2018

Learning in the wild: Coding for learning and practice on Reddit

Caroline A. Haythornthwaite

Priya Kumar
Ryerson University

Anatoliy Gruzd
Ryerson University

Sarah Gilbert
University of British Columbia

Marc Esteve del Valle
University of Groningen

See next page for additional authors

Recommended Citation

Haythornthwaite, C., Kumar, P., Gruzd, A., Gilbert, S., Esteve del Valle, M., & Paulin, D. (2018). Learning in the wild: Coding for learning and practice on Reddit. Authors original manuscript for paper published in Learning, Media and Technology, 43(3), 219–235. DOI: doi.org/10.1080/17439884.2018.1498356.

Learning in the Wild: Coding for Learning and Practice on Reddit

This is an original manuscript / preprint of an article published by Taylor & Francis in Learning, Media and Technology, 2018, available online: https://www.tandfonline.com/doi/full/10.1080/17439884.2018.1498356

Haythornthwaite, C., Kumar, P., Gruzd, A., Gilbert, S., Esteve del Valle, M., & Paulin, D. (2018). Learning in the wild: Coding for learning and practice on Reddit. *Learning, Media and Technology*, *43*(3), 219–235. DOI: doi.org/10.1080/17439884.2018.1498356

Caroline Haythornthwaite, School of Information Studies, Syracuse University, Syracuse, United States (chaythor@syr.edu)

Priya Kumar, Social Media Lab, Ryerson University, Toronto, Canada,

Anatoliy Gruzd, Ted Rogers School of Management, Ryerson University, Toronto, Canada

Sarah Gilbert, The iSchool, University of British Columbia, Vancouver, Canada

Marc Esteve Delle Valle, Department of Media Studies and Journalism, University of Groningen, Netherlands

Drew Paulin, School of Information, University of California, Berkeley, United States

Abstract

Learning on and through social media is becoming a cornerstone of lifelong learning, creating places not only for accessing information, but also for finding other self-motivated learners. Such is the case for Reddit, the online news sharing site that is also a forum for asking and answering questions. We studied learning practices found in 'Ask' subreddits AskScience, Ask_Politics, AskAcademia, and AskHistorians to develop a coding schema for informal learning. This paper describes the process of evaluating and defining a workable coding schema, one that started with attention to learning processes associated with discourse, exploratory talk, and conversational dialogue, and ended with including norms and practices on Reddit and the support of communities of inquiry. Our 'learning in the wild' coding schema contributes a content analysis schema for learning through social media, and an understanding of how knowledge, ideas, and resources are shared in open, online learning forums.

Keywords: informal learning, social media, coding, content analysis, Reddit

This work was supported by the Social Sciences and Humanities Research Council of Canada(SSHRC) under Insight Grant scheme.

Learning in the Wild: Coding for Learning and Practice on Reddit

Introduction

The Internet provides a wealth of ways to learn, from crowdsourced resources of online encyclopedias such as Wikipedia, how to videos on YouTube, online news, e-books, and open access journals to interactive learning opportunities, such as open courses and online interest groups. Then there are the wilds of open online discussions on sites such as Digg, Snapzu, Stacksity, Voat, and Reddit (Sankin 2017). These social media sites offer discussion that is led and moderated by contributors to the site. Discussions can be for play, social interaction, curiosity and learning. This is learning where there is no instructor, syllabus, or mandate to cover essential texts; no one earns a university degree or a workplace promotion from this kind of teaching or learning (at least not directly). Yet, they are sites where questions are asked, where crowds of participants comment, correct, and argue about answers, and where those who answer make the effort to present information in informed, accessible ways, often with citations to further resources. This informal learning takes place outside traditional educational environments, based on crowdsourced interest in just-in-time answering of posted questions. It is what we call 'learning in the wild' (with due acknowledgement of Hutchins' *Cognition in the Wild*). It is informal and non-formal learning taking place outside classroom settings, with what is asked about, answered, and learned at the discretion of those who ask and answer.

This paper investigates this type of open, online, informal learning, using the online news site Reddit as our case study. It focuses specifically on conversational patterns in 'Ask' subreddits, sites for discussion, engaged with knowledge dissemination and learning. We describe the application of content analysis to online social learning practices and the resultant coding schema. The latter is intended both for further use, testing, and extension by other researchers and as a basis for creation of automated classification systems for online learning conversations.

We ask these research questions:

- What patterns of online discourse operate in open learning environments?
- What do these patterns suggest are important discourse practices for operation of such environments?

- What do these patterns suggest are important for open, online, informal learning in these environments?
- What is the same or different across different discussion sites?

To address these questions, we draw on research addressing learning, communication, group behavior, and virtual communities to understand the practices that maintain these communities of inquiry.

Research in Open, Online Learning

Attention to open, online learning brings together research in communication, group behavior, information science, education and Internet research. This range is necessary because "learning, as a contemporary practice, fuses together a core set of related constructs: knowledge and information, tools, learners (from individual to collective), and spaces of engagement" (Ahn and Erickson, 2016, 81). Research at this intersection seeks to identify ways in which Internet technologies afford greater opportunities for connecting learning to personal experience and interest. Such ideas attend to connections among resources and people, including connecting online places and spaces into personal learning platforms, resources and people into personal learning networks, and experiences into connected learning (Authors 2015; Authors 2016; Luckin 2010). Benefits are seen in ways to manage individual learning, access to a wider range of information and lived experiences, engaging in social interaction in support of learning, and extending information worlds and practices outside formal educational settings (Davis and Fullerton, 2016, 110).

Trends in education, career trajectories, and the pace of change in knowledge point to the need for learning that is both lifelong and lifewide (Jackson 2011). Learning has always taken place outside classrooms, but the development of open, online forums provides the opportunity to study this 'learning in the wild'. Mindful of the growing importance of open learning for career and personal needs, and the range of learning occurring in online communities and groups, we set out to explore how learning unfolds in open, online environments.

Learning and Interaction in Reddit

To explore open, online learning, we directed our attention to Reddit. This site suggests itself as an ideal setting for examining learning practices because participation engages self-motivated

learners, occurs outside traditional settings (e.g., academic research, universities, workplaces), combines perspectives from experts and non-experts alike (Moore and Chuang 2017), and covers topics chosen and responded to according to member contributions.

Reddit is an online news sharing site, founded in 2005 by Steve Huffman and Alexis Ohanian. Reddit has become increasingly popular since its launch, ranking in 2018 as the third most visited site in the US. It maintains a relative stronghold as the go-to, self-organized community site for people interested in current affairs, social commentary, and Internet subcultures. Until a recent redesign, the site has appeared as a message board, with threaded conversations organized by topic (a design still available as 'Old Reddit'). Participants, known as 'Redditors', can post stories, links, photos, and videos that are openly shared online.

Redditors can also create a 'subreddit' for conversation around a particular topic, forming subcommunities with norms of their own. These can be started by any Redditor, and include such subreddits as "Change My View", where individuals post and challenge others to change their view (Heffernan 2018); or, The_Donald, for posting, discussion and commentary on Donald Trump. Reddit also includes a variety of "Ask" subreddits that tap into crowd knowledge, covering topics from science to professional practice.

The many, user-generated and user-managed subreddit communities afford learners opportunities to stay updated on a multiplicity of subjects, and engage with others to discuss, argue, and clarify positions. Two key features define the character of Reddit: posts are anonymous, although tied to a user identifier; and participants can 'upvote' or 'downvote' a posting to raise or lower its profile, resulting in crowd-driven attention that influences the visibility of postings. The anonymity has the great potential to lead to transgressions. Yet, the site is manageable due to member adherence to the rules and norms known as 'reddiquette' (Loudon 2014), and proper behavior is learned through observing the rewarding of behaviors that are consistent with site-wide and subreddit subcultures (Anderson 2015), such as upvoting and downvoting posts. These features contribute to our characterization of Reddit as a site for 'learning in the wild'. This signifies both the way the site supports user-initiated learning, and how Reddit culture privileges open, communication. These aspects highlight the importance of the internal behavior management of communal norms and moderator sanctioning that make it possible for subreddits to operate.

This research aims to understand practices and patterns of conversation, interaction and learning that support learning in the wild. To address this, we began by applying content analysis to online social learning practices to create a general coding schema for understanding open, online learning

practices. The development process included iterative code refinement as members of the team piloted and pre-tested the schema across four 'Ask' subreddits communities, that invite participants "to ask and answer questions that elicit thought-provoking discussions, as well as some lighter questions which will hopefully entertain and help you learn a little about your fellow redditors" (Ask Reddit Rules 2017, see: https://about.reddit.com). We chose subreddits 'AskScience', 'Ask_Politics', 'AskAcademia', and 'AskHistorians' to cover a range of Ask types. The multi-stage process drew on previous literature and schemas to address considerations we were aware deserved attention in online learning conversations, while also striving for a parsimonious schema that could be applied first by independent human coders and later in automated text analysis.

Framing for Coding Learning

Our team has been working for several years studying practices associated with learning online, observing and researching trends toward more learner-centered participation. While beyond the scope of this paper to review, in developing our coding schema we kept in mind what is known about group and community formation and maintenance, offline and online (Authors 2011; Authors 2009; McGrath and Hollingshead 1994; Preece 2000), how adults learn (Bransford, Brown, and Cocking 1999; Hase and Kenyon 2000), trends in e-learning (Haythornthwaite et al. 2016), connected learning (Siemens 2005), earlier work on discourse in learning environments (Gunawardena, Lowe, and Anderson 1997; Mercer 2004) and emerging research in learning analytics (Haythornthwaite, de Laat, and Dawson 2013; Lang et al. 2017).

Together, this literature reveals the way open, online participatory practices merge with learning practices and norms associated with these online settings. The following describes three key ideas that capture this combination and framed our analysis: *social learning*, *online community maintenance*, and *community of inquiry*.

Social Learning

Social learning holds that learning occurs through observation of and reaction to behaviors; the learner (e.g., a child), chooses whether to imitate the behavior according to reactions observed (Bandura 1977). For adults, apprenticeships provide a framework for learning by observing and doing in communities of practice, with master craftsmen modeling appropriate practice, and newcomers observing and learning through 'legitimate peripheral participation' (Lave and Wenger

1991). In open, online environments similar learning processes occur, as individuals lurk before posting, and as they observe others responding to and addressing inappropriate behavior (Haythornthwaite and Andrews 2011).

Others propose that online communities of practice manage a group Zone of Proximal Development (Gunarwardena et al. 2009). Collaboration among such peers build from multiple viewpoints and ideas that are actively shared, clarified, and contested by the individuals within the group (Goos, Galbraith and Renshaw 2002). Online learning outside formal settings also expands social activity, widening the scope of social learning. In analyzing such interactive practices, "the focus … is on processes in which learners are not solitary, and are not necessarily doing work to be marked, but are engaged in social activity, either interacting directly with others (for example, messaging, friending or following), or using platforms in which their activity traces will be experienced by others (for example, publishing, searching, tagging or rating)" (Buckingham Shum and Ferguson, 2012, 5).

In online learning environments, social learning occurs through discussion. Analyzing conversations offers much promise for identifying patterns of activity that indicate meaningful learning and knowledge construction (De Liddo et al. 2011). Previous research coding learning processes has focused on addressing formal settings (e.g. educational courses, conferences, teams), and have applied techniques and computational tools to a specific case or online phenomenon. For both, the aim has been to understand learning processes to suggest ways of improving practices. Studies have used quantitative predictive modeling to show how knowledge is constructed, disseminated and validated in open online settings (Ezen-Can and Boyer 2015); and how automated dialogue assessment tools improve collaboration in virtual classrooms, academic communities, and communities of practice (Iglesias-Pradas, Ruiz-de-Azcárate, and Agudo-Peregrina 2015; Nistor et al. 2015).

As the amount of online text in conversations increases, newer techniques aim to automate detection of interaction patterns. Discourse Centered Learning Analytics (DCLA) is an emerging area stemming from earlier work in computer mediated communication that analyzed the quality of interactions and learning experiences in collaborative environments (e.g., Gunawardena, Lowe, and Anderson 1997). Much of the work in DCLA is focused on analysis of contributions and contributors, by mapping contributions from participants to roles, or categories of discourse to productive, explanatory-seeking discussion (Chen and Resendes 2014). Other work focuses on statistical analysis of discourse, which allows for both modeling of individual contributions, and

modeling relationships among messages within an online community or network, and considers other variables such as demographics (Chiu and Fujita 2014).

Efforts to build automated process are still in the formative stage. Keeping previous work in mind in developing our coding schema, and our goal of eventual development of automated processing, we focused on work by others that best fit the online conversational style of open forums. In early exploration of the data, we found that the Interaction Analysis Model of Gunawardena, Lowe, and Anderson (1997), while highly applicable to formal settings for teaching and learning, was less well suited to the free-wheeling style of Reddit. Rather, we felt that approaches that stressed dialogue would be most appropriate. Thus, we began by following the efforts of Ferguson and colleagues who set out to identify elements of *exploratory dialogue* in a manner suitable for machine learning (Ferguson et al. 2013).

The idea of exploratory dialogue comes from the work of Mercer (2004), who identified three kinds of talk promoting learning in a classroom setting: *Disputational*, "characterised by disagreement and individualised decision making"; *Cumulative*, "in which speakers build positively but uncritically on what the others have said"; and *Exploratory*, "in which partners engage critically but constructively with each other's ideas" (Mercer, 2004, 146). In keeping with Ferguson et al, we built on Mercer's (2004) exploratory talk because it represents the kind of constructive interaction that reflects adult, collaborative learning most likely to advance both individual and group knowledge.

As Mercer describes it, in exploratory talk, statements and opinions are open for joint discussion and debate, and can be publicly challenged through alternative methods of reasoning and hypotheses (Mercer 2004). We expect this kind of exploratory talk to support informal learning because online textual discussions involve active processes of co-reasoning and negotiation, and knowledge, idea or resource sharing (Ferguson et al. 2013). We assume that where we find exploratory talk, we are finding learning to have occurred. However, we stress that our aim is to understand online processes in the service of learning and we are not addressing individual learning outcomes.

Online Community Maintenance

Beyond subject learning, online conversations also contribute to group practice and the maintenance of the online community. Learners entering online conversations join or create new communities of practice where rules and norms are defined and reinforced. Research on virtual

communities, group behavior, and professional apprenticeship emphasize how norms are (re)created through awareness and interaction, with new users learning how to become members of the community (Authors 2016; Preece and Maloney-Krichmar 2005). The need for such learning is evident even in the terms used for new users – newbies, apprentices, lurkers – and for more advanced users – experts, wizards, gurus; and in the support mechanisms created for new user integration, such as FAQ lists (Frequently Asked Questions), and practices of lurking as a means of learning the practices of an online community (Preece, Nonnecke, and Andrews 2004), each supporting legitimate peripheral participation (Eberle, Stegmann, and Fischer 2014). Group maintenance practices include sanctioning those who do not follow the rules, keeping participants in line about appropriate language and genre of postings, allowing newbies to observe the consequences of not following the norms. In applying this background to our coding schema, we looked for practices that paid attention to and reinforced community norms about conversation style, topic, citation practices, etc.

Community of Inquiry

The community of inquiry (CoI) framework defined by Garrison and colleagues (Garrison 2009; Garrison, Anderson, and Archer 2001) provides a more focused view of community practice and maintenance specifically addressing online learning contexts. The framework combines attention to learning processes with the roles and practices of the community, and particularly the active role of both instructor and learner. While the framework was first developed to make sense of practices in online education programs, it has been usefully applied to address pedagogical strengths and weaknesses in Massively Open Online Courses (MOOCs) (Amemado and Manca 2017), to theorize about the role of instructors in building knowledge-sharing communities (Tomkin and Charlevoix 2014), and to consider how social networking sites and social media can support learning communities for students (Keles 2018; Lim and Richardson 2016).

The open, online learning setting of Reddit is a significant departure from Garrison and colleagues' initial CoI learning environment. Accordingly, we use the framework to delve further into how communicative, social, and personalized signals of online learning appear in Reddit text-based discussions (Borup, West, and Graham 2012). In Reddit, as in other learning communities, active engagement is key to maintaining the community. Thus, active engagement is expected as part of the learning process, as it is in the CoI framework, which emphasizes that online teaching and learning entails much, "more than simply accessing information and participating in chat rooms" (Garrison, 2003, 2). Mercer's exploratory talk is thus an integral part of the kind of engagement

and interaction that is expected for a learning community, enabling reflective inquiry and communication, and intertwining the public, personal and private worlds of the learner (Garrison 2003).

Thus, in building our coding schema for Reddit text-based discussions, we analyze the online content for evidence of CoI signals of learning. In CoI, these comprise in-depth, collaborative, and constructivist learning experiences through three interdependent elements: *cognitive presence*, *social presence,* and *teaching presence*. Cognitive presence refers to "the extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication" (Garrison, Anderson, and Archer, 2001, 11), and can include phases of *triggering* (identifying an issue), *exploration* (brainstorming), *integration* (construct meaning), and *resolution* (testing or implementing solutions). For our study, we note that when Redditors engage in a process of individual reflection and knowledge development, they are also collectively contributing to the wider subreddit community discourse.

Social presence is "the ability of participants to identify with the community …, communicate purposefully in a trusting environment, and develop inter-personal relationships by way of projecting their individual personalities" (Garrison, 2009, 352). Redditors who express their opinions and share insights openly can play an active role in shaping the discourse of learner-learner interactions in group-based online learning environments, providing a way for others to get to know who is asking and answering a question. Such opinions can also include matters of how the group or subreddit operates, both developing and sustaining interactive practices that allow participants to understand the way to behave in the subreddit, and to trust the behavior of others will be managed.

Finally, *teaching presence* constitutes "the design, facilitation and direction of cognitive and social processes for the purpose of realizing personally meaningful and educationally worthwhile learning outcomes" (Anderson et al., 2001, 5). Online teaching activities help set the tone for learning through curriculum choices and course organization, instructional design, and discourse facilitation, all of which contribute to meaning making. In Reddit, moderators set rules, norms, and codes of behavior. These play a role in shaping the broader learning climate of the online community, and signal the status and presence of those who take on the teaching role, both for topic content (e.g., experts) and for group maintenance (e.g., moderators).

Examining Reddit

While our overall aim is to develop a general coding schema that will hold across different online, informal learning settings, at first instance we defined and refined our coding by working with these four Ask subreddits (descriptions from the subreddit sites, June 2018):

- AskScience. "[A] forum for answering science questions. It aims to promote scientific literacy by helping people understand the scientific process and what it can achieve"; created in September 2008; a default subreddit to which users are automatically subscribed. As of June 2018, AskScience had 15,568,080 subscribers;
- Ask_Politics. "The goal of this subreddit is the promotion of political knowledge by disseminating knowledge of law and policy considerations that drive our representatives and other government actors"; created October 2011; 29,157 subscribers;
- AskAcademia: "This subreddit is for discussing academic life, and for asking questions directed towards people involved in academia, (both science and humanities)"; created January 2011; 46,803 subscribers;
- AskHistorians: "Questions about the past: Answered!"; created August 2011; 762,558 subscribers.

Development of the Coding Schema

The coding schema was developed through three stages of iteration. In all stages, the coders were researchers from our team, each aware of the literature in this area, the kinds of learning processes that might occur, and the research aims. Coders included two doctoral students, one post-doctoral fellow, and three faculty holding university positions. One was a long-time Reddit user, researching motivations to participate in open, online initiatives, who acted as the group's Reddit cultural advisor. The post-doctoral fellow was designated as the 'primary coder' with responsibility for managing the coding process and gathering input individually and collectively from coders. In general, the research team met weekly in a team Skype meeting and coding experiences were shared. The coders applied each version of the schema to subreddit datasets, and then engaged in discussion about the pros and cons of particular codes, the range of activity that should be coded, and how the codes should be refined. The resulting redefined coding schema was then used as the basis of the next stage of coding.

Stage I: Exploratory Dialogue and Intra-group Behavior

In Stage I, we adopted Ferguson et al's (2013) cue phrases framework comprising seven categories, described in Table 1: *Critique*; *Discussion of Resources*; *Evaluations*; *Explanations*; *Explicit Reasoning*; *Justifications*; *Others' Perspectives*. These cue phrases were developed and piloted by Ferguson and Buckingham Shum (2011) and colleagues in a series of studies that added a qualitative layer to quantitative data through self-trained (automatic) detection of exploratory and non-exploratory dialogue. A key benefit of their research agenda is that it combines manual cue-phrase coding with computational linguistics/machine learning classification techniques, a future direction for our work. Because of the open nature of the Reddit environment, and its greater similarity to online group behavior and virtual community practices (Authors 2006), our schema was extended with two additional categories addressing group behavior. *Learning the Rules* was added to capture the dialogue acts and content submissions pertaining to community maintenance, e.g., following subreddit norms and guidelines that explain how to be an effective contributor or member of the community. *Socializing* was added to capture the human context of Reddit conversations, which reflect forming and reinforcing social bonds with others, e.g., through positive expressions of gratitude or approval, and negative expressions relating to confrontation or opposition (see Table 1. Reddit Codebook Version 1).

In Stage I coding, we used DiscoverText (http://discovertext.com), a cloud-based text-analysis software that allowed assignment of multiple coders to the same dataset. The first cycle of coding was undertaken on a dataset of one percent of 2015 subreddit posts (excluding parent submissions) from three subreddits: AskScience (n=163), Ask_Politics (n=189), and AskAcademia (n=197). Each sample was coded by three coders.

Our Stage I coding did not provide a satisfactory result. Intercoder reliability showed low agreement among coders, with Krippendorf's alpha scores from .16 to .22 where .67 to .80 is considered a good level of agreement (AskScience .22; Ask_Politics .16; AskAcademia .2). Coders had difficulties distinguishing between Ferguson et al's (2013) cue phrases for *Explanation* versus *Explicit Reasoning*, and *Discussion of Resources* versus *Others' Perspectives*, and coding for dialogue that could be described as information seeking and knowledge sharing. Coders were confused with dialogue in the form of questions on whether these were rhetorical, conversational, or seeking further clarification. Finally, coders were unable to distinguish between Socializing, Critique (negative commentary or disagreement) and Evaluation (positive commentary or agreement).

Stage II: Reducing and Refining Codes

In Stage II, we aimed to capture more precisely the socializing, and resource and information elements of informal online learning, refine codes relating to discussion of resources, and delete little used codes (Table 2). The *Socialization* code was refined to capture the valence of feelings using codes *Explanation* (neutral), *Evaluation* (positive/agree), and *Critique* (negative/disagree); *Justification*, and *Others' Perspectives* were removed due to lack of use; and *Information Seeking* was added to address general inquiry, asking for help or clarification. In Stage II, we allowed for multiple coding of posts (up to three per comment) because many single Reddit comments exhibited several different dialogue processes.

Despite these efforts, sufficient coding issues remained at the end of Stage II that we decided on a different approach. The two stages had provided increased understanding of the elements of learning dialogue in the Ask subreddits. Given this knowledge, and the need to arrive at a repeatable coding scheme, we made the collective decision to revise and rewrite our codebook in its entirety.

Stage III: Fully Revised Codebook

Version 3, our fully revised coding schema, is a significant departure from the premise of Ferguson et al.'s (2013) coding used in the previous stages. In Stage III, we simplified the categories to facilitate use of the codes, standardize multi-coder agreement, and address the types of exploratory learning dialogue we were observing.

Version 3 of the codebook (Table 3) addressed and captured two trends observed in reading Reddit posts: the positive expressions, supportive dialogue and information provision that pull participants toward each other and foster topic-specific discussions; and the more negative exchanges that monitor and sanction behavior, silence participants, and can stifle online learner dialogue. Accordingly, the revised schema extended the identification of the valence of emotion to three explicit categories for Explanation, *Neutral, Agreement* and *Disagreement,* and two for Socializing, *Positive* and *Negative*. Coding was refined to identify two distinct types of information exchange, *Information Seeking,* and *Providing Resources*. Only one aspect of learning about internal Reddit culture was coded for: *Subreddit Rules and Norms*.

Our test of the Version 3 coding schema resulted in a more acceptable level of agreement between coders, with Krippendorf's alpha of .52 to .67 (AskScience, .67, Ask_Politics, .52, and AskAcademia, .64; Table 4). Thus, we settled on Version 3 as our final coding schema. We then

extended testing to apply the schema to a sample from the subreddit AskHistorians from 2015 (n=267). Agreement between coders was an alpha of .57. While these values are of moderate agreement, they are much stronger than our Version 1 coding schema. Along these lines, we note that Ferguson et al.'s (2013) binary classification (exploratory or non-exploratory dialogue) recorded an inter-annotator agreement score of .597, which they understood as having 'moderate agreement', and thus reliable enough to train an automated classifier. In designing our study on exploratory learning dialogue, we anticipated that adding multiple coders (3) and codebook categories (8) to our methodology could potentially produce lower levels of intercoder agreement (DeCuir-Gunby, Marshall, and McCulloch 2011; Krippendorff 2004). Yet, our results still fall in line with moderate agreement. At the end of this stage, we decided to test the validity of our coding schema with independent coders on a much larger and more recent sample of Reddit data.

Applying the Coding Schema

Data for the schema testing with independent coders was collected using a custom web application (available at: https://collector.socialmediadata.org) that used Reddit's public API (https://www.reddit.com/dev/api/). Since Reddit users do not use their real names and we only collected publicly available data, consent was not considered necessary to solicit from Reddit users or platform intermediaries. We sampled one percent of public Reddit comments posted in 2016 from our four Ask subreddits (Table 5; since the dataset was collected retroactively, it does not include comments deleted by authors or moderators.) The sample comments were then manually coded by three independent coders each of whom had first completed a schema tutorial training-module.

Results (Table 6) from the three independent coders showed agreement statistics of acceptable levels from 72-79 percent (Krippendorff's alpha: AskScience .69, 78 percent; Ask_Politics .60, 72 percent; AskAcademia .64, 77 percent; and AskHistorians .76, 79 percent). We regard these alpha levels as acceptable considering that coders could apply up to three codes per comment. For exploratory studies like ours, alpha levels between .67 and .80 are considered reliable enough to draw out and develop cautionary conclusions (Hayes and Krippendorff 2007; Krippendorff 2004).

We note the comparatively lower levels of agreement in the Ask_Politics and AskAcademia subreddits. Both of these communities exhibit a more conversational and personalized style of dialogue than the more transactional question and answer discourse of AskScience and AskHistorians. We believe these cognitive elements are more challenging for coders to breakdown and categorize. On multiple occasions, we found that sample posts from these subreddits could be

argued to display different levels of deliberation and types of dialogue (e.g., subtle disagreements that were neutral in tone). According to Hayes and Krippendorf (2007), human coders as observers are trained to make judgments of *kind* (e.g., what category does this unit belong to?), *magnitude* (e.g., how pronounced is the unit attribute?), or *frequency* (how often is it occurring?). It is likely our coders differed in their judgments of *kind*, *magnitude* and *frequency*, that is, they did judge similarly the overall prominence of a type of exploratory talk or conversational dialogue being communicated through content. Further, Riff et al. (2014) add that the degree of connotative and denotative meanings attached to words and symbols can present complex challenges when attempting to achieve high levels of intercoder reliability. For example, coding news stories for different topics and subjects would be much easier and likely to achieve higher levels of intercoder agreement than coding the valence (positive or negative) of said news stories (Lacy et al. 2015). These are areas to pursue in the future.

Despite the difficulties and moderate agreement levels among coders, we felt these results were sufficient to give insight into the learning and community processes in these subreddits, along with the type and range of expressions associated with 'learning in the wild'.

To illustrate how the coding schema identified learning processes in Reddit, we present the final count results for the 2016 data where two or more coders agreed on the same code, and discuss what these tell us about learning processes in Reddit. We then conclude by offering additional insights on the communicative, collaborative and knowledge-rich learning environments being fostered in social media.

Results

Results of the application of the coding schema to the four subreddits by independent coders reveal subtle nuances in the way people converse and participate across different subreddit communities. Much like a strand of DNA, each subreddit maintains its own unique signatures that contributes to the discourse of the online community (see Table 6; Figure 1).

Results show that all four subreddits demonstrate a substantial proportion of neutral comments (43-50 percent), with differences found in the balance of positive and negative explanation, and in information seeking and resource provision. Unsurprisingly, Ask_Politics has the greatest negative valence, with the highest percentage of comments coded as Explanation with Disagreement (18 percent), Socializing with Negative Intent (5 percent), and Subreddit Rules and Norms (10 percent). AskAcademia and AskHistorians lead on positive valence interactions: AskAcademia

has the highest Explanation with Agreement (12 percent), Socializing with Positive Intent (17 percent), and the fewest on Subreddit Rules and Norms (1 percent); AskHistorians explanation remains primarily neutral (with only 6 and 4 percent with disagreement or agreement), but with Socializing with Positive Intent (17 percent) equal to that of AskAcademia. AskScience and AskHistorians lead on Information Seeking requests (18 and 22 percent), and backing that up with Providing Resources (12 and 21 percent).

These results reflect the norms and rules associated with each subreddit, but also the nature of the topic, the community, and interaction practices of each. Ask_Politics rules and norms stipulate that posts should be reputable, civil, sourced and remain on-topic, but the very personal and normative nature of politics seems to fuel more volatile comments on a topic where it may be said there is no 'right' or 'wrong' answer in an objective sense. By contrast, both AskScience and AskHistorians are seeking to explain through evidence rather than opinion; norms and rules encourage civility, evidence, external sources and academic-level answers. These two subreddits bear similarity to the professionally-oriented AskAcademia in supporting an apprentice-type inclusion, with common future goals and practices.

Discussion

Our aim in this research has been two-fold: to define a working coding schema for examining open, online learning in the wild; and to see how patterns of interaction in such a learning environment differ across discussion sites. This has been framed with attention to social learning, community maintenance, and the community of inquiry framework, for the case of Reddit 'Ask' subreddits. In creating our final coding schema, we began by looking at the coding schemas of others such as the Interaction Analysis Model of Gunarwardena and colleagues (1997), and Mercer's exploratory dialogue as applied by Ferguson and colleagues (2013). However, through preliminary evaluation and two rounds of schema testing, neither of these approaches lent themselves well to a set of codes that captured the nature of interaction in Reddit reliably across coders. Yet, this background and our continued evaluation of our coding efforts and subreddit communications informed the development of our final coding schema.

This final coding schema includes codes that show the way learning happens in these subreddits: discussion begins with topic-oriented postings *seeking information*; topics are further explored and evaluation with *explanation* with a *positive, negative* or *neutral* valence that provide comment on previous comment and/or adding new ideas or facts to the discussion; veracity of answers is

supported through *providing references.* Community practices are maintained explicitly through postings about *subreddit rules and norms*. As well, non-topic *socializing* postings add with either *positive* (praise, irony, humour) or *negative* (insult, abuse) valence add to the informality of the venue.

Social learning is demonstrated in a number of ways in these Ask communities. Explanations represent the practice of learning from and with others. Equally important are the opening forays into seeking information, where individuals begin the process of engaging with others in the service of learning. Experts who respond, e.g., through explanation, do so in a reciprocal social learning role, the teacher role in response to the learner. In keeping with ideas of apprenticeship, experts and moderators also model and instruct in proper answering, e.g., in providing resources to justify claims, and sanctioning off-topic or non-conforming answers.

The success of this final coding schema allows for further interpretation associated with learning in open, online environments. Reddit, and the Ask subreddits, are, perhaps, a prime example of user-managed discussion and self-regulation, particularly given the 'free-speech' ethos associated with the site. Yet, the subreddits examined operate successfully as sites for information, learning, and knowledge sharing. Without both continued usefulness of topic information, and useful management of discussion practices, the sites would be unlikely to remain as active information environments. Thus, online community maintenance looms large in making these sites viable, and this is learned and achieved through online interaction practices.

Since direct comments on subreddit rules and norms takes up only a small proportion of the discussion (even with the 10 percent for Ask_Politics), our coding suggests that Ask subreddit participants both create and maintain their community of inquiry through participating and structuring of information and learning practices. Cognitive presence appears to be represented in the explanation codes. Explanations – regardless of valence – promote continued attention to and development of a topic, and retain engagement in evaluation and learning. Moreover, different kinds of explanations can expand the number of views on a subject and/or provide different explanations for understanding a particular topic, providing more ways to engage with a topic. Social presence, the ability to identify with a community, is represented in 'Socializing with positive intent' which is more strongly evident in the AskAcademia and AskHistorians. However, we can also see both positive and negative ways that Redditors 'project their individual personalities' in other sites, both in offering explanations – whether good, bad or neutral – and in 'Socializing with negative intent'. Each of these does project a personality, whether through the altruism of a detailed explanation, or the forceful expression of a personal reaction to others' ideas.

Teaching presence is also expressed through explanations, and instructing others about rules and norms, as well as being in the duties of those in the designated moderator roles.

Limitations

Overall, we find that the coding schema picks up in a general way on social learning, community maintenance, and cognitive, social and teaching presence. It captures well the extent and valence of explanation and socializing, community practices of information seeking, answering with reference to resources, and learning and following social norms. However, unlike other coding schemas for learning interactions, we were unable to reliably distinguish in more detail the process of argumentation (whether based on Gunarwardena's Interaction Analysis Model, or Mercer's exploratory dialogue). This may reflect the nature of the anonymous, just-in-time, question-and-answer engagement in topic development that happens in these Ask subreddits, yet it represents a limitation in our schema for exploring in-depth learning processes. However, the schema does give an overview of online discourse practices in the studied open learning environments, showing which of these were most important as well as the variation across different discussion sites. Results show the importance of managing conversation practices in such environments, as well as the way these are successfully managed in the open, online Reddit environment.

Future Directions

Future plans are to this research by validating the proposed coding schema with a wider sample of subreddits (for example, 'Explain Like I'm Five' and 'Today I learned'), and later to other social media platforms. Further, while hand coding was applied in the first instance, it is the aim of this research to apply Natural Language Processing (NLP) techniques to allow the analysis of the large datasets found for learning in open, online settings. 'Supervised' machine learning is a commonly used approach in NLP, which entails coding a sample set of data (as done here) and then creating an algorithm that classifies sufficiently accurately on the training dataset to provide confidence that the coding of the full dataset will also be suitably accurate. NLP techniques are beginning to be brought into analysis of learning and online argumentation. They have been used to automatically identify learning versus social conversation in MOOCs (Wise et al. 2017); to address linguistic indicators of an online comment's persuasive power in Reddit (Khazaei, Xiao, and Mercer 2017). NLP can be used for sentiment analysis to identify the valence of comments, i.e., positive or negative, agreement or disagreement; and for Argumentation Mining, which aims to

detect “all the arguments involved in the argumentation process, their individual or local structure… and the interactions between them” (Mochales-Palau and Moens 2009, 98).

Conclusion

As more learning goes online, and as more resources and venues spring up to support learning in the wild, the more we will depend on learning via asynchronous collaborative spaces of engagement similar to Reddit. Examining learning processes in such forums can extend our understanding of how learning and conversation around both objective and subjective topics happens outside the classroom. Our ‘learning in the wild’ coding schema was designed to account for learning that takes place outside traditional educational institutions, where there may not be a clear distinction between teachers, experts and groups of students. The schema has been useful for demonstrating the ‘DNA’ of different learning communities, highlighting processes associated with informal social learning, cognitive, social and teaching presence, and community maintenance. As we work further to validate our coding schema, and develop automated analysis techniques, we invite other scholars to apply our schema to their research.

References

Ahn, J., and I. Erickson. 2016. “Revealing Mutually Constitutive Ties between the Information and Learning Sciences.” *The Information Society* 32 (3): 81-84.

Amemado, D., and S. Manca. 2017. “Learning from Decades of Online Distance Education: MOOCs and the Community of Inquiry Framework.” *Journal of E-Learning and Knowledge Society* 13 (2). https://www.learntechlib.org/p/180225/.

Anderson, K. E. 2015. “Ask Me Anything: What Is Reddit?” *Library Hi Tech News* 32 (5): 8–11. https://doi.org/10.1108/LHTN-03-2015-0018.

Anderson, T., and D. R. Garrison. 2003. *E-Learning in the 21st Century: A Framework for Research and Practice*. New York, NY: Routledge.

Anderson, T., R. Liam, D.R Garrison, and W. Archer. 2001. “Assessing Teaching Presence in a Computer Conferencing Context.” https://auspace.athabascau.ca/handle/2149/725.

Bandura, A. 1977. “Self-Efficacy: Toward a Unifying Theory of Behavioral Change.” *Psychological Review* 84 (2): 191–215.

Borup, J., R. E. West, and C. R. Graham. 2012. “Improving Online Social Presence through Asynchronous Video.” *The Internet and Higher Education* 15 (3): 195–203. https://doi.org/10.1016/j.iheduc.2011.11.001.

Bransford, J. D., A. L. Brown, and R. R. Cocking. 1999. *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy Press.

Chen, B., and M. Resendes. 2014. "Uncovering What Matters: Analyzing Transitional Relations Among Contribution Types in Knowledge-Building Discourse." Proceedings of the 4th International Conference on Learning Analytics and Knowledge, ACM, 226-230.

Chiu, M. M., and N. Fujita. 2014. "Statistical Discourse Analysis of Online Discussions: Informal Cognition, Social Metacognition and Knowledge Creation." Proceedings of the 4th International Conference on Learning Analytics and Knowledge, ACM, 217–225.

Davis, K., and S. Fullerton. 2016. "Connected Learning in and after School: Exploring Technology's Role in the Learning Experiences of Diverse High School Students." *The Information Society* 32 (2): 98–116. https://doi.org/10.1080/01972243.2016.1130498.

DeCuir-Gunby, J. T., P. L. Marshall, and A. W. McCulloch. 2011. "Developing and Using a Codebook for the Analysis of Interview Data: An Example from a Professional Development Research Project." *Field Methods* 23 (2): 136–55. https://doi.org/10.1177/1525822X10388468.

De Liddo, A., S. Buckingham Shum, I. Quinto, M. Bachler, and L. Cannavacciuolo. 2011. "Discourse-centric Learning Analytics." Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 23-33.

Eberle, J., K, Stegmann, and F. Fischer. 2014. "Legitimate Peripheral Participation in Communities of Practice: Participation Support Structures for Newcomers in Faculty Student Councils." *Journal of the Learning Sciences*, 23(2), 216-244.

Ezen-Can, A., and K. E. Boyer. 2015. "Understanding Student Language: An Unsupervised Dialogue Act Classification Approach." *JEDM - Journal of Educational Data Mining* 7 (1): 51–78.

Ferguson, R., and S. Buckingham Shum. 2011. "Learning Analytics to Identify Exploratory Dialogue Within Synchronous Text Chat." Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 99-103.

Ferguson, R., Z. Wei, Y. He, and S. Buckingham Shum. 2013. "An Evaluation of Learning Analytics to Identify Exploratory Dialogue in Online Discussions." Proceedings of the 3rd International Conference on Learning Analytics and Knowledge, ACM, 85–93.

Garrison, D. 2003. *Cognitive Presence for Effective Asynchronous Online Learning: The Role of Reflective Inquiry, Self-Direction and Metacognition*. Vol. 4.

Garrison, D. R. 2009. "Communities of Inquiry in Online Learning." In *Encyclopedia of Distance Learning, Second Edition*, 352-355, IGI Global.

Garrison, D. R., T. Anderson, and W. Archer. 2001. "Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education." *American Journal of Distance Education* 15 (1): 7–23. https://doi.org/10.1080/08923640109527071.

Authors. 2011. Removed for blind review.

Goos, M., P. Galbraith, and P. Renshaw. 2002. "Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving." *Educational Studies in Mathematics* 49 (2):193-223.

Authors. 2016. Removed for blind review.

Gunawardena, C.N., M.B Hermans, D. Sanchez, C. Richmond, M. Bohley, and R. Tuttle. 2009. "A Theoretical Framework for Building Online Communities of Practice with Social Networking Tools." *Educational Media International,* 46 (1), 3-16.

Gunawardena, C. N., C. Lowe, and T. Anderson. 1997. "Analysis of a Global Online Debate and the Development of an Interaction Analysis Model for Examining Social Construction of

Knowledge in Computer Conferencing." *Journal of Educational Computing Research*, 17 (4), 397–431.
Hase, S., and C. Kenyon. 2000. "From Andragogy to Heutagogy." *Ulti-BASE In-Site*. https://epubs.scu.edu.au/gcm_pubs/99.
Hayes, A. F., and K. Krippendorff. 2007. "Answering the Call for a Standard Reliability Measure for Coding Data." *Communication Methods and Measures* 1 (1): 77–89. https://doi.org/10.1080/19312450709336664.
Authors. 2006. Removed for blind review.
Authors. 2009. Removed for blind review.
Authors. 2015. Removed for blind review.
Haythornthwaite, C., and R. Andrews. 2011. *E-Learning Theory and Practice*. SAGE.
Haythornthwaite, C., R. Andrews, J. Fransman, and E. M. Meyers. 2016. *The SAGE Handbook of E-Learning Research*. SAGE.
Haythornthwaite, C., M. de Laat, and S. Dawson. 2013. "Introduction to the Special Issue on Learning Analytics." *American Behavioral Scientist* 57 (10): 1371–79. https://doi.org/10.1177/0002764213498850.
Heffernan, V. 2018. "Our Best Hope for Civil Discourse Online is on … Reddit." *Wired*. Retrieved from: https://www.wired.com/story/free-speech-issue-reddit-change-my-view/
Iglesias-Pradas, S., C. Ruiz-de-Azcárate, and Á. F. Agudo-Peregrina. 2015. "Assessing the Suitability of Student Interactions from Moodle Data Logs as Predictors of Cross-Curricular Competencies." *Computers in Human Behavior*, 47: 81–89. https://doi.org/10.1016/j.chb.2014.09.065.
Jackson, N. J. 2011. *Learning for a Complex World: A Lifewide Concept of Learning, Education and Personal Development*, Author House.
Keles, E. 2018. "Use of Facebook for the Community Services Practices Course: Community of Inquiry as a Theoretical Framework." *Computers & Education* 116: 203–24. https://doi.org/10.1016/j.compedu.2017.09.003.
Khazaei, T., L. Xiao, and R. Mercer. 2017. "Writing to Persuade: Analysis and Detection of Persuasive Discourse." Proceedings of the 2017 iConference. Retrieved June 20 from: http://hdl.handle.net/2142/96673
Krippendorff, K. 2004. "Reliability in Content Analysis." *Human Communication Research* 30 (3): 411–33. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x.
Lacy, S., B. R. Watson, D. Riffe, and J. Lovejoy. 2015. "Issues and Best Practices in Content Analysis." *Journalism & Mass Communication Quarterly* 92 (4): 791–811. https://doi.org/10.1177/1077699015607338.
Lang, C., G. Siemens, A. Wise, and D. Gašević. 2017. *Handbook of Learning Analytics*. First Edition. https://solaresearch.org/hla-17/.
Lave, J., and E. Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press.
Lim, J., and J. C. Richardson. 2016. "Exploring the Effects of Students' Social Networking Experience on Social Presence and Perceptions of Using SNSs for Educational Purposes." *The Internet and Higher Education* 29 (April): 31–39. https://doi.org/10.1016/j.iheduc.2015.12.001.
Loudon, M. 2014. "'Research in the Wild' in Online Communities: Reddit's Resistance to SOPA." *First Monday* 19 (2). http://dx.doi.org/10.5210/fm.v19i2.4365.

Luckin, R. 2010. *Re-Designing Learning Contexts: Technology-Rich, Learner-Centred* Ecologies, Routledge.

McGrath, J., and A. Hollingshead. 1994. *Groups Interacting with Technology: Ideas, Evidence, Issues, and an Agenda*. Thousand Oaks, Ca: SAGE Publications.

Mercer, N. 2004. "Sociocultural Discourse Analysis: Analysing Classroom Talk as a Social Mode of Thinking." *Journal of Applied Linguistics* 1 (2): 137–68.

Mochales-Palau, R., and M.F. Moens. 2009. "Argumentation mining: the detection, classification and structure of arguments in text." Proceedings of the 12th International Conference on Artificial Intelligence and Law, ACM, 98-107.

Moore, C., and L. Chuang. 2017. "Redditors Revealed: Motivational Factors of the Reddit Community." Proceedings of the 50th Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2017.279.

Nistor, N., Ş. Trăuşan-Matu, M. Dascălu, H. Duttweiler, C. Chiru, B. Baltes, and G. Smeaton. 2015. "Finding Student-Centered Open Learning Environments on the Internet: Automated Dialogue Assessment in Academic Virtual Communities of Practice." *Computers in Human Behavior*, 47: 119–27. https://doi.org/10.1016/j.chb.2014.07.029.

Authors. 2016. Removed for blind review.

Preece, J. 2000. *Online Communities: Designing Usability and Supporting Socialbilty*. New York, NY, USA: John Wiley & Sons, Inc.

Preece, J., and D. Maloney-Krichmar. 2005. "Online Communities: Design, Theory, and Practice." *Journal of Computer-Mediated Communication* 10 (4). https://doi.org/10.1111/j.1083-6101.2005.tb00264.x.

Preece, J., B. Nonnecke, and D. Andrews. 2004. "The Top Five Reasons for Lurking: Improving Community Experiences For Everyone." *Computers in Human Behavior,* 20(2), 201–223. http://dx.doi.org/10.1016/j.chb.2003.10.015

Riff, D., S. Lacy, and F. Fico. 2014. *Analyzing Media Messages: Using Quantitative Content Analysis in Research*. Routledge.

Sankin, A. 2017. "7 Sites to Try during Reddit's Meltdown | The Daily Dot." 2017. https://www.dailydot.com/layer8/reddit-alternatives-goodbye-cruel-world/.

Shum Buckingham, S., and R. Ferguson. 2012. "Social Learning Analytics." *Journal of Educational Technology & Society* 15 (3): 3-126.

Siemens, G. 2005. "Connectivism: A Learning Theory for the Digital Age." ELearnSpace. April 5. http://www.elearnspace.org/Articles/connectivism.htm.

Tomkin, J. H., and D. Charlevoix. 2014. "Do Professors Matter?: Using An A/B Test To Evaluate The Impact Of Instructor Involvement On MOOC Student Outcomes." Proceedings of the First ACM Conference on Learning@ Scale Conf., ACM, 71-78.

Wise, A. F., Y. Cui, W. Jin, and J. Vytasek. 2017. "Mining for Gold: Identifying Content-Related MOOC Discussion Threads across Domains through Linguistic Modeling." *The Internet and Higher Education* 32: 11-28.

Tables and Figures

Table 1. Reddit Codebook Version 1

Code	Definition	Linguistic Dialogue Example
1. Critique	The comment suggests disagreement; something may be wrong, faulty or in need of correction/ revision/ reassessment.	'However', 'not sure', 'maybe', 'hmm not really', 'think it through', 'actually, not exactly'
2. Discussion of Resources	The comment references and provides details of additional outside resources (e.g: links to external websites, forums, books, articles) to support understanding or extend discussion.	'Have you read', 'more links', 'check this out', 'look at', 'read this'…BOTH online and offline resources
3. Evaluations	The comment appraises and assesses the merit, worth and/or significance of something.	'Likely', 'good point/example', 'could be', 'fair enough'
4. Explanations	The comment has a descriptive quality and undertakes a process of 'thinking it through' by explaining, brainstorming and justifying a position or idea.	'Means that', 'our goals', 'the aim is', 'meaning', 'it depends, for example'
5. Explicit Reasoning	The comment works out ideas in a logical manner, often reaching a conclusion or proving a point through example based inferences. This includes taking the same line of argument further through questions/objections.	'Next steps', 'relates to', 'that's why', 'then you would', conditional 'if X then Y', 'along these lines'
6. Justifications	The comment reasons/expresses/offers judgment in terms of something already known or found.	'I mean', 'we learned', 'we observed', 'based on'
7. Others' Perspectives	The comment extends discussion by putting forward additional/alternative views and positions, increasing the range of an idea.	'Agree', 'another way to look at it', scholar/public figure argument, 'their research focuses on', 'through this lens'
8. Learning the Rules	The comment references the Reddit platform and may remind users of the protocol/code of conduct for the particular subreddit.	'See/don't forget subreddit link', 'this post doesn't belong here', up-/downvote mentions, acknowledging OP redditors
9. Socializing	The comment follows an informal, small-talk and conversational-like structure between users.	'Thank you', 'much appreciated', gratitude, positive/negative informal conversations, sarcastic one-liners and jokes, personal attacks/criticisms 'you know nothing', 'you are dumb'
Codes 1-7 from Ferguson et al. exploratory dialogue cue phrases (2013); Codes 8-9 added.		

Table 2. Reddit Codebook Version 2

Code	Definition	Linguistic Dialogue Example
1. Critique	The comment suggests disagreement; something may be wrong, faulty or in need of correction/revision/reassessment. Formal/informal negative conversations, personal attacks, criticisms without explanation/discussion.	'However', 'not sure', 'maybe', 'hmm not really', 'what about', 'seems to me', 'actually, not exactly', 'you know nothing', 'you're dumb'
2. Discussion of Resources	The comment references and provides explicit details of additional outside resources (e.g: links to external websites, forums, books, articles) to support understanding or extend discussion.	'Have you read', 'more links', 'check this out', 'look at', 'read this'…BOTH online and offline resources
3. Evaluations	The comment appraises and assesses the merit, worth or significance of something. Formal/informal personal view or positive affirmation/expression of gratitude.	'Likely', 'good point/example', 'agree', 'could be', 'fair enough', 'thank you', 'much appreciated'
4.Explanations	The comment has a descriptive quality and undertakes a process of 'thinking it through' by explaining, brainstorming and justifying a position or idea.	'Meaning/means that', 'our goals', 'aim is', 'it depends, for example', 'that's why', 'another way to look at it', 'through this lens', 'I'd argue', 'same logic would apply'
5. Explicit Reasoning	The comment works out ideas in a logical manner, often reaching a conclusion or proving a point through example based inferences. This includes taking the same line of argument further through questions/objections.	'Next steps', 'relates to', 'then you would', conditional 'if X then Y', 'along these lines', 'maybe/maybe it's because'
6. Information Seeking	The comment asks a specific question, seeks clarification, posts a general inquiry, asks for help on a topic, issue or idea.	'Tell me more about', 'how do you', 'anyone know', 'any advice on how to'
7. Referencing Reddit	The comment references and cites the Reddit platform and may remind users of the protocol/code of conduct for the particular subreddit.	'See/don't forget subreddit link', 'this post doesn't belong here', up-/downvote mentions, acknowledging OP redditors

Table 3. Reddit Codebook Version 3 (FINAL)

Code	Definition	Linguistic Dialogue Example
1. Explanation with Disagreement	Expresses a NEGATIVE take on the content of the previous comment by adding new ideas or facts to discussion thread.	'But', 'I disagree', 'not sure', 'not exactly' with explanation/ judgment/ reasoning/ etc.
2. Explanation with Agreement	Expresses a POSITIVE take on the content of the previous posts by adding new ideas or facts to discussion thread.	'Indeed', 'also', 'I agree', with explanation/ judgment/ reasoning/ etc.
3. Explanation with Neutral Presentation	Expresses a NEUTRAL explanation/judgment/reasoning/etc. with neither negative nor positive reference to the content of the previous comments, nor necessarily any reference to previous comments.	Comments with non-judgmental language. Advice, brainstorming and first hand experiences are framed neutrally. 'I can understand', 'interesting', 'depends on…' or statement responses.
4. Socializing with Negative Intent	Socializing that expresses NEGATIVE affect through tone, words, insults, expletives intended as abusive.	'no', 'you're an idiot', 'this has been explained multiple times'
5. Socializing with Positive Intent	Socializing that expresses POSITIVE affect tone, words, praise, humor, irony intended in a positive way.	'thanks', 'great feedback', 'you're correct'
6. Information Seeking	Comments asking questions or soliciting opinions, resources, etc. ('Does anyone know …?' 'How does this work?'). This does not include questions answered rhetorically within the comment, e.g., if a question is asked and answered.	'First you have to think what happens if …?' and then you can see what happens', 'does anyone know', 'can anyone explain'
7. Providing Resources	Comments that include direct reference to a URL, book, article, etc.; comments that call upon a well-known theory or the name of a well-known figure.	Link to resource copied (book, URL, article, audio/video file). Referencing theory/theorists, scholar or public work (Einstein, Newton, Freud).
8. Subreddit Rules and Norms	Comments on topics such as what is the appropriate sub-reddit for a particular discussion, what language is appropriate to use, how to back up claims by using resources, etc.	'See/don't forget subreddit link', 'this post doesn't belong here', upvote/downvote mentions, acknowledging OP redditors, and bots.

Table 4. Testing Phase Coding Results, Version 3 Schema (2015 data)

	AskScience	**ask_Politics**	**askAcademia**	**askHistorians**
Sample Size	164	190	198	267
1.Explanation with Disagreement	16 (10%)	91 (48%)	21 (11%)	34 (13%)
2.Explanation with Agreement	10 (6%)	11 (6%)	20 (10%)	4 (1%)
3.Explanation with Neutral Presentation	100 (61%)	45 (24%)	102 (52%)	67 (25%)
4.Socializing with Negative Intent	0 (0%)	37 (19%)	5 (3%)	0 (0%)
5.Socializing with Positive Intent	19 (12%)	2 (1%)	44 (22%)	31 (12%)
6.Information Seeking	23 (14%)	22 (12%)	13 (7%)	29 (11%)
7.Providing Resources	33 (20%)	20 (11%)	13 (7%)	64 (24%)
8.Subreddit Rules and Norms	2 (1%)	3 (2%)	6 (3%)	0 (0%)
Krippendorff's alpha (% agreement)	**0.67**	**0.52**	**0.64**	**0.57**

Note: Counts represent an agreement between two or more research coders. Comments where two or more coders did not agree were not counted or included.

Table 5. Subreddit Descriptive Statistics

Subreddit Community	Number of Moderators	Number of Subscribers (at time of data collection)	Number of Posts from 2016	Coded Sample (%1)
AskScience	433	14,000,000	223,000	2,235
ask_Politics	8	26,000	46,000	464
askAcademia	3	32,000	26,900	269
askHistorians	40	600,000	122,000	1,227

Table 6. Coding Results for Independent Coders Phase (2016 data)

	AskScience	**ask_Politics**	**askAcademia**	**askHistorians**
Sample Size	2,235	464	269	1,227
1.Explanation with Disagreement	398 (9%)	164 (18%)	32 (6%)	71 (6%)
2.Explanation with Agreement	323 (7%)	66 (7%)	62 (12%)	45 (4%)
3.Explanation with Neutral Presentation	1890 (43%)	398 (44%)	253 (50%)	592 (48%)
4.Socializing with Negative Intent	43 (1%)	47 (5%)	9 (2%)	4 (0%)
5.Socializing with Positive Intent	360 (8%)	46 (5%)	86 (17%)	204 (17%)
6.Information Seeking	767 (18%)	97 (11%)	50 (10%)	274 (22%)
7.Providing Resources	522 (12%)	78 (9%)	18 (4%)	260 (21%)
8.Subreddit Rules and Norms	49 (1%)	10 (10%)	1 (1%)	66 (5%)
Krippendorff's alpha (% agreement)	**0.69 (78%)**	**0.60 (72%)**	**0.64 (77%)**	**0.76 (79%)**

Note: Counts represent an agreement between two or more independent coders. Percentages may be higher than 100% when coders have assigned multiple (maximum three) codes per comment.

Figure 1: Subreddit Intercoder Agreement Distribution