shared interests; common vernacular; and interactivity.

Platform affordances. Researchers often designate communities by using particular platform affordances. For example, social network analysis analyzes connections between users who have “added” or “followed” each other. Users are considered “nodes” or “vertices” in the network, connections between them are “edges,” and researchers detect communities based on communication patterns, edge weights, *etc.* [42] Some studies use hashtags to indicate communities, such as Black Twitter [64, 104] or communities of users on Tumbler [46] or Instagram [5, 75] who use the hashtag #depression. On other platforms, communities are designated by more concrete digital “containers” that users elect to join, e.g., Facebook Groups [122], online health blogs [69, 106], or Twitch channels [50]. In line with this “container” view, Reddit’s website copy and many papers (e.g., [2, 24, 66, 86, 87, 111, 116]) refer to each subreddit as its own community. Other papers don’t mention community (e.g., [52, 67]) or identify families of subreddits [111] with overlapping membership [112]. Moreover, most users do not perceive individual subreddits as the best “units” of community, instead reporting greater SOVC across *multiple* subreddits [105], each of which may meet different needs [54]. These studies suggest that users’ SOVC often extends beyond the boundaries of digital containers, and that research will benefit from more fine-grained delineations.

Shared interests. Classic works in HCI denote online community as a group of users with a shared interest or goal [15, 68, 75, 88]. If users simply exist in the same online space with nothing in common, it is unlikely that they will form community-like connections. However, when users have a common interest or goal driving them together, community bonds are much more likely to form. Shared interests can be such a strongly unifying force that users may attempt to migrate their communities to new platforms when current platforms are insufficient or threatened. For example, communities of fandom users have often migrated across platforms [39] and hundreds of thousands of Twitter users rapidly migrated to Mastodon after the 2022 Musk acquisition [61].

Common vernacular. Another approach from natural language processing that is commonly used in HCI focuses on the common vernacular that users employ when interacting in online spaces. Use and familiarity with specific language gives users the sense that they are part of an “in-group.” [58] Therefore, the more that specific forms of language, discursive conventions, or terminologies exist in an online space, the more likely it is to be a community [34, 58].

Interactivity. Finally, HCI researchers have focused broadly on interactivity: the more activity that exists in a group of users, the stronger the community [34, 75, 88]. Some HCI work considers interactivity as a mechanism for designating user groups, such as core v.s. peripheral community members [22, 49], old-timers v.s. newcomers [31, 118], or differing user roles [57, 79, 120]. Diverging from considerations of interactivity as a determination of membership or as an *antecedent* to SOVC however, much work refers to the type and volume of interaction instead as a *success* metric. For example, Cunha et al. describe four success measures: (1) growth in the number of members; (2) retention of members; (3) long-term survival of the community; (4) volume of activity [33]. Many studies take a similar approach, with community size or survival as major targets of prediction.

Recent work has also sought more refined metrics, such as measuring *pro-social* behaviors, rather than identifying *all* interaction as positive, community-oriented behavior [12]. This prior work provides important insights, however such HCI methods can be refined through: (1) a more coherent alignment with advances in organizational psychology theory that distinguish between antecedents, SOVC, and outcomes of SOVC; and (2) a community-centered approach that reflects users' actual experiences of SOVC in the design of research methods. These considerations motivate our research questions and goal of creating an empirical framework to guide future community data science techniques.

3 METHODS

Figure 1 summarizes our research approach. We interviewed 21 researchers to understand how they conceptualize and operationalize community on Reddit. We also observed and surveyed 12 different subreddits to gain insight into users' experiences of SOVC. We then used Grounded Theory Method [78] to analyze the data. This study was reviewed and deemed exempt by our institutional IRB office. We begin with a positionality statement to disclose how our own personal identities have influenced our study before describing our methods in detail.

3.1 Positionality & Ethical Stance

We view the increasing research attention on Reddit [91] as an urgent opportunity to ensure that new methods are conscientiously, respectfully, and ethically designed for purposes of service and accountability to communities rather than extraction or manipulation. All authors are regular Reddit users.³ Consequently, we are influenced by our own on- and offline experiences related to our usage of Reddit. Most subreddits were selected somewhat randomly according to criteria we will soon describe. However, our own membership and prior engagement with moderators contributed to our ability to successfully recruit a few of the subreddits in Table 2. For example, the second author is a member of S8. Due to research and personal interests in mental health, she felt it vital to include at least one mental health subreddit. Given an otherwise arbitrary set of candidates, it was most sensible to recruit a familiar subreddit. As academic researchers with values and ethics oriented toward equity and inclusion, we also felt it was vital to recruit subreddits for underrepresented populations (*e.g.*, S6, S7, S11), even though we do not identify as members. We acknowledge that a complex set of privacy, safety, and equity concerns must be addressed in the development of new methods that can measure and impact human communities. Our research team values collaborative efforts with moderation teams and users, and we suggest that future researchers continue to work directly with users and communities, especially ensuring that such research efforts are allowed and desired.

3.2 Interviews with Reddit Researchers

During our literature review, we observed that most papers on Reddit seem to be based on unstated assumptions about subreddits existing as communities. However, literature review does not enable us to make claims about Reddit researchers' actual mental models of community, how these impact selected methodologies, or other unwritten limitations or opportunities they have considered. We chose to interview Reddit researchers in order to ensure the empirical validity and rigor of our claims and to directly compare researcher and user perspectives.

3.2.1 Recruitment. A systematic review of 727 Reddit publications from 2010–2020 shows that 57.4% are first-authored by researchers in Computer Science, Math, and Engineering units, with disciplines

³Three team members, including the first author and two assistants listed in the acknowledgements, were originally recruited to join this research project through a subreddit post.

ID	Career Stage	Discipline	Methods	Reddit User?
R1	Postdoc	CS, Math, Eng.	Quant	prior to research
R2	PhD Student/Industry Researcher	CS, Math, Eng.	Mixed	prior to research
R3*	Industry Researcher	CS, Math, Eng.	Qual	prior to research
R4	Faculty	CS, Math, Eng.	Mixed	prior to research
R5*	Faculty	CS, Math, Eng.	Qual	prior to research
R6	PhD Student	CS, Math, Eng.	Mixed	since research
R7	Industry Researcher	CS, Math, Eng.	Quant	prior to research
R8	Faculty	CS, Math, Eng.	Quant	never
R9	Faculty	Humanities	Mixed	prior to research
R10	Faculty	Humanities	Qual	since research
R11	Faculty	Humanities	Qual	prior to research
R12	Faculty	Humanities	Mixed	prior to research
R13	Faculty	Humanities	Qual	never
R14	Faculty	Medicine & Health	Mixed	prior to research
R15	PhD Student/Industry Researcher	Medicine & Health	Quant	prior to research
R16	Faculty	Social Science	Qual	since research
R17*	Faculty	Social Science	Qual	prior to research
R18	Faculty	Social Science	Mixed	prior to research
R19	Faculty	Social Science	Mixed	prior to research
R20	PhD Student	Social Science	Quant	prior to research
R21	Faculty	Social Science	Quant	prior to research

Table 1. Summary of Reddit Researcher Participants. Most participants indicated that they have published between 1–5 papers on Reddit; * indicates researchers who have published >5 papers on Reddit.

like the Humanities (18.2%), Medicine & Health (9.6%), and Social Science (7.8%) accounting for most of the rest. 482 (66.3%) papers used quantitative methods (primarily computational techniques such as machine learning and natural language processing), 183 (25.2%) used qualitative methods (e.g., discourse analysis, ethnography, grounded theory), and 56 (7.7%) used mixed methods [91]. We manually collated a list of email addresses for all first and last authors in this set of papers, on the assumption that first and last authors are most influential in methods decisions. We emailed the list a recruitment message and screening survey and selected researchers across disciplines, methods, and seniority, purposefully ensuring representation of as balanced and diverse a collection of voices as possible. We continued recruiting until we reached data saturation—i.e. new interviews were no longer providing new information [78]. Table 1 summarizes participant information of 21 participants: 13 faculty, 1 postdoc, 2 industry researchers, and 4 Ph.D. students (2 of whom also indicated industry researcher). 7 participants (33.3%) self-identified Computer Science, Math, and Engineering (including HCI) as their primary discipline, 6 (28.6%) Social Science, 5 (23.4%) Humanities, and 2 (9.5%) Medicine & Health. Of these, 6 primarily used quantitative methods, 7 qualitative, and 8 mixed. 14 participants (66.7%) identified as white, 6 as Asian (28.6%), and 1 as Middle Eastern. 16 participants (76.2%) were at a US institution, 4 in Europe (19%), and 1 in Asia. 13 identified as male (including 1 trans) (61.9%), and 8 female (38.1%).

3.2.2 Interview Protocol. Interviews were conducted remotely over Zoom by the two co-first authors. Interviews ranged from 43 to 73 minutes long and lasted an average length of 56 minutes. Interviews had two parts. The first used a semi-structured approach with questions regarding

the researchers' mental model of community, what a strong community might look like, specific definitions/operationalizations of community, and methods decisions in the interviewees' work. (The complete set of interview questions is available in supplemental section 3.) The second part used a talk-aloud structure (similar to [41, 109]) in which we screen-shared slides, asking participants to reflect on four known antecedents to SOVC (interactivity, membership boundaries, homogeneity, and norm enforcement). We defined each antecedent on-screen and asked how researchers might measure it using any qualitative, quantitative, or mixed methods. A final slide asked about any additional variables or ideas we had not asked about. Next, we describe why we chose these four particular antecedents.

3.2.3 Selected antecedents to sense of virtual community. Organizational psychologists have demonstrated interactivity, membership boundaries, homogeneity, and norm enforcement to be fundamental antecedents to the formation of SOVC [16]. This list is not exhaustive, however, it does provide a strong theoretical bedrock.⁴ We opted for more generic rather than overly-specified constructs to allow for the broadest possible interpretations of user behavior. For example, we asked about *interactivity* rather than *support exchange* or *information sharing*, since both of those are more specific forms of interactivity which may or may not be *required* for SOVC in different virtual contexts. We displayed the following four definitions to participants:

Interactivity: The degree to which users are interacting within the community.

Membership Boundaries: The degree to which users can be considered members of a community, *v.s.* not being considered members of a community.

Homogeneity: The degree of similarity of users related to their values, attitudes, goals, or other personal characteristics (*e.g.*, demographics, personality traits, etc.).

Norm Enforcement: The degree to which norms are enforced within the community.

3.2.4 Interview limitations. One limitation is that we could not interview researchers who directly studied the *same* subreddits we surveyed, however, we addressed this limitation by ensuring that diverse subreddits were included in our surveys, including the same basic types studied by prior researchers. Standard interview limitations apply: (1) Our small sample of Reddit researchers may not be representative of all researchers; and (2) We asked researchers to describe prior work, some of which was completed years ago, and recall of past events is imperfect.

3.3 Surveys of Reddit Users

Surveys were run in collaboration with the moderation teams of twelve different subreddits (See Table 2).⁵ In addition to the reasons mentioned above in sec. 3.1), we selected subreddits using the following criteria. We avoided NSFW (Not Safe For Work) and hate- or abuse-centered subreddits for two main reasons: (1) preserving the safety of our research team; (2) focusing our research efforts on *pro-social* communities—*i.e.*, the types of communities that we hope this work can support. We aimed to capture as diverse a collection of subreddits as possible, with high variability in size, topic, and observable community activities (*e.g.*, weekly or daily threads, annual customs, special AMA events, different types of user flairs or threads, etc.) to ensure that our results are as broadly applicable across Reddit as possible. We also chose *active* subreddits in which surveys were allowed and in which users appear to post every day (rather than those with obviously inactive userbases)

⁴See supplemental section 2 for in-depth discussion of the theoretical bases for these antecedents.

⁵This paper uses the same survey data collected in [105], which performs a purely statistical analysis on Likert rating data for psychometric scale validation. The present paper uses and analyzes qualitative data from the surveys in order to understand new dimensions of sense of virtual community. Complete survey text is available at https://bit.ly/gov-bots_pdf.

ID	Description	Size	N	Community Archetypes
S1	Posts with niche absurdist humor style	10-50K	16	ContentGen
S2	Corrections of poor scientific reporting	10-50K	28	Learning
S3	Discuss fundamental elements of music	100K-1M	77	Q&A; Learning
S4	Help with job application materials	100K-1M	48	Q&A; ContentGen
S5	Support for pregnancy-related issues	<10k	39	Support
S6	POC/allies discuss race/intersectionality	10-50K	41	Learning; Support
S7	Discuss ironic matters regarding race	<10k	21	Learning
S8	Support for a mental health issue	50-100K	13	Support
S9	A US university and geographical area	10-50K	43	Affiliation; Q&A
S10	Q&A with experts on an academic topic	1-5+M	25	Q&A; Learning
S11	Screenshots of social media posts from POC users	1-5+M	250	ContentGen; Learning; Support
S12	Learning about indigenous spiritual beliefs	50-100K	7	Learning; Support

Table 2. Descriptions of subreddits included in this study. Abbreviations included in the Community Archetypes column are fully specified in Section 5.2; Archetypes are listed in order of highest to lowest relevance. Note: “POC” is an acronym for People of Color.

to maximize response rates. We posted surveys to each subreddit using a method preferred by each moderation team, including strategies such as:

- Moderators create a post about the survey and pin it to the top of the subreddit.
- Researchers post the survey in a weekly thread for surveys and/or personal promotion (and mods “approve” or highlight the post).
- Researchers post the survey to a thread specifically for surveys in that subreddit.

Survey links were posted from October through December of 2021, and each link was open to responses for a maximum of two weeks; some surveys were closed before that time due to satisfactory response rates. Survey participants could provide their email to opt into a drawing for \$10 eGift Cards. In addition to basic demographics, surveys included three free response questions that provide insight into users’ experience of SOVC with an eye toward RQ2 and RQ3:

- What motivates you to visit `r/[subreddit name]`?
- What makes `r/[subreddit name]` feel like a community to you?
- What (if anything) could be done to improve your experience in `r/[subreddit name]`?

3.3.1 Survey participants. We manually inspected all survey responses and eliminated those which contained nonsensical or copy/pasted free responses. In total, we collected 608 valid responses across the twelve subreddits in Table 2. Respondents skew male (58.8%), white (61.5%), residing in North America (81%), and relatively young (50.1% of respondents selected 25–34 years of age, and another 21.3% 18–24); these demographic trends are broadly consistent across the Reddit userbase.

3.3.2 Survey limitations. The most important survey limitation lies in our selection of subreddits. Since every subreddit can have unique rules, norms, interaction styles, bots, *etc.*, it is impossible to select a perfectly representative set. For instance, many subreddits do not allow posts on unrelated topics, including research advertisement, thus a systematic difference might exist between the types of communities that allow research *v.s.* those that do not. (For example, perhaps norm enforcement is more tightly coupled with SOVC in subreddits that do not allow research.) We intentionally selected diverse subreddits but may have inadvertently missed certain types of communities; future work should periodically revisit our research questions to enrich and update our understanding of the community types we will describe, as well as appending new types we may have missed or which may emerge in the future.

In order to encourage respondents to finish the survey, we limited survey length and did not ask many detailed questions—*e.g.*, about each antecedent individually. We note that respondents provided responses of varying length and detail and that these responses *organically* included information related to antecedents, SOVC, and other community-related concepts. Standard survey limitations apply, including the possibility of misalignment between self-reported data and actual behaviors and opt-in selection bias. Our survey advertisement was posted with different levels of visibility in different subreddits, ranging from high visibility (*e.g.*, pinned post) to lower visibility (*e.g.*, survey thread). In general, people who elected to take the survey may or may not be representative of the user base of each subreddit. However, respondents from subreddits where there was lower visibility of the recruitment post might be less representative due to finding our study in a more niche way, while respondents who found the survey in a more visible location might be more representative.

3.4 Observation of subreddits

We visited every subreddit in our sample to observe how users are currently interacting. Rather than programmatically collecting top-voted posts, we scrolled through at least 20 recent posts, along with the comments on those posts, in their order of appearance on the user interface. This technique provides us with a sample of the data near the time when surveys were collected and avoids biasing our understanding of the content toward all-time popular posts (which may not accurately capture the day-to-day diversity of content and interactions on the subreddit). We wrote memos and noted the types of topics users wrote about and whether there were any particular formats or patterns in the content.⁶ We also conducted a preliminary analysis of the types of the rules present in each subreddit to inform our observations; the details of this tangential analysis are included in the supplemental materials.

3.5 Analysis

We used Grounded Theory Method (GTM) [78] conducted virtually by the two co-first authors. For the interviews, we used transcripts automatically generated by Zoom as a basis, and then corrected errors while re-listening to recordings. We wrote memos and notes during data collection and transcript correction in order to capture our immediate reflections and ideas about the data. Sensitized by concepts from prior literature, while also allowing new concepts to emerge inductively from the data [29], we then systematically open-coded all interview transcripts. Over the course of approximately two months, the two co-first authors met weekly to analyze and collaboratively cluster all open codes, continuously discussing any uncertainties or disagreements until both authors agreed upon a final clustering arrangement and placement for each individual code. We conducted this affinity mapping using Miro software to first cluster all interview open codes into major axial themes, each with several sub-themes. Similarly, we next open-coded all survey responses and clustered the survey codes in a separate space on the same Miro board. After clustering the interview and survey codes independently, we then iterated upon our clusters and rearranged the Miro board to highlight overlapping ideas between the two groups. Overall, this process generated a total of 1,149 open codes, which we organized into 45 major axial themes with 83 sub-themes. We report on approximately half of these themes/codes that directly address our research questions on SOVC by synthesizing them within the framework presented by this paper.

Importantly, a variety of users' ideas generated distinct clusters that had not emerged prominently in our mapping of researchers' data—thus leading to our conceptualization of these clusters as five

⁶We considered completing a formal, systematic content analysis of posts and comments, however this degree of precision is not required for our research questions, which are better addressed through analysis of survey responses.

“Community Archetypes” (see Table 4). Upon closer inspection of these clusters and the participant data they were derived from, our discussions inductively revealed that there were clear user groups specific to each archetype for which we assigned the title of “roles;” these roles directly led to the interaction patterns which resulted in users’ experiences of SOVC. Finally, we cross-referenced the archetype clusters with our subreddit observations in order to assign each subreddit in our sample to one or more archetypes (see Table 2). The two co-first authors conducted every step of this analytical work together, and then discussed and edited our presentation of results with all authors for clarity and concision.⁷

4 RQ1: RESEARCHER CONCEPTIONS OF COMMUNITY & HOW TO MEASURE IT

This section seeks to address two central lines of inquiry: (Section 4.1) When can researchers understand whether a particular subreddit is experienced as a community or not?; (Section 4.2) What types of behavioral data traces may indicate how and whether users are experiencing SOVC (or not)—either within a particular subreddit (or across Reddit more broadly)?

4.1 Researcher conceptions of Reddit communities

Despite frequently calling subreddits “*communities*,” our participants have not specifically analyzed whether users experience subreddits as communities. We asked whether they draw upon any literature on the theory or praxis of online communities to inform their understanding and methodological approaches; 8 out of 21 participants (1 in computer science, 3 in humanities, and 4 in social science) mentioned at least one paper (e.g., [88]). However, in most cases, researchers have made a certain set of assumptions to frame and constrain their studies, and have used anywhere from one to thousands of subreddits as a source of data for their projects. Although some researchers preferred terms like “*forum*” or “*discourse/linguistic setting*”, consensus across participants’ responses indicates that it is reasonable for researchers to refer to all subs as “*topical affinity groups*” or “*common interest groups*.”

Moreover, most researchers made it clear that they do not believe all subreddits are actually communities. As R12 noted, subreddits are simply the “*neutral descriptor of the actual digital space*.” In order to qualify as a community, there must exist some additional set of criteria beyond the content alone, such as: participation in conversations or communication networks, formation of meaningful relationships between users (even occasionally extending into off-platform activities), users’ sense of belonging in what they perceive as a community space, and willingness to act on shared values. Many researchers disagreed with a binary approach to whether a subreddit is a community, instead favoring a spectrum model where “*you can have more or less [community] and there’s no threshold*” (R21). Others provided further nuance, pointing out that a subreddit could be one big community, but it might also contain “*pocket communities*” (R12)—i.e., smaller communities within larger spaces which do not encompass every user. Although researchers agreed that deciding how to measure or delineate community is quite challenging and imperfect, the next section offers synthesized insights toward considering and assessing SOVC antecedents.

4.2 Researchers’ strategies for assessing SOVC antecedents

Here, we summarize researchers’ suggested strategies for understanding and measuring four antecedent constructs to SOVC: interactivity, membership, homogeneity, and norm enforcement. Although this portion of results is largely recapitulating known methods, it should function as a compact and

⁷We collected consent to participate in this research according to our exempted IRB protocol, however we did not request permission in advance from participants to make interview transcripts and survey data publicly available. Our researcher data would be impossible to de-identify given the nature of our questions. If future researchers are interested to re-analyze these data, our research team will gladly provide assistance in submitting appropriate IRB protocols for re-use of the data.

Construct	Assessment Strategies
Interactivity (sec. 4.2.1)	<ul style="list-style-type: none"> • Volume of observable interaction • Degree of conversation occurring
Membership Boundaries (sec. 4.2.2)	<ul style="list-style-type: none"> • Tiers of membership • Prolonged participation • Linguistic membership boundaries
Homogeneity (sec. 4.2.3)	<ul style="list-style-type: none"> • Demographic • Situational • Goals or values • Linguistic similarity
Norm Enforcement (sec. 4.2.4)	<ul style="list-style-type: none"> • Vertical moderation (by moderators) • Horizontal norm enforcement (by users)

Table 3. Assessment strategies for antecedent constructs to sense of virtual community.

useful reference point for researchers who would like to design methods in more conscientious alignment with the Community Archetypes we will soon introduce in section 5.

4.2.1 Interactivity. We identified two major ways that researchers can quantitatively operationalize interactivity, the first of which is more common in the literature despite being less relevant to users' experience of SOVC than the second:

- (1) **Volume of observable interaction:** “A community is nothing without the interaction of the users.” (R6) In line with this, researchers often measure the level at which users participate—e.g., the aggregate volume, rate, and topic matter of posts, comments, and votes. An important caveat is that the volume of interactivity alone is an overly blunt instrument. A certain level of interaction is necessary, yet insufficient, to indicate SOVC.
- (2) **Degree of conversation:** Researchers agree that there must be “some robust, sustainable form of communication” (R16) in order for a subreddit to qualify as a community. People participating in communities are more interactive and discursive than their non-community counterparts; the way to measure that difference is to look for people exchanging back and forth, rather than counting individual comments without interaction between commenters. Researchers should look to identify nested comments and/or repeated user interactions across multiple comments.

4.2.2 Membership Boundaries. Deciding how to count users as *members* of a community, versus those who are not, is often a pivotal methods decision with important implications for results.

- (1) **Membership tiers:** Participants often assign users to tiers based on: the amount of individual actions taken by the user (from complete inaction (lurking) up to moderating or engaging heavily with the moderation team to advocate for change); the number of posts, comments, and/or amount of time a user spends on the subreddit; or directly asking how users view their position in the community (e.g., in surveys, interviews).
- (2) **Prolonged participation:** Users who return repeatedly to a subreddit understand and develop community culture. These core members help a subreddit maintain its distinctiveness and attract newcomers. Researchers can identify these users by searching for users that post and/or comment a disproportionate amount and have been active on the subreddit for a prolonged amount of time.

- (3) **Linguistic membership boundaries:** Tailored uses of language are also helpful for differentiating members *v.s.* non-members. If users can comfortably use the specific language of a given subreddit, then they have both been there long enough to be familiar with the vernacular, and they are contributing relevant content. Users exhibiting markers such as in-jokes, industry- or topic-specific terms, or particular abbreviations, spellings, or even new words that do not tend to appear elsewhere, are more likely to be members. Researchers also may evaluate whether a user adheres to the style of discourse, including tone, shared opinions, and overall sentiment.

4.2.3 *Homogeneity.* Measurements of homogeneity will be highly dependent on the researchers' purposes for the project, and the type of subreddit being studied. We discuss four types of homogeneity in ascending levels of perceived utility or accessibility for research.

- (1) **Demographic:** Unless demographics (*e.g.*, race, age, gender, education, *etc.*) are directly related to the research question, researchers warned against collecting or inferring too much information from demographic homogeneity because there are, likely more often than not, hidden confounds behind demographics than can lead researchers toward faulty claims if they attribute results to demographics rather than the underlying confounds.
- (2) **Situational:** Situational homogeneity considers the user's shared life experiences or situational circumstances in relation to others in the subreddit. Situational homogeneity can sometimes be related to demographics, but it is a more powerful and unifying concept, and is especially important in subreddits focused on events in one's personal life.
- (3) **Goals or values:** Homogeneity in shared goals/values can illustrate why people are in a subreddit and what they hope to accomplish, and is of very high utility for research. It can deviate from other kinds of homogeneity because it is often better to have people with quite different identities and life experiences coming together under a certain goal.
- (4) **Linguistic similarity:** The degree to which a single user's language use adheres to group-level patterns may be useful for indicating their own *individual* membership status. However, assessing the degree to which *groups* of users use similar language may indicate a form of linguistic homogeneity that can be directly calculated from trace data.

Given the absence of demographic information on anonymous accounts, and an inability to accurately infer users' personal situations or goals/values, surveys or interviews should always remain the gold standards for assessing those three types of homogeneity. Qualitative methods with smaller sample sizes can be useful for first deriving the types of personal situations or goals and values that are most relevant to later include in surveys. Ethnographic observation of posts and comments over a period of time can also help to determine what types of situational or shared goals/values exist. Linguistic homogeneity, on the other hand, does not rely upon users' internal states, and may therefore be a convenient metric for studies that cannot directly query users.

4.2.4 *Norm Enforcement.* Norms vary substantially across Reddit, as do the ways they are enforced. Researchers described two broad ways of assessing this.

- (1) **Vertical:** Vertical norm enforcement refers to top-down moderation by appointed moderators. Researchers suggested quantitative methods for assessing the degree of vertical moderation occurring, such as counting the number of moderators and their volume of activity in the sub⁸, measuring the proportion of removed content, frequency of banning users, and volume of activity by governance bots, or observing how often rules are modified.

⁸Note that moderators can also participate more casually, such as by adding their commentary to a discussion, apart from official moderation. A refinement of this strategy requires classifying the purpose of moderators' messages.

- (2) **Horizontal:** Although users have less power than moderators, they can nonetheless enforce norms horizontally. They can reply to others' posts or comments to point out norm violations (e.g., inappropriate behavior or irrelevant/disallowed content, sometimes even suggesting an alternate subreddit) or report issues to moderators. One strategy for assessing this could be to develop classifiers for identifying this behavior and assessing its frequency. A lack of engagement (e.g., posts receiving no comments) may indicate content that doesn't align with subreddit norms. Finally, up/downvoting might also be an enforcement mechanism, although the meaning of a vote is subjective and not universal, making it an unreliable measure.

Having described researchers' concepts of community on Reddit, and their ideas for evaluating the constructs of interactivity, membership boundaries, homogeneity, and norm enforcement, we will next explore users' perspectives. By juxtaposing researchers' ideas with users' experiences, we can refine and specify assessment strategies for a more human-centered understanding of SOVC.

5 RQ2: USER EXPERIENCES OF SENSE OF VIRTUAL COMMUNITY

In this section, we present our analysis of our surveys of Reddit users. We find evidence of several broad *qualities of community experience* that are likely to apply across any community. These qualities align closely with prior work and researchers' ideas; therefore we start by briefly highlighting this overlap (sec. 5.1). Importantly, we also find that users attribute their SOVC to specific aspects of community experiences that are closely related to specific *types* of subreddits. Researchers' comments largely did not capture these ideas, thus we introduce the framing of *Community Archetypes* and focus most of this section on detailed characterizations of the core ways in which users experience SOVC differently across different archetypes (sec. 5.2).

5.1 General qualities of community experience

There are three major ways users' responses align with approaches suggested by both our researcher participants, as well as by prior work in the social sciences and HCI:

- (1) **Socializing** is a major factor that improves SOVC; users frequently referenced building meaningful relationships through discussions on threads, iterating on each other's ideas or jokes, and engaging in other subreddit traditions, offline activities, or affiliated online spaces. This user data strongly validates researchers' discussions of the importance of conversation between users rather than volume of interaction alone (sec. 4.2.1).
- (2) Users referenced **regular participants**, or other users with special flairs as core members. Aligning with prior work and researchers' comments on prolonged membership (sec. 4.2.2), it is clear that the visible activities of core community members indicate SOVC.
- (3) Users pointed to **shared experiences, identities, mindsets, values, or views** that bind them together as an in-group. Their comments closely mirror researchers' discussions of homogeneity (sec. 4.2.3) and suggest that demographics are not especially relevant, however shared situational circumstances, and shared goals or values, are essential.

We also found that users' perspectives highlighted how the specific configurations of affordances, rules, norms, and behaviors within specific types of subreddits caused them to experience SOVC in unique ways, leading us toward the concept of Community Archetypes.

5.2 Community Archetypes

Prior works such as [101, 120] have introduced the language of archetypes to describe conceptual frameworks for how an abstract set of qualities may apply across individual instances. Rather than hard classification categories, boundaries between archetypes are fuzzy, yet each nonetheless retains a unique flavor that may manifest across various identifiable subreddit features and user

Archetype	Roles	Content Patterns
Sec. 5.2.1 Topical Q&A (Q&A)	<ul style="list-style-type: none"> • Expert • Novice 	<p>Posts: questions</p> <p>Comments: answers, discussions and enrichments of others' answers</p>
Sec. 5.2.2 Learning & Broadening Perspective (Learning)	<ul style="list-style-type: none"> • Insider • Outsider 	<p>Posts: news, events, publications, personal stories, questions on a particular topic</p> <p>Comments: celebrating or disagreeing, making jokes, contrasting or similar personal stories, elaborative ideas</p>
Sec. 5.2.3 Social Support (Support)	<ul style="list-style-type: none"> • Support seeker • Supporter 	<p>Posts: personal experiences, sensitive disclosures, health questions, milestone announcements, reflection, venting, resources</p> <p>Comments: support, validation, reflection, commiseration, resources</p>
Sec. 5.2.4 Content Generation (ContentGen)	<ul style="list-style-type: none"> • Producer • Consumer 	<p>Posts: original content from a sub member, or contributions from other users/platforms that exemplify a specific content style</p> <p>Comments: Opinions on the content, extra information, commiseration</p>
Sec. 5.2.5 Affiliation with an Entity (Affiliation)	<ul style="list-style-type: none"> • Current affiliate • Prior affiliate • Future affiliate 	<p>Posts: entity-specific news, events, and questions</p> <p>Comments: Feelings about news or events, answers or advice about the entity</p>

Table 4. Summary of Community Archetypes. We occasionally refer to these archetypes by the (parenthetical abbreviations) included in the first column.

behaviors. Whereas the authors of [101, 120] apply the term archetype to individuals, we use it to reference *communities*. This terminology thus helps us to understand how abstract community characteristics lend themselves to certain roles and functions in a user's content consumption or other forms of participation. Given that subreddits are the selected unit of analysis for this paper, we apply these archetypes as labels to individual subreddits, while also suggesting that future work should apply them to other units of analysis on Reddit—e.g., groups of subreddits, or subgroups of users—or moreover, on other platforms that host online communities.

Our work reveals five community archetypes (Table 4). Multiple archetypes may overlap in any subreddit, however, there is typically a more distinct focus on one over others. For example, Table 2 designates which archetypes best match each subreddit in our study, and shows how some embody one major archetype, whereas others have multiple. Next, we describe each of these archetypes in detail, specifying important **user roles**, **content patterns**, and **rules** associated with each. Table 4 includes an overview of archetype roles and content patterns.⁹

5.2.1 Topical Q&A. Some subreddits are explicitly set up with strict rules to create a Q&A format around specific topics. For example, P335 declares that S10 is “*the best place on the internet to find reliable, well-sourced, in-depth information on various [academic discipline] topics.*” In our sample of Q&A subreddits, 55.2% of rules relate to content/behavior (mainly around respectful

⁹Section 4 in the supplemental materials includes more detailed results of our analysis of subreddit rules.

engagement with others and following the Q&A format), with advertising and commercialization (which allows for users to ask for help without being targeted as consumers), low-quality content (which encourages generating content that is useful for the subreddit), and off-topic (which keeps content relevant) rules being the next largest categories, accounting for 10.3% each. We observe the content pattern that most top-level posts are questions, while most comments are answers, or discussions and enrichments of others' answers. There are two clear user roles: **novices** (those who ask questions) and **experts** (those who answer them). These groups may possibly overlap—*i.e.*, a user who usually answers questions might occasionally ask one, or vice versa. Our survey respondents typically indicated their motivations for using the subreddit in an either/or manner; the distinction between these two roles appears to be rather sharp. Expert users enjoy helping people and sharing their knowledge about a subject they are passionate about, while novices sincerely appreciate the special access to experts and the information they can provide. For example, P333 visits S10 to “*provide answers where I have specific knowledge or expertise,*” whereas in S3, P53 enjoys that “*questions can always be answered constructively.*”

Users explain that SOVC arises when the community is effectively fulfilling its Q&A function, providing rich, multi-faceted discussions on current topics, while also generating a useful archive of past inquiries and information. For example, P337 appreciates the collaboration and diversity on S10 when “*people with different training and background will often read the same question in a different way and bring different perspectives.*” However, some users, such as P163 mentioned that they don't feel SOVC in S4 because it isn't “*used in a way that could foster a community...it's closer to a help or Q&A board.*” In general, we posit that each user experiences her own affinity for particular forms of interaction or community; there is no such thing as a one-size-fits-all online community, and it is impossible to claim that SOVC will always arise for every user when *X* phenomenon occurs. However, across many users, our results suggest that SOVC is more *likely* to arise when *X* phenomenon occurs, if *X* is consistent with its associated archetype.

5.2.2 Learning & Broadening Perspective. As our most commonly observed archetype, users frequently mentioned learning or broadening their perspective as a primary reason for visiting a subreddit. Content patterns are less formulaic than Q&A subs, however top-level posts are often pointers to current events, publications, or relevant news, or relatable personal stories, experiences, and questions. Comments tend to elaborate upon the ideas raised, for example by celebrating or disagreeing, making jokes, providing contrasting or similar personal stories and experiences, or adding new ideas and references into the mix. These subreddits have a central focus around a particular culture—not only for people *within* that culture (who benefit from shared experiences of it), but also for users who are *not* part of that culture themselves, yet appreciate it and want to know more. For example, P433 (S11) said, “*I want to hear new and/or different opinions on various topics important to the [redacted] community. It helps me grow my allyship.*” Given this cultural focus, we refer to two roles as **insiders** and **outsiders**. For example, on S2, insiders are scientists or people with extensive scientific background and training, whereas outsiders are curious and engaged members of the lay public. On S6, insiders are members of historically marginalized racial groups, whereas other users (mostly white) self-identify as outsiders who want to learn how to be better allies. It may be that this less formulaic, diverse structure is the reason that this category of subs has one of the highest instances of content/behavior rules (62.5%). Guidance around treating others well and generating content that serves the purpose of the subreddit allows for the content to be more free form while still cultivating a desirable space. They also have higher rates of rules about harassment and hate speech (15% each) which allows for insiders and outsiders to interface respectfully and in a way that fosters a learning environment.

Users in these subs report feeling SOVC because of shared mindset or values. For example, P609 (S12) said, *“I think the users of this subreddit are a community because of similar ideas of approach to spirituality.”* Having one big similarity that holds users together allows users to learn from each other’s differences. For another example, in S2, users share a value system that places high importance on correct reporting of academic subjects. S2 users have differing expertise and levels of subject familiarity. Therefore, they can inform others who may not have the same background, but who also care deeply about information accuracy. However, some users do not feel SOVC on these types of subreddits because their *“interest in the subreddit is in the content and information more than the interpersonal relationships”* (P341, S10). Such users may not develop SOVC, even if others do, because they are simply not looking for it.

5.2.3 Social Support. Much prior work has examined social support behaviors in online communities [3, 4, 106]. At a higher level, our work suggests a social support community archetype, in which socially supportive behaviors are the main purpose and organizing principle for gathering. Many users are seeking spaces that can help them navigate their own difficulties or illnesses; others are seeking information and resources to help a loved one who is struggling. For example, P189 said that S7 helps them *“get a better insight as to how to support [their loved one] and what not to do.”* Rules in these subs seek to protect users and keep the conversation focused on the issue, with 61.7% regulating content/behavior to create a subreddit that is safe and focused, 14.9% warning against off-topic content (*esp.* content that is unhelpful or triggering), and 10.6% disallowing hate speech in order to create a safe space. Top-level posts are often specific personal experiences or sensitive disclosures, questions about a health issue, announcements of milestones, reflections or venting, or sharing of resources, artwork, and encouraging thoughts and memes/jokes. These subs often contain flairs that specify content warnings, symptoms, or labels of users’ intent (*e.g.*, “seeking support” or “venting”). Comments generally offer support, reflection, commiseration, resources, and validation. In this archetype, roles include **support seekers** and **supporters**, yet the distinction is fuzzy and inconsistent, with heavy overlap in how users occupy them. In one post, a user might seek support. In the next, she might become a supporter, returning the kindness she received from others. It is certainly possible to build classification schemes that label individual posts and comments as support seeking or supporting (and sometimes both), however it would be difficult to accurately reflect the role in a static manner per user; it shifts continually.

Users report SOVC when they see their own hardship reflected in other’s posts and the community provides true kindness, support, and validation. P238 (S8) said, *“it feels like my experience was real and valid.”* P189 also said their support community, S5, offers *“sympathy that no one but them can provide properly and meaningfully”* due to other users’ intimate familiarity with the issue. Users call these subs a safe space to express feelings, be themselves, and get meaningful help. A few respondents also recognized that they didn’t always use these subreddits in a healthy way. P277 often visits S8 for support, but sometimes they *“want to force myself to get over it by reading about people who have it worse,”* a behavior they do not view as healthy. It is possible to stay too long in support subs, thus promoting lingering in painful memories rather than recovery (unless the user is intentionally returning to help others). It is important to understand when SOVC is beneficial to well-being versus when it could be holding users back.

5.2.4 Content Generation. Users who visit ContentGen subreddits are interested in particular types of content with a certain sense of humor, viewpoint, format, purpose, or type of expression. For example, P4 appreciates that S1 has *“the expectation that you will ‘act oddly.’”* In this archetype, user roles include **producers** and **consumers**, where producers create top-level posts that exemplify that sub’s specific content style, and consumers are specifically there to view and respond to it. Top level posts might be created by original posters themselves, or they may be contributions from

other users or platforms that the poster would like to share or discuss on the subreddit. Comments often include peoples' opinions on the content, extra information about the content, or sometimes commiseration with the original poster. As a result of this content-focused culture, content/behavior is the subject of 62.5% of rules, low-quality content is 16.7%, and hate speech and advertisements & commercialization are 12.5% each, both of which serve to focus users solely on the content rather than things like ads or hate speech.

Some users expressed feeling SOVC due to the niche nature of the content which brings people together because, as P271 (S11) states, *"We share the same sense of humor and appreciate the same content."* Another user, P6 illustrates how niche subreddits like S1 provide a kind of camaraderie that doesn't exist for them in real life: *"It's like finding out there's a 'splashing ice-cold water in people's faces' club after years of failing to convince your family and friends that being splashed in the face with ice-cold water is fun."* However, similarly to the Learning & Perspective Broadening archetype, users may not feel SOVC if they are simply using the subreddit as a resource rather than looking for community. P142 (S3) illustrated this point, saying, *"It feels much more transactional to me... I do not frequently see updates or multiple posts from a given user."*

5.2.5 A Iliation with an Entity. Some subreddits have explicit affiliations with particular entities, such as geographical places or organizations (cities, universities), popular media (book or fan series, TV shows), sports teams, *etc.* In our limited data set, we only observed this archetype once, in affiliation with a particular geographical area and university campus (S9). Consequently, we observed user roles of **current**, **prior**, and **future** campus residents. However, we will refer to these roles more broadly as **"affiliates."** In our usage of Reddit beyond this sample, we have informally observed these affiliate roles across different types of entities; broadly, they can be defined as a user's level of active, ongoing knowledge and investment in a particular entity. Current affiliates have a good grasp on up-to-date "local" knowledge by virtue of ongoing attention to or residency within the affiliated entity. Future affiliates are more like "prospective community members" who are curious, while prior affiliates enjoy retaining a connection even though they are no longer focusing as much resource or attention. In terms of content patterns, posts often feature local or breaking news and events related to the entity. Future and current affiliates often pose questions, while past (and other current affiliates) tend to answer them. In the comments, users generally express their feelings about news or events, and answer questions or offer advice and opinions about the entity. There is notably less regulation on content/behavior than in other archetypes (40%), possibly because it is already constrained by the geographical affiliation (however, this may also be because only one subreddit in our sample fits this archetype).

In S9, the specific tie to a geographical entity makes users feel SOVC. P286 (S9) said, *"[City] as a whole has always been a community to me and this is a microcosm of that community."* Users such as P307 (S9) also appreciate *"people sharing their experiences and advice in a genuine and helpful way,"* because it's likely to be applicable due to users' geographic closeness. Some users felt it was easier to get good information from the subreddit because the city and university administration did a poor job of circulating it. Thus, the subreddit seems to supplement the offline space it is tied to. As P297 put it, *"everything posted there [S9] has to relate to [US university], which helps me feel more connected to the offline campus community."* In fact, P305 also noticed that *"some memes from [S9] leak onto campus every now and then."* However, several users stated that this kind of subreddit isn't the place they would go to feel SOVC. P297 (S9) said, *"If I wanted to feel like part of a community, I wouldn't be looking for it on a subreddit for a large state school."* Another user, P310, remarked that a subreddit like S9 is *"more like a collage of experiences"* than a community.

6 DISCUSSION

In this paper, we set out to understand researchers' mental models of community, juxtapose these models against users' experiences, and produce recommendations for refining research methods to better align with users' experiences of participating in online communities. Using the Reddit platform as a case study, we found that researchers tend to view subreddits as topical affinity groups, however each subreddit individually sits somewhere along a spectrum of "community-likeness." This result evokes work by Bruckman, who has suggested that groups be categorized as communities on a spectrum based on how closely they resemble a baseline or "prototypical" concept of a community [20, 21]. However, that introduces the problem of what exactly the baseline idea of a community should be. Our work suggests that there is no single prototype that can describe all communities. Rather, our analysis reveals at least five *Community Archetypes* that can be embodied by online communities to varying degrees: Topical Q&A, Learning & Broadening Perspective, Social Support, Content Generation, and Affiliation with an Entity (see Table 4). Each archetype is distinguished by specific behaviors involving user roles, content patterns, and group rules, and users are most likely to experience a sense of virtual community (SOVC) when these behaviors are *strong* and *frequent*. Although we will use Reddit-specific terms and affordances throughout much of our discussion, we will conclude with reflections on how we expect these archetypes to persist across other online social platforms, given the fundamental psychological and social needs shaping human behavior.

Our results suggest that subreddits can embody one or two archetypes, however, no single subreddit can fully embody *all* archetypes. To illustrate this, we will compare two popular subreddits: *r/aww* and *r/wallstreetbets*.¹⁰ Figure 2 depicts screenshots of the subreddits, while Figure 3 contains a hypothetical spiderplot depicting how Community Archetypes may be present within them to differing degrees. *r/aww* was a "default sub" until 2017 when Reddit discontinued this automatic subscription mechanism for new users [94] and is described as "*a place for really cute pictures and videos*." Most rules are restrictive against any other content, and every single post features a picture or video of a cute animal. Thus *r/aww* appears to be well-described almost exclusively by the Content Generation archetype (possibly with a hint of Learning & Perspective Broadening). In *r/wallstreetbets*, on the other hand, we observe a dominant affiliation with finance and stock trading.¹¹ The community description ("*Like Achan found a Bloomberg Terminal*") is more cryptic than *r/aww*, and the rules focus on a variety of activities that are tolerated (or not) related to stock purchases. Content patterns are strongly evocative of the Affiliation and Learning archetypes, often revolving around discussions of specific stocks and whether users should "bet" by buying risky stocks for humor, thrills, or financial gain. There is also an edgy flavor of Social Support in which users who lost large amounts of money on failed bets are simultaneously made fun of, consoled, and glorified. Although a few posts resemble Q&A or ContentGen, these are not nearly as consistent as in *r/aww*. These examples highlight how some subreddits have a single or dominant Community Archetype, whereas others have more hybrid, flexible, or nuanced Community Archetypes.

¹⁰We picked large and well-known subreddits in order to illustrate how our framework can impact methods decisions in future studies. We chose these examples because they were *not* included in our sample and they are already well-known among Reddit users, reducing possible exposure of less well-known subreddits to unwanted attention [38, 107].

¹¹"Wall Street" is a metonym using the road in Manhattan where the New York Stock Exchange is located to refer more generally to financial markets. *r/wallstreetbets* became influential in 2021 due to driving massive upswings in prices of "meme-stocks." [11, 73]

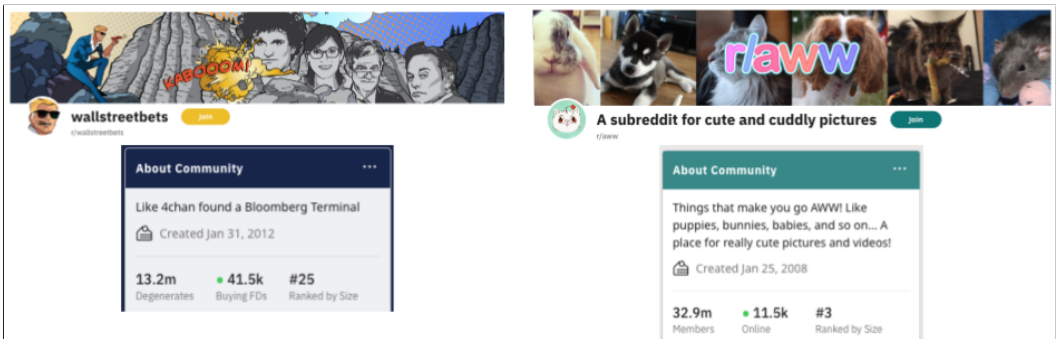


Fig. 2. Screenshots of banner images and community descriptions at r/wallstreetbets (left) and r/aww (right). Retrieved in December 2022.

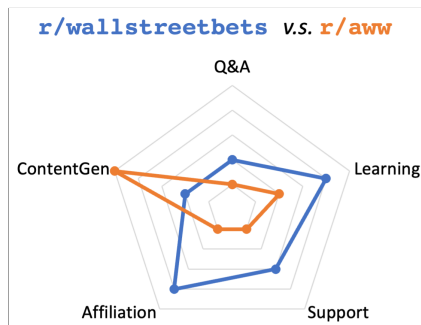


Fig. 3. Hypothetical spiderplot comparing Community Archetypes in r/wallstreetbets (blue) v.s. r/aww (orange) for purposes of illustration only. Rather than factual graphs, these plots are suggestive illustrations based on our estimated observations of the proportional frequency of visible content patterns and user behaviors. The outermost ring represents 100% frequency, with concentric rings descending to an innermost ring of near 0% frequency. Future work could systematically calculate such plots by tabulating the actual frequency of posts and user behaviors in each archetype.

6.1 Implications for community-centered data science

We turn our attention first to translating our findings for a data science audience. We will present three reflections on prior work, pointing to implications where Community Archetypes can help to better leverage insights from organizational psychology and more closely align data science with users' SOVC.

6.1.1 Rethinking "community size" as a success metric. HCI papers often use *community size* as a success metric (e.g., [33, 60, 111]) on the assumption that the larger a community has grown, the more successful it is. This metric *does* indicate how effectively an online group has recruited users, and the recruitment of a "critical mass" of users is a necessary pre-condition and known challenge for community survival [68]. However, our work suggests that community size is too blunt of an instrument for assessing SOVC. Rather, users often participate in small subreddits to meet their preferences and needs [54] and if subreddits grow too large, studies have noted user complaints that they may have been "*better when they were smaller.*" [71] As in Usenet's "Eternal September" [47], influxes of new users may seriously disrupt cherished community dynamics, unless there are sufficient sociotechnical affordances and policies to maintain established practices [62]. For example, r/wallstreetbets rapidly attracted several million new users after its meme-stocks

disrupted the market in 2021 [1]. By 2022, longtime members felt that *r/wallstreetbets* lost its distinctive character, since new users' discussions mostly revolved around old, tired topics [11]. On the other hand, Lin et al. showed that when subreddits were added to default subscription lists by Reddit admins (as was the case in *r/aww*), there were *momentary* fluctuations in community activity following newcomer influxes, but the subreddits more-or-less returned to pre-established patterns following the disruption [71]. Newer work similarly explores community resilience after a subreddit is featured on *r/popular* (Reddit's replacement for default subreddits), finding that community behaviors were differentially and strongly impacted in smaller rather than larger subreddits [26]. These examples show that community size does not predict SOVC, even though it can *impact* SOVC.

Our work suggests that alternative metrics may be better aligned with users' SOVC. As reported in Section 5.1, data scientists should consider community success metrics derived from behavioral traces such as: (1) the frequency and depth of conversations occurring among users; (2) the interactions between well-established community members and newer community members; or (3) the persistence of users' engagement over time. These strategies should be widely applicable across online communities beyond Reddit as forms of interactivity and membership that are antecedents to SOVC. This concept is supported by prior work demonstrating that greater thread depth on Reddit is associated with language markers indicating positive community outcomes like stability, cohesiveness, and sociability [75]. However, the Community Archetypes framework also suggests that more specialized forms of interactivity are expected within specific archetypes, and that other antecedents like membership boundaries, norm enforcement, and homogeneity may be quite distinct between archetypes. Our next reflection explores how to operationalize this specificity by re-purposing methods proposed by prior work.

6.1.2 Rethinking "SOVC antecedents" as universal. The Community Archetypes framework suggests that users form SOVC as a consequence of community behaviors that differ across archetypes. In other words, different archetypes have specialized antecedents, and this impacts how researchers should interpret different categories of behavior across different communities. For example, Bao et al. describe computational measures for a set of eight pro-social behaviors expected in healthy online communities: information sharing, gratitude, esteem enhancement, social support, social cohesion, fundraising and donating, mentoring, and the absence of anti-social behavior [12]. Although these all seem beneficial at face value, they are not necessarily required antecedents of SOVC across all archetypes. For instance, social support is fundamental to the Support archetype, however it may not be expected, needed, or even appropriate in others. If researchers assume that all healthy communities are marked by high levels of social support, they risk inaccurate conclusions that community spaces where social support is low or absent also lack a sense of virtual community. Therefore, for computational metrics related to community features such as pro-social behaviors [12] or types of relationships between users [32], researchers need to choose *which* measures align with users' normative experiences, needs, and values in that particular community and archetype. In *r/wallstreetbets*, for instance, the edgy style of social support could easily be mis-categorized or perceived as anti-social if the researcher was unfamiliar with the norms of the community. On the other hand, in *r/aww*, the purpose of the subreddit is to generate cute memes; social support is not the point. Therefore, low levels of social support would not necessarily indicate poor SOVC in this "ContentGen" Community Archetype.

6.1.3 Rethinking "community boundaries" across Reddit. As in [105], we acknowledge that a single subreddit is not the best unit of community. Rather, there exist mutualistic, overlapping groups of members across subreddits [112] with similar topics or genealogical relationships [111] and norms or rules [28, 40]. News events such as the #RedditBlackout in 2023 due to the introduction of new API

fees [77] and academic studies have also examined migratory events in which groups of users moved between different spaces on Reddit (or off of the platform entirely) due to moderation conflicts [27, 36, 80, 114]. For example, when Reddit banned the hate-based subreddits *r/fatpeoplehate* and *r/CoonTown*, many users left the platform, whereas others moved to new subreddits but mostly did not bring their hate speech with them [27]. These studies highlight how the same communities of users exist beyond the structures of individual subreddits and how these communities may or may not reconstitute themselves elsewhere after major events.

The Community Archetypes framework provides new strategies for delineating communities that may help to describe these phenomena and better match users' experiences of SOVC than individual subreddits. For example, community detection methods that rely upon network structure [35, 42, 82] may benefit from incorporating archetypal community features. One possibility is a “top-down” strategy in which researchers could assess and label the archetypal composition of each subreddit in their sample, next analyzing the interaction and relationships specific to each archetype separately. This strategy will allow researchers to look at similar behaviors across multiple different archetypes, but to interpret them more coherently within the context of each archetype. On the other hand, a “bottom-up” strategy could de-emphasize subreddits as units and instead look primarily to users' interaction patterns and styles to delineate Reddit communities. For example, researchers could define communities by setting participation thresholds or criteria guided by questions like: Which users of which subreddits are regularly interacting with each other (and/or longer term super users) in nested discussions, regardless of which subreddit the interaction takes place in? What types of *roles* [23] and *norms* [28] do these behaviors indicate? Within this community of users interacting across subreddits, what archetype(s) best explain their community's activities? Although this strategy may be computationally taxing, it should enable a more fine-grained analysis with better alignment to users' SOVC.

Finally, it is also important to highlight the temporal boundaries of communities [13]. Rather than looking at “all time” activities across a subreddit, researchers may better align their methods with users' SOVC if they choose time frames that align with members' actual engagements with the community [44]. Users often flow in and out of online spaces, taking their ideas and norms with them. This creates “cohorts”—*i.e.*, groups of people who entered the community near the same time, often as the result of an event that brought their attention to the topic or to the specific subreddit [13]. For example, in Affiliation subreddits (*e.g.*, with a sports team or television show) the community may be healthy and thriving, with a great deal of activity during events related to the entity (*e.g.*, live games or the release of new episodes), even if it is dormant during the “off season.” Alternatively, even if activity is quite consistent in a given subreddit, our participants also observed that content may start to feel repetitive when discussions return to the same topics over time. For example, in the Learning archetype, outsiders may have similar questions that regularly percolate to the surface, causing cyclical patterns in the content. Given these types of considerations, we suggest that analyzing cohorts of users who were active during the same period will be more aligned with their SOVC within a given archetypal community.

6.2 Applying the Community Archetypes framework in research

In cases when researchers hope to study, model, or support users' experiences of SOVC in online communities, the Community Archetypes framework can help researchers to structure and justify their selection of methods. This section synthesizes a workflow for doing so (Fig. 4).

6.2.1 Identify archetype(s). Researchers should first consider which archetype(s) best fit the purposes of their study. (Alternatively, if subreddit(s) are already identified, assessing which archetype(s) those subreddit(s) embody is necessary to analyze them effectively.) Data science models built

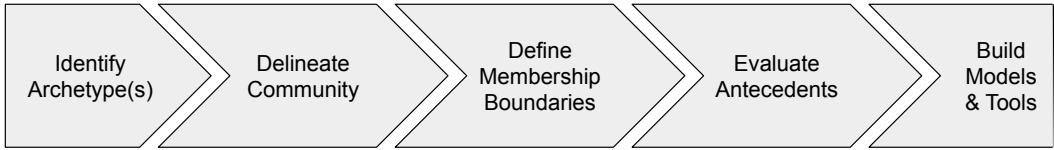


Fig. 4. Workflow for applying the Community Archetypes framework.

across many different subreddits without specification of archetype should only rely upon broadly applicable analytical strategies (as in Sec. 6.1.1 above), whereas more sophisticated techniques can be designed if researchers have conscientiously selected particular archetype(s). For methods with smaller sample sizes (e.g., ethnography, interviews, surveys, content analysis, or mixed methods), researchers can use the framing of archetypes, content patterns, and user roles to help structure and scope their protocols, questions posed to users, analytical codebooks, *etc.*, using the archetypal category as justification for asking about certain types of norms, behaviors, *etc.*, and not others.

6.2.2 Delineate community. Once archetype(s) have been chosen, researchers can delineate community boundaries at the intersection of those archetype(s), the topic of interest, and a sensible timeframe. This can be done through examination of groups of candidate subreddits *or* groups of interacting users across subreddits (sec. 6.1.3). Researchers can choose qualitative or quantitative techniques for analyzing content patterns, user roles, and rules/norms (*see* Table 4) depending on the size and scope of their study. However, in order to ensure that assumptions hold across their sample, they should select communities of similar archetypal composition.

Most researchers in our study opted to study the largest or most popular possible subreddit(s) on their topic of interest. However, large subreddits have distinct moderation and governance challenges relative to smaller subreddits. The question of what size subreddit is “most representative” or “more of a community” is difficult given that long-tailed distributions have no central tendency: are many small subreddits more representative of a user’s typical experience or a few large subreddits? The largest subreddit may not always be the best choice, particularly when it is important that a community exists or a specific kind of data is needed. For example, one researcher described how many fandoms have multiple Affiliate subreddits, the largest of which is often for more general information, casual fans, and people just discovering the fandom (*i.e.*, future affiliates), whereas smaller ones are dedicated to more specific sub-topics for much more highly invested current affiliates or have more stringent moderation rules. More community-like behaviors may exist in these smaller, more specific subreddits than in the larger “on-boarding” subreddit. Therefore looking at smaller, more specialized subreddits or groups of users interacting across the whole family tree of fandom Affiliate subreddits may be valuable tactics.

6.2.3 Define membership boundaries. The next step is to identify how users should be classified as members. Our framework suggests a tiered membership classification scheme because a binary view of membership may be overly simplistic and provides an incomplete interpretation of actual community dynamics. An ideal way to determine membership in qualitative and mixed methods is to first observe what user roles exist, and then ask users directly about how they view their own membership and role in the community. In quantitative and data science-based methods, rather than a simple activity threshold (e.g., “members” are defined by making X number of posts), researchers should identify markers of expected user roles within the specified archetype(s). Prior work on identifying social roles in online communities [23, 57, 79, 120, 123] and techniques that combine behavioral and content features [72] will be especially useful in this regard.

6.2.4 Evaluate remaining antecedents. The first three steps set the stage for more effectively assessing the remaining antecedents (interactivity, homogeneity, and norm enforcement) with respect to the appropriate archetype.

Interactivity. Different *forms* of interactivity are more essential in different archetypes (sec. 6.1.2). Prior work provides examples of ways to assess some antecedents (*e.g.*, [12]). New methods can be tailored to distinctive archetypal content patterns:

- **Q&A:** Researchers can assess whether almost all posts are formatted as questions, whether the majority are receiving answers, and how in-depth and varied the answers are.
- **Support:** Support exchange is fundamental and SOVC is likely when users regularly switch between support-seeking and support-providing roles.
- **Affiliation:** Frequent references to entity-specific terms, people, and topics encourage SOVC; the community should become more active during important entity-affiliated events.
- **Learning:** SOVC is likely when in-depth, respectful discussions feature many contrasting perspectives between insiders and outsiders.
- **ContentGen:** Posts and in-jokes should adhere to expected content formats and not be repetitive re-posts. Posts that refer to and build on prior content should indicate SOVC.

Homogeneity. Participants suggested that demographic homogeneity (race, sex, age, *etc.*) is less salient than other forms of homogeneity, such as shared interests, experiences, and goals. Even these latter forms may be differentially important across archetypes. In qualitative studies, researchers may want to focus on shared interests for Q&A and ContentGen communities, shared experiences for Support and Affiliation communities, and shared goals in Learning communities. In quantitative studies, linguistic homogeneity is important to community cohesion and should be useful across all archetypes.

Norm Enforcement. There are *macro* norms (universal to much of Reddit), *meso* norms across particular groups of subreddits, and highly specialized *micro* norms within individual subreddits [28]. Community Archetypes offers a new lens for interpreting meso norms: we expect that subreddits of similar archetypal composition are likely to be marked by similar meso norms, and that users' SOVC is related to how these norms are vertically and/or horizontally enforced. Future work should explore this question in more depth, particularly by using the taxonomy of Reddit rules by Fiesler *et al.* (as in our cursory rules analysis included in the supplemental materials) [40]. We expect differences across archetypes in terms of: what types of rules/norms exist; how strictly users *want* norms to be enforced; and how strictly moderators enforce norms in actual practice. For instance, users of Support subreddits especially emphasized how their SOVC is contingent upon stricter enforcement and mods' commitment to keeping the space safe; these subs may need more rules overall and higher enforcement both vertically and horizontally to protect users' SOVC. Q&A and ContentGen subreddits may need stronger adherence to prescriptive content rules, whereas Learning and Affiliation subreddits may need better mechanisms to govern behaviors like harassment, hate speech, and influence operations.

6.2.5 Build models and community support tools. Community Archetypes can sensitize our research methods, data models and community support tools to meaningfully nourish the human need for sense of community in our modern world. For example, users told us that strategies like increasing the activity of moderators, keeping subreddits small but fostering more activity and content, and structuring regular activities both on Reddit (*e.g.*, routine threads for catching up with other users) and off Reddit (*e.g.*, on Discord) would contribute to improving SOVC. They also highlighted opportunities for better governance bots [105] for patrolling hate speech and troll activities, specific tasks in that subreddit's interest area, answering repetitive posts on common questions, or responding to posts that never received any comments.

Tailoring governance support to Community Archetypes. Similarly to how rules and norms vary by archetype, our work suggests that there is no “one-size-fits-all” set of moderation strategies, policies, or tools that are likely to work equally as well for all communities. Styles and types of moderation labor should differ in reliable ways across different archetypes; thus, a useful future contribution could be baseline tooling tailored appropriately to each archetype. Rather than one u/AutoModerator to rule them all, subreddit mods could be provided with options configured and tuned to more sophisticated patterns of activity in the subreddit—not only for enforcing rules restrictively, but also for spurring positive antecedent behaviors, interactions, and communication that lead to SOVC in that archetype. For example, imagine an u/AutoSupportMod for Social Support subs, an u/AutoMemeMod for ContentGen, etc.

6.3 Community Archetypes beyond Reddit

6.3.1 Archetypes across platforms. Reddit is now a prominent platform studied by academics, however that could easily change given the ever-shifting landscape of social media. The archetypes uncovered in our study relate to basic human needs for information, entertainment, socialization, and support, and that they likely feel familiar to readers. Entire platforms with more constrained affordances and contexts of use than Reddit exist specifically for Q&A (e.g., StackOverflow, Quora), Social Support (e.g., CaringBridge, PatientsLikeMe), Content Generation (e.g., Imgur, MemeGenerator), etc. At the same time, platforms like Facebook, Instagram, Twitter, or analogous Fediverse options can support similar Community Archetypes depending on how users and/or moderators appropriate them. Reddit communities likely learned a thing or two from their predecessors, just as future platforms will iterate on today’s models. We expect this Community Archetypes framing to remain fundamentally relevant to research because the underlying community needs will remain relevant to users. Yet future work will need to account for new sociotechnical circumstances, such as issues raised during the #RedditBlackout in June 2023.

Although Reddit’s API was free and open for many years, new API rules have restricted access to data, impacting researchers, developers, and users. Moderators of thousands of popular subreddits made their subs private (and therefore inaccessible to millions of members) or participated in other forms of protest such as bizarre or humorous content restrictions [77]. Although Reddit granted mods a limited tier of free API usage for running moderation bots [51], it also threatened to remove mods who refused to re-open their subs [85]. Issues like this may cause users to abandon the platform, migrate to other subreddits, or alter their community participation behaviors. Amidst the instability of shifting platform incentives and migrations, Community Archetypes offers a promising way to conceptualize, study, and support real human communities—regardless of the platform they currently inhabit.

6.3.2 Future work. Given the rapid rate at which algorithms, affordances, behaviors, and governance of tech platforms continue to evolve, the Community Archetypes framework is certainly not complete. We did not specifically set out to find these archetypes, thus different archetypes could exist or emerge on other current or future social media platforms. Future work may identify additional archetypes missed by this study, as well as methodological considerations that could differ accordingly. Additionally, there is a need for case studies in which we can empirically use and validate the Community Archetypes framework not only on the level of individual “digital containers” of communities (e.g., subreddits) but even more especially across *groups* of such containers that provide users with SOVC. Moreover, trace data can easily be misinterpreted and can never tell the whole story of users’ internal states or feelings. In order to truly understand how users are feeling, it will always be necessary to ask them directly, thus future work should continue

engaging directly with community members to understand their experiences and perceptions as platforms change.

6.4 Conclusion

In this study, we interviewed 21 Reddit researchers and surveyed users of 12 subreddits to explore SOVC. We identified five Community Archetypes with distinctive forms of activity, user roles, and content patterns. Accordingly, we contribute methodological recommendations for identifying and studying each of these archetypes. This work will support researchers to understand, select, and justify methods that best suit their projects, as well as to design community support tools that can promote healthy community formation and provide members with nourishing SOVC.

ACKNOWLEDGMENTS

We thank our participants for sharing their insights and experiences. We are also appreciative for the supportive feedback from Stevie Chancellor, Loren Terveen, Charles Chuankai Zhang, Sanjay Kairam, and the dedicated anonymous reviewers who supported us to improve this manuscript. Undergraduate assistants Tamer Shahwan, Sahand Setareh, and Vikki Wong helped us to collect our survey data and were also instrumental to the success of this project. This work was supported by NSF Award #1910225.

REFERENCES

- [1] Chris Alcantara. 2021. Reddit's /r/wallstreetbets astronomical rise. *Washington Post* (Jan. 2021). <https://www.washingtonpost.com/technology/2021/01/29/wallstreetbets-reddit-gamestop/>
- [2] Jisun An, Haewoon Kwak, Oliver Posegga, and Andreas Jungheer. 2019. Political Discussions in Homogeneous and Cross-Cutting Communication Spaces. *Proceedings of the International AAAI Conference on Web and Social Media* 13 (July 2019), 68–79. <https://doi.org/10.1609/icwsm.v13i01.3210>
- [3] Nazanin Andalibi, Oliver L. Haimson, Munmun De Choudhury, and Andrea Forte. 2018. Social Support, Reciprocity, and Anonymity in Responses to Sexual Abuse Disclosures on Social Media. *ACM Transactions on Computer-Human Interaction* 25, 5 (Oct. 2018), 28:1–28:35. <https://doi.org/10.1145/3234942>
- [4] Nazanin Andalibi, Oliver L. Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). Association for Computing Machinery, New York, NY, USA, 3906–3918. <https://doi.org/10.1145/2858036.2858096>
- [5] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2017. Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of #Depression. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1485–1500. <https://doi.org/10.1145/2998181.2998243>
- [6] Benedict R. O'G Anderson. 2006. *Imagined communities: reflections on the origin and spread of nationalism* (rev. ed ed.). Verso, London ; New York.
- [7] Anderson, Benedict. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*.
- [8] Blake E. Ashforth and Fred Mael. 1989. Social Identity Theory and the Organization. *Academy of Management Review* 14, 1 (Jan. 1989), 20–39. <https://doi.org/10.5465/amr.1989.4278999>
- [9] Kenneth M. Bachrach and Alex J. Zautra. 1985. Coping with a Community Stressor: The Threat of a Hazardous Waste Facility. *Journal of Health and Social Behavior* 26, 2 (1985), 127–141. <https://doi.org/10.2307/2136602>
- [10] Tanvi Bajpai, Drshika Asher, Anwesa Goswami, and Eshwar Chandrasekharan. 2022. Harmonizing the Cacophony with MIC: An Affordance-aware Framework for Platform Moderation. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 385:1–385:22. <https://doi.org/10.1145/3555110>
- [11] Gunjan Banerji and Caitlin McCabe. 2022. Reddit's WallStreetBets Was the GameStop Kingmaker, but Longtime Users Say the Thrill Is Gone. *The Wall Street Journal* (Jan. 2022). <https://www.wsj.com/articles/reddits-wallstreetbets-was-the-gamestop-kingmaker-but-longtime-users-say-the-thrill-is-gone-11643025602>
- [12] Jiajun Bao, Junjie Wu, Yiming Zhang, Eshwar Chandrasekharan, and David Jurgens. 2021. Conversations Gone Alright: Quantifying and Predicting Prosocial Outcomes in Online Conversations. In *Proceedings of the Web Conference 2021*. Association for Computing Machinery, New York, NY, USA, 1134–1145. <https://doi.org/10.1145/3442381.3450122>

- [13] Samuel Barbosa, Dan Cosley, Amit Sharma, and Roberto M. Cesar. 2016. Averaging Gone Wrong: Using Time-Aware Analyses to Better Understand Behavior. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 829–841. <https://doi.org/10.1145/2872427.2883083>
- [14] Hoda Baytiyeh and Jay Pfaffman. 2010. Volunteers in Wikipedia: Why the Community Matters. *Journal of Educational Technology & Society* 13, 2 (2010), 128–140. <https://www.jstor.org/stable/jeductechsoci.13.2.128> Publisher: International Forum of Educational Technology & Society.
- [15] Michael Bieber, Douglas Engelbart, Richard Furuta, Starr Roxanne Hiltz, John Noll, Jennifer Preece, Edward A. Stohr, Murray Turoff, and Bartel Van De Walle. 2002. Toward Virtual Community Knowledge Evolution. *Journal of Management Information Systems* 18, 4 (March 2002), 11–35. <https://doi.org/10.1080/07421222.2002.11045707>
- [16] Anita L. Blanchard, Leann E. Caudill, and Lisa Slattery Walker. 2020. Developing an entitativity measure and distinguishing it from antecedents and outcomes within online and face-to-face groups. *Group Processes & Intergroup Relations* 23, 1 (Jan. 2020), 91–108. <https://doi.org/10.1177/1368430217743577>
- [17] Anita L. Blanchard and M. Lynne Markus. 2004. The experienced "sense" of a virtual community: characteristics and processes. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* 35, 1 (Feb. 2004), 64–79. <https://doi.org/10.1145/968464.968470>
- [18] Michael G. Blight, Erin K. Ruppel, and Kelsea V. Schoenbauer. 2017. Sense of Community on Twitter and Instagram: Exploring the Roles of Motives and Parasocial Relationships. *Cyberpsychology, Behavior, and Social Networking* 20, 5 (May 2017), 314–319. <https://doi.org/10.1089/cyber.2016.0505> Publisher: Mary Ann Liebert, Inc., publishers.
- [19] Gianluca Bonifazi, Silvia Cecchini, Enrico Corradini, Lorenzo Giuliani, Domenico Ursino, and Luca Virgili. 2022. Investigating community evolutions in TikTok dangerous and non-dangerous challenges. *Journal of Information Science* (Aug. 2022). <https://doi.org/10.1177/01655515221116519>
- [20] Amy Bruckman. 2006. A new perspective on "community" and its implications for computer-mediated communication systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. Association for Computing Machinery, New York, NY, USA, 616–621. <https://doi.org/10.1145/1125451.1125579>
- [21] Amy S. Bruckman. 2022. *Should You Believe Wikipedia?: Online Communities and the Construction of Knowledge*. Cambridge University Press, New York, NY. <https://doi.org/10.1017/9781108780704>
- [22] Susan L. Bryant, Andrea Forte, and Amy Bruckman. 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work - GROUP '05*. ACM Press, Sanibel Island, Florida, USA, 1. <https://doi.org/10.1145/1099203.1099205>
- [23] Cody Buntain and Jennifer Golbeck. 2014. Identifying social roles in reddit using network structure. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, Seoul Korea, 615–620. <https://doi.org/10.1145/2567948.2579231>
- [24] Lauren E. Cagle and Carl Herndl. 2019. Shades of denialism: discovering possibilities for a more nuanced deliberation about climate change in online discussion forums. *Communication Design Quarterly* 7, 1 (May 2019), 22–39. <https://doi.org/10.1145/3331558.3331561>
- [25] Donald T. Campbell. 1958. Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science* 3, 1 (1958), 14–25. <https://doi.org/10.1002/bs.3830030103>
- [26] Jackie Chan, Aditi Atreyasa, Stevie Chancellor, and Eshwar Chandrasekharan. 2022. Community Resilience: Quantifying the Disruptive Effects of Sudden Spikes in Activity within Online Communities. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–8. <https://doi.org/10.1145/3491101.3519813>
- [27] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (Dec. 2017), 31:1–31:22. <https://doi.org/10.1145/3134666>
- [28] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. The Internet's Hidden Rules: An Empirical Study of Reddit Norm Violations at Micro, Meso, and Macro Scales. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 32:1–32:25. <https://doi.org/10.1145/3274301>
- [29] Kathy Charmaz. 2014. *Constructing Grounded Theory* (2nd ed.). SAGE.
- [30] David M. Chavis and J. R. Newbrough. 1986. The meaning of "community" in community psychology. *Journal of Community Psychology* 14, 4 (1986), 335–340. [https://doi.org/10.1002/1520-6629\(198610\)14:4<335::AID-JCOP2290140402>3.0.CO;2-T](https://doi.org/10.1002/1520-6629(198610)14:4<335::AID-JCOP2290140402>3.0.CO;2-T)
- [31] Boreum Choi, Kira Alexander, Robert E. Kraut, and John M. Levine. 2010. Socialization tactics in wikipedia and their effects. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 107–116. <https://doi.org/10.1145/1718918.1718940>
- [32] Minje Choi, Luca Maria Aiello, Krisztián Zsolt Varga, and Daniele Quercia. 2020. Ten Social Dimensions of Conversations and Relationships. In *Proceedings of The Web Conference 2020*. ACM, Taipei Taiwan, 1514–1525.

<https://doi.org/10.1145/3366423.3380224>

- [33] Tiago Cunha, David Jurgens, Chenhao Tan, and Daniel Romero. 2019. Are All Successful Communities Alike? Characterizing and Predicting the Success of Online Communities. In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 318–328. <https://doi.org/10.1145/3308558.3313689>
- [34] Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No country for old members: user lifecycle and linguistic change in online communities. In *Proceedings of the 22nd international conference on World Wide Web - WWW '13*. ACM Press, Rio de Janeiro, Brazil, 307–318. <https://doi.org/10.1145/2488388.2488416>
- [35] Srayan Datta and Eytan Adar. 2019. Extracting Inter-Community Conflicts in Reddit. *Proceedings of the International AAAI Conference on Web and Social Media* 13 (July 2019), 146–157. <https://ojs.aaai.org/index.php/ICWSM/article/view/3217>
- [36] Cai Davies, James Ashford, Luis Espinosa-Anke, Alun Preece, Liam Turner, Roger Whitaker, Mudhakar Srivatsa, and Diane Felmlee. 2021. Multi-scale user migration on Reddit. AAAI, Virtual. <https://doi.org/10.36190/2021.13>
- [37] Robert J. Doolittle and Donald Macdonald. 1978. Communication and a sense of community in a metropolitan neighborhood: A factor analytic examination. *Communication Quarterly* 26, 3 (June 1978), 2–7. <https://doi.org/10.1080/01463377809369297>
- [38] Brianna Dym and Casey Fiesler. 2020. Social Norm Vulnerability and its Consequences for Privacy and Safety in an Online Community. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 155:1–155:24. <https://doi.org/10.1145/3415226>
- [39] Casey Fiesler and Brianna Dym. 2020. Moving Across Lands: Online Platform Migration in Fandom Communities. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (May 2020), 042:1–042:25. <https://doi.org/10.1145/3392847>
- [40] Casey Fiesler, Jialun Jiang, Joshua McCann, Kyle Frye, and Jed Brubaker. 2018. Reddit Rules! Characterizing an Ecosystem of Governance. *Proceedings of the International AAAI Conference on Web and Social Media* 12, 1 (June 2018). <https://ojs.aaai.org/index.php/ICWSM/article/view/15033> Number: 1.
- [41] Kathi Fisler and Francisco Enrique Vicente Castro. 2017. Sometimes, Rainfall Accumulates: Talk-Alouds with Novice Functional Programmers. In *Proceedings of the 2017 ACM Conference on International Computing Education Research*. ACM, Tacoma Washington USA, 12–20. <https://doi.org/10.1145/3105726.3106183>
- [42] Santo Fortunato and Darko Hric. 2016. Community detection in networks: A user guide. *Physics Reports* 659 (Nov. 2016), 1–44. <https://doi.org/10.1016/j.physrep.2016.09.002>
- [43] Seth Frey, Qiankun Zhong, Beril Bulat, William D. Weisman, Caitlyn Liu, Stephen Fujimoto, Hannah Wang, and Charles M. Schweik. 2022. Governing Online Goods: Maturity and Formalization in Minecraft, Reddit, and World of Warcraft Communities. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 300:1–300:23. <https://doi.org/10.1145/3555191>
- [44] R. Stuart Geiger and Aaron Halfaker. 2013. Using edit sessions to measure participation in wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 861–870. <https://doi.org/10.1145/2441776.2441873>
- [45] Thomas J. Glynn. 1981. Psychological Sense of Community: Measurement and Application. *Human Relations* 34, 9 (Sept. 1981), 789–818. <https://doi.org/10.1177/001872678103400904>
- [46] Frances J. Griffith and Catherine H. Stein. 2021. Behind the Hashtag: Online Disclosure of Mental Illness and Community Response on Tumblr. *American Journal of Community Psychology* 67, 3-4 (2021), 419–432. <https://doi.org/10.1002/ajcp.12483>
- [47] Wendy Grossman. 1997. *Net.wars*. NYU Press.
- [48] Joseph R. Gusfield. 1975. *Community: A critical response*. Harper & Row New York.
- [49] Aaron Halfaker and Dario Taraborelli. 2013. Making Peripheral Participation Legitimate: Reader Engagement Experiments in Wikipedia. (Feb. 2013), 849–860. <https://doi.org/10.1145/2441776.2441872>
- [50] William A. Hamilton, Oliver Garretson, and Android Kerne. 2014. Streaming on twitch: fostering participatory communities of play within live mixed media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1315–1324. <https://doi.org/10.1145/2556288.2557048>
- [51] Reddit Help. 2023. Moderation Bots & Tooling. <https://mods.reddithelp.com/hc/en-us/articles/16693988535309-Moderation-Bots-Tooling>
- [52] Joke Hermes and Leonie Stoete. 2019. Hating Skyler White: audience engagement, gender politics and celebrity culture. *Celebrity Studies* 10, 3 (July 2019), 411–426. <https://doi.org/10.1080/19392397.2019.1630155>
- [53] Rick H. Hoyle (Ed.). 2013. *Handbook of Structural Equation Modeling* (1 ed.). Guilford Press. <https://www.routledge.com/Handbook-of-Structural-Equation-Modeling/Hoyle/p/book/9781462516797>

- [54] Sohyeon Hwang and Jeremy D. Foote. 2021. Why do People Participate in Small Online Communities? *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–25. <https://doi.org/10.1145/3479606>
- [55] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (WebKDD/SNA-KDD '07)*. Association for Computing Machinery, New York, NY, USA, 56–65. <https://doi.org/10.1145/1348549.1348556>
- [56] Jialun Aaron Jiang, Charles Kiene, Skyler Middler, Jed R. Brubaker, and Casey Fiesler. 2019. Moderation Challenges in Voice-based Online Communities on Discord. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 55:1–55:23. <https://doi.org/10.1145/3359157>
- [57] Leigh Jin, Daniel Robey, and Marie-Claude Boudreau. 2007. Beyond Development: A Research Agenda for Investigating Open Source Software User Communities. *Information Resources Management Journal* 20, 1 (Jan. 2007), 68–80. <https://doi.org/10.4018/irmj.2007010105>
- [58] Barbara Johnstone. 2008. *Discourse Analysis*. Blackwell Publishing, Malden, MA.
- [59] Sanjay Ram Kairam, Steven Sumner, and Melissa C. Mercado-Crespo. 2022. A Social-Ecological Approach to Modeling Sense of Virtual Community (SOVC) in Livestreaming Communities. *CSCW '22: ACM Conference On Computer-Supported Cooperative Work And Social Computing* 6, CSCW2 (2022), 1–35. <https://doi.org/10.1145/3555081>
- [60] Sanjay Ram Kairam, Dan J. Wang, and Jure Leskovec. 2012. The life and death of online groups: predicting group growth and longevity. In *Proceedings of the fifth ACM international conference on Web search and data mining (WSDM '12)*. Association for Computing Machinery, New York, NY, USA, 673–682. <https://doi.org/10.1145/2124295.2124374>
- [61] Brian C. Keegan. 2022. What is Mastodon? A social media expert explains how the 'federated' network works and why it won't be a new Twitter. <http://theconversation.com/what-is-mastodon-a-social-media-expert-explains-how-the-federated-network-works-and-why-it-wont-be-a-new-twitter-194329>
- [62] Charles Kiene, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. Surviving an "Eternal September": How an Online Community Managed a Surge of Newcomers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1152–1156. <https://doi.org/10.1145/2858036.2858356>
- [63] Jennifer G. Kim, Robert E. Kraut, and Karrie Karahalios. 2022. Designing a Medical Crowdfunding Website from Sense of Community Theory. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 290:1–290:25. <https://doi.org/10.1145/3555181>
- [64] Shamika Klassen, Sara Kingsley, Kalyn McCall, Joy Weinberg, and Casey Fiesler. 2021. More than a Modern Day Green Book: Exploring the Online Community of Black Twitter. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 458:1–458:29. <https://doi.org/10.1145/3479602>
- [65] Joon Koh, Young-Gul Kim, and Young-Gul Kim. 2003. Sense of Virtual Community: A Conceptual Framework and Empirical Validation. *International Journal of Electronic Commerce* 8, 2 (Dec. 2003), 75–94. <https://doi.org/10.1080/10864415.2003.11044295>
- [66] Yubo Kou and Colin M. Gray. 2017. Supporting Distributed Critique through Interpretation and Sense-Making in an Online Creative Community. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (Dec. 2017), 60:1–60:18. <https://doi.org/10.1145/3134695>
- [67] Yubo Kou, Xinning Gui, Yunan Chen, and Kathleen Pine. 2017. Conspiracy Talk on Social Media: Collective Sense-making during a Public Health Crisis. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (Dec. 2017), 61:1–61:21. <https://doi.org/10.1145/3134696>
- [68] Robert E. Kraut and Paul Resnick. 2012. *Building Successful Online Communities: Evidence-Based Social Design*. <https://doi.org/10.7551/mitpress/8472.001.0001>
- [69] Zachary Levonian, Marco Dow, Drew Erikson, Sourojit Ghosh, Hannah Miller Hillberg, Saumik Narayanan, Loren Terveen, and Svetlana Yarosh. 2021. Patterns of Patient and Caregiver Mutual Support Connections in an Online Health Community. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 275:1–275:46. <https://doi.org/10.1145/3434184>
- [70] Brian Lickel, David L. Hamilton, and Steven J. Sherman. 2001. Elements of a Lay Theory of Groups: Types of Groups, Relational Styles, and the Perception of Group Entitativity. *Personality and Social Psychology Review* 5, 2 (May 2001), 129–140. https://doi.org/10.1207/S15327957PSPR0502_4
- [71] Zhiyuan Lin, Niloufar Salehi, Bowen Yao, Yiqi Chen, and Michael Bernstein. 2017. Better When It Was Smaller? Community Content and Behavior After Massive Growth. *Proceedings of the International AAAI Conference on Web and Social Media* 11, 1 (May 2017), 132–141. <https://doi.org/10.1609/icwsm.v11i1.14884> Number: 1.
- [72] Yezheng Liu, Fei Du, Jianshan Sun, Thushari Silva, Yuanchun Jiang, and Tingting Zhu. 2019. Identifying social roles using heterogeneous features in online social networks. *Journal of the Association for Information Science and Technology* 70, 7 (2019), 660–674. <https://doi.org/10.1002/asi.24160>

- [73] Elizabeth Lopatto. 2021. How r/WallStreetBets gamed the stock of GameStop. *The Verge* (Jan. 2021). <https://www.theverge.com/22251427/reddit-gamestop-stock-short-wallstreetbets-robinhood-wall-street>
- [74] Terri Mannarini and Angela Fedi. 2009. Multiple senses of community: the experience and meaning of community. *Journal of Community Psychology* 37, 2 (March 2009), 211–227. <https://doi.org/10.1002/jcop.20289>
- [75] Bree McEwan. 2016. Communication of communities: linguistic signals of online groups. *Information, Communication & Society* 19, 9 (Sept. 2016), 1233–1249. <https://doi.org/10.1080/1369118X.2016.1186717> Publisher: Routledge _eprint: <https://doi.org/10.1080/1369118X.2016.1186717>.
- [76] David W. McMillan and David M. Chavis. 1986. Sense of community: A definition and theory. *Journal of Community Psychology* 14, 1 (Jan. 1986), 6–23. <https://doi.org/10.1002/1520-6629>
- [77] Ivan Mehta. 2023. Reddit communities adopt alternative forms of protest, as the company threatens action on moderators. <https://techcrunch.com/2023/06/19/reddit-communities-adopt-alternative-forms-of-protest-as-the-company-threats-action-on-moderators/>
- [78] Michael Muller. 2014. Curiosity, Creativity, and Surprise as Analytic Tools: Grounded Theory Method. In *Ways of Knowing in HCI*, Judith S. Olson and Wendy A. Kellogg (Eds.). Springer New York, New York, NY, 25–48. https://doi.org/10.1007/978-1-4939-0378-8_2
- [79] Claudia Müller-Birn, Benjamin Karran, Janette Lehmann, and Markus Luczak-Rösch. 2015. Peer-production system or collaborative ontology engineering effort: what is Wikidata?. In *Proceedings of the 11th International Symposium on Open Collaboration (OpenSym '15)*. Association for Computing Machinery, New York, NY, USA, 1–10. <https://doi.org/10.1145/2788993.2789836>
- [80] Edward Newell, David Jurgens, Haji Mohammad Saleem, Hardik Vala, Jad Sassine, Caitrin Armstrong, and Derek Ruths. 2016. User Migration in Online Social Networks: A Case Study on Reddit During a Period of Community Unrest. In *Tenth International AAAI Conference on Web and Social Media*. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13137>
- [81] Patricia Obst, Sandy G. Smith, and Lucy Zinkiewicz. 2002. An exploration of sense of community, Part 3: Dimensions and predictors of psychological sense of community in geographical communities. *Journal of Community Psychology* 30, 1 (2002), 119–133. <https://doi.org/10.1002/jcop.1054>
- [82] Randal S. Olson and Zachary P. Neal. 2015. Navigating the massive world of reddit: using backbone networks to map user interests in social media. *PeerJ Computer Science* 1 (May 2015), e4. <https://doi.org/10.7717/peerj-cs.4> Publisher: PeerJ Inc..
- [83] Christian Pentzold. 2011. Imagining the Wikipedia community: What do Wikipedia authors mean when they write about their ‘community’? *New Media & Society* 13, 5 (Aug. 2011), 704–721. <https://doi.org/10.1177/1461444810378364> Publisher: SAGE Publications.
- [84] D. D. Perkins and R. B. Taylor. 1996. Ecological assessments of community disorder: their relationship to fear of crime and theoretical implications. *American Journal of Community Psychology* 24, 1 (Feb. 1996), 63–107. <https://doi.org/10.1007/BF02511883>
- [85] Jay Peters. 2023. Reddit will remove mods of private communities unless they reopen. *The Verge* (June 2023). <https://www.theverge.com/2023/6/29/23778997/reddit-remove-mods-private-communities-unless-reopen>
- [86] Liza Potts, Rebekah Small, and Michael Trice. 2019. Boycotting the Knowledge Makers: How Reddit Demonstrates the Rise of Media Blacklists and Source Rejection in Online Communities. *IEEE Transactions on Professional Communication* 62, 4 (Dec. 2019), 351–363. <https://doi.org/10.1109/TPC.2019.2946942> Conference Name: IEEE Transactions on Professional Communication.
- [87] Naveena Prakasam and Louisa Huxtable-Thomas. 2021. Reddit: Affordances as an Enabler for Shifting Loyalties. *Information Systems Frontiers* 23, 3 (June 2021), 723–751. <https://doi.org/10.1007/s10796-020-10002-x>
- [88] Jenny Preece. 2000. *Online communities: Designing usability, supporting sociability*. John Wiley New York. <https://www.cs.umd.edu/hcil/pubs/books/online-communities.shtml>
- [89] Grace M. H. Pretty, Colleen Conroy, Jason Dugay, Karen Fowler, and Diane Williams. 1996. Sense of community and its relevance to adolescents of all ages. *Journal of Community Psychology* 24, 4 (1996), 365–379. [https://doi.org/10.1002/\(SICI\)1520-6629\(199610\)24:4<365::AID-JCOP6>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1520-6629(199610)24:4<365::AID-JCOP6>3.0.CO;2-T)
- [90] Miretta Prezza and Stefano Costantini. 1998. Sense of community and life satisfaction: investigation in three different territorial contexts. *Journal of Community & Applied Social Psychology* 8, 3 (1998), 181–194. [https://doi.org/10.1002/\(SICI\)1099-1298\(199805/06\)8:3<181::AID-CASP436>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-1298(199805/06)8:3<181::AID-CASP436>3.0.CO;2-4)
- [91] Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. 2021. Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media + Society* 7, 2 (April 2021), 20563051211019004. <https://doi.org/10.1177/20563051211019004>
- [92] John E. Puddifoot. 1995. Dimensions of community identity. *Journal of Community & Applied Social Psychology* 5, 5 (Dec. 1995), 357–370. <https://doi.org/10.1002/casp.2450050507>

- [93] Mark Rapley and Grace M. H. Pretty. 1999. Playing procrustes: The interactional production of a “psychological sense of community”. *Journal of Community Psychology* 27, 6 (1999), 695–713. [https://doi.org/10.1002/\(SICI\)1520-6629\(199911\)27:6<695::AID-JCOP5>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1520-6629(199911)27:6<695::AID-JCOP5>3.0.CO;2-M)
- [94] Reddit. 2017. Default subreddit. https://reddit.fandom.com/wiki/Default_subreddit
- [95] María-Carmen Ricoy and Tiberio Feliz. 2016. Twitter as a Learning Community in Higher Education. *Journal of Educational Technology & Society* 19, 1 (2016), 237–248. <https://www.jstor.org/stable/jeductechsoci.19.1.237> Publisher: International Forum of Educational Technology & Society.
- [96] Eric W. Rothenbuhler, Lawrence J. Mullen, Richard DeLaurell, and Choon Ryul Ryu. 1996. Communication, Community Attachment, and Involvement. *Journalism & Mass Communication Quarterly* 73, 2 (June 1996), 445–466. <https://doi.org/10.1177/107769909607300214>
- [97] Dana Rotman, Jennifer Golbeck, and Jennifer Preece. 2009. The community is where the rapport is – on sense and structure in the youtube community. In *Proceedings of the fourth international conference on Communities and technologies (C&T '09)*. Association for Computing Machinery, New York, NY, USA, 41–50. <https://doi.org/10.1145/1556460.1556467>
- [98] Seymour B. Sarason. 1974. *The psychological sense of community: Prospects for a community psychology*. Jossey-Bass, Oxford, England. Pages: xii, 290.
- [99] Nathan Schneider, Primavera De Filippi, Seth Frey, Joshua Z. Tan, and Amy X. Zhang. 2021. Modular Politics: Toward a Governance Layer for Online Communities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 16:1–16:26. <https://doi.org/10.1145/3449090>
- [100] Joseph Seering, Juan Pablo Flores, Saiph Savage, and Jessica Hammer. 2018. The Social Roles of Bots: Evaluating Impact of Bots on Discussions in Online Communities. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 157:1–157:29. <https://doi.org/10.1145/3274426>
- [101] Joseph Seering, Geoff Kaufman, and Stevie Chancellor. 2022. Metaphors in moderation. *New Media & Society* 24, 3 (March 2022), 621–640. <https://doi.org/10.1177/1461444820964968>
- [102] Joseph Seering, Michal Luria, Connie Ye, Geoff Kaufman, and Jessica Hammer. 2020. It Takes a Village: Integrating an Adaptive Chatbot into an Online Gaming Community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376708>
- [103] Dhavan V. Shah, Jack M. McLeod, and So-Hyang Yoon. 2001. Communication, Context, and Community: An Exploration of Print, Broadcast, and Internet Influences. *Communication Research* 28, 4 (Aug. 2001), 464–506. <https://doi.org/10.1177/009365001028004005>
- [104] Sanjay Sharma. 2013. Black Twitter? Racial Hashtags, Networks and Contagion. *New Formations* 78, 78 (July 2013), 46–64. <https://doi.org/10.3898/NewF.78.02.2013>
- [105] C. Estelle Smith, Irfanul Alam, Chenhao Tan, Brian Keegan, and Anita Blanchard. 2022. The Impact of Governance Bots on Sense of Virtual Community: Development and Validation of the GOV-BOTs Scale. *CSCW '22: ACM Conference On Computer-Supported Cooperative Work And Social Computing* 6, 2 (2022).
- [106] C. Estelle Smith, Zachary Levonian, Haiwei Ma, Robert Giaquinto, Gemma Lein-McDonough, Zixuan Li, Susan O'conner-Von, and Svetlana Yarosh. 2020. "I Cannot Do All of This Alone": Exploring Instrumental and Prayer Support in Online Health Communities. *ACM Transactions on Computer-Human Interaction* 27, 5 (Aug. 2020), 38:1–38:41. <https://doi.org/10.1145/3402855>
- [107] C. Estelle Smith, Eduardo Nevarez, and Haiyi Zhu. 2020. Disseminating Research News in HCI: Perceived Hazards, How-To's, and Opportunities for Innovation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376744>
- [108] C. Estelle Smith, Bowen Yu, Anjali Srivastava, Aaron Halfaker, Loren Terveen, and Haiyi Zhu. 2020. Keeping Community in the Loop: Understanding Wikipedia Stakeholder Values for Machine Learning-Based Systems. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376783>
- [109] Chandra Sripada and Aman Taxali. 2020. Structure in the stream of consciousness: Evidence from a verbalized thought protocol and automated text analytic methods. *Consciousness and Cognition* 85 (Oct. 2020). <https://doi.org/10.1016/j.concog.2020.103007>
- [110] Standage, Tom. 2015. *Writing on the Wall: Social Media – The First 2,000 Years*. Bloomsbury Publishing USA.
- [111] Chenhao Tan. 2018. Tracing Community Genealogy: How New Communities Emerge from the Old. *Proceedings of the International AAAI Conference on Web and Social Media* 12, 1 (June 2018). <https://ojs.aaai.org/index.php/ICWSM/article/view/15003> Number: 1.
- [112] Nathan TeBlunthuis and Benjamin Mako Hill. 2022. Identifying Competition and Mutualism between Online Groups. *Proceedings of the International AAAI Conference on Web and Social Media* 16 (May 2022), 993–1004. <https://doi.org/10.1145/3531111.3531112>

[//ojs.aaai.org/index.php/ICWSM/article/view/19352](https://ojs.aaai.org/index.php/ICWSM/article/view/19352)

- [113] Veronica L Thomas, Marisol Chavez, Erica N Browne, and Alexandra M Minnis. 2020. Instagram as a tool for study engagement and community building among adolescents: A social media pilot study. *Digital Health* 6 (Jan. 2020), 2055207620904548. <https://doi.org/10.1177/2055207620904548>
- [114] Amaury Trujillo and Stefano Cresci. 2022. Make Reddit Great Again: Assessing Community Effects of Moderation Interventions on r/The_Donald. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 526:1–526:28. <https://doi.org/10.1145/3555639>
- [115] Jean M. Twenge, Brian H. Spitzberg, and W. Keith Campbell. 2019. Less in-person social interaction with peers among U.S. adolescents in the 21st century and links to loneliness. *Journal of Social and Personal Relationships* 36, 6 (June 2019), 1892–1913. <https://doi.org/10.1177/0265407519836170>
- [116] Galen Weld, Amy X. Zhang, and Tim Althoff. 2021. Making Online Communities 'Better': A Taxonomy of Community Values on Reddit. *arXiv:2109.05152 [cs]* (Nov. 2021). <http://arxiv.org/abs/2109.05152> arXiv: 2109.05152.
- [117] Wellman, Barry, Milena Gulia, and Marilyn Tremaine. 1999. Net Surfers Don't Ride Alone: Net Surfers Don't Ride Alone. *Communities in Cyberspace* (1999), 167–194.
- [118] Diyi Yang, Robert Kraut, and John M. Levine. 2017. Commitment of Newcomers and Old-timers to Online Health Support Communities. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6363–6375. <https://doi.org/10.1145/3025453.3026008>
- [119] Amy X. Zhang, Grant Hugh, and Michael S. Bernstein. 2020. PolicyKit: Building Governance in Online Communities. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. Association for Computing Machinery, New York, NY, USA, 365–378. <https://doi.org/10.1145/3379337.3415858>
- [120] Charles Chuankai Zhang, Mo Houtti, C. Estelle Smith, Ruoyan Kong, and Loren Terveen. 2022. Working for the Invisible Machines or Pumping Information into an Empty Void? An Exploration of Wikidata Contributors' Motivations. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (March 2022), 1–21. <https://doi.org/10.1145/3512982>
- [121] Shaoke Zhang, Hao Jiang, and John M. Carroll. 2011. Integrating online and offline community through facebook. In *2011 International Conference on Collaboration Technologies and Systems (CTS)*. 569–578. <https://doi.org/10.1109/CTS.2011.5928738>
- [122] Yan Zhang, Dan He, and Yoonmo Sang. 2013. Facebook as a Platform for Health Information and Communication: A Case Study of a Diabetes Group. *Journal of Medical Systems* 37, 3 (April 2013), 9942. <https://doi.org/10.1007/s10916-013-9942-7>
- [123] Haiyi Zhu, Robert E. Kraut, Yi-Chia Wang, and Aniket Kittur. 2011. Identifying shared leadership in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 3431–3434. <https://doi.org/10.1145/1978942.1979453>

A INTERVIEW QUESTIONS

We used the following interview questions as the basis for semi-structured interviews, asking additional follow-up questions as required:

- (1) How have you personally used Reddit in the past?
- (2) What led you to begin studying Reddit?
- (3) How would you describe your mental model of what a community is on Reddit?
- (4) Do you use a specific definition of “community” in your research?
- (5) What experience or literature (if any) are you drawing from in your conceptualization or definition of what a community is?
- (6) How do you operationalize this concept of community in your research?
- (7) In your work, which users do you consider to be (or include as) members of this community? (Or: How do you think about the boundaries of community membership)
- (8) (If quant) When you’re choosing how to structure predictive or descriptive tasks, how do you select which variables to include?
- (9) (If quant) How do you assign values to parameters of your operationalization that don’t necessarily have pre-existing “ground truths”? (Possibly: How do you justify your selections of methods or models?)
- (10) What do you think are the strengths/limitations of your operationalization?
- (11) Are there any variables you wish you had that you think would strengthen or validate your decisions?
- (12) To what degree do you feel that you explain these considerations in your papers? Is there anything you don’t write about—e.g., due to space, disciplinary norms, presentation or ethical concerns?
- (13) If a subreddit were to have a strong “sense of community,” what types of observable characteristics might you expect in that sub?
- (14) Do you have a sense of what types of features, norms, or mechanisms cause these characteristics?
- (15) How would you assess whether different subreddits or communities on Reddit are related or similar to each other?
- (16) Do you have ideas about mechanisms or reasons that cause users to: form new subreddits? migrate to other subreddits? leave the platform?
- (17) Anything else you’d like to discuss that I haven’t asked about?

Received January 2023; revised July 2023; accepted November 2023