Interactional Stancetaking in Online Forums

Scott F. Kiesling
University of Pittsburgh
Department of Linguistics
Kiesling@Pitt.edu

Umashanthi Pavalanathan
Georgia Institute of Technology
School of Interactive Computing
umashanthi@gatech.edu

Jim Fitzpatrick
University of Pittsburgh
Department of Linguistics
jim.fitzpatrick@gmail.com

Xiaochuang Han
Georgia Institute of Technology
School of Interactive Computing
xc@gatech.edu

Jacob Eisenstein
Georgia Institute of Technology
School of Interactive Computing
jacobe@gmail.com

Language is shaped by the relationships between the speaker/writer and the audience, the object of discussion, and the talk itself. In turn, language is used to reshape these relationships over the course of an interaction. Computational researchers have succeeded in operationalizing sentiment, formality, and politeness, but each of these constructs captures only some aspects of social and relational meaning. Theories of interactional stancetaking have been put forward as holistic accounts, but until now, these theories have been applied only through detailed qualitative analysis of (portions of) a few individual conversations. In this article, we propose a new computational operationalization of interpersonal stancetaking. We begin with annotations of three linked stance dimensions—affect, investment, and alignment—on 68 conversation threads from the online platform Reddit. Using these annotations, we investigate thread structure and

Submission received: October 15, 2017; revised version received: May 4, 2018; accepted for publication: August 20, 2018.

doi:10.1162/coli_a_00334

linguistic properties of stancetaking in online conversations. We identify lexical features that characterize the extremes along each stancetaking dimension, and show that these stancetaking properties can be predicted with moderate accuracy from bag-of-words features, even with a relatively small labeled training set. These quantitative analyses are supplemented by extensive qualitative analysis, highlighting the compatibility of computational and qualitative methods in synthesizing evidence about the creation of interactional meaning.

1. Introduction

When people interact online, they often do so in the context of a conversational sequence and within the context of a community. Interactants also take stances in conversation—how they claim relationships to their talk, the entities in their talk, and their audience and interlocutors. Such stancetaking is likely affected by both the sequential and community contexts, but as yet there has been little investigation of how stancetaking and these contexts interact to produce linguistic patterns.[1] This article focuses on the intersection of sequencing of stancetaking in conversations on Reddit as they relate to different communities, or subreddits.

Our focus on stancetaking is related to a line of research on the annotation of interpersonal and extra-propositional aspects of language, which encompass topics such as affect, certainty, formality, politeness, and subjectivity. Interpersonal stancetaking represents an attempt to unify many of these threads into a single theoretical framework (Jaffe 2009; Kiesling 2009). The notion of stancetaking, based on the stance triangle approach of Du Bois (2007), captures the speaker's (or writer)'s relationship to (a) the topic of discussion, (b) the interlocutor or audience, and (c) the talk (or writing) itself. Various configurations of these three stance dimensions can account for a range of phenomena. For example, epistemic stance indicates the speaker's certainty about what is being expressed, and affective stance indicates the emotional position of the speaker with respect to the content (Ochs 1993). Until now, the stancetaking framework has been applied only through qualitative analysis of small corpora. In this article, we attempt to operationalize stancetaking in a formal annotation framework, and to use these annotations to analyze social media conversations at scale, bringing insights from qualitative interactional analysis into a quantitative analysis.

Our project therefore has as its main goal the development of a method to annotate stance in conversations of any type. We have used Reddit comment threads as a test bed, and found that stance can be reliably annotated in three dimensions: AFFECT, ALIGNMENT, and INVESTMENT. Building on these annotations, we investigate several properties of stancetaking, taking both qualitative and quantitative perspectives.

First, we investigate patterns of stance "stickiness" through a conversation. We find that some types of stance tend to persist more through a conversation than in others, for both structural and interpersonal reasons. In addition, we find that these "stickiness patterns" are also sensitive to the subreddit in which they occur, which we take to be a proxy for a difference in community norms of stancetaking. Finally, we show that these patterns are connected to different uses of stance markers in conversation (Biber and Finegan 1989; Pavalanathan et al. 2017), enabling classification of stance dimensions from lexical features. These findings demonstrate that the concept of stance and

1 Interactional stancetaking is distinct from **argumentative stances**, a term used to characterize positions in debate (Anand et al. 2011).

stancetaking in conversations is a useful way to explore how interpersonal relationships are created in conversations, and moreover how stancetaking can be computationally extracted from such conversations. To summarize, the article makes the following contributions:

- We introduce interactional stancetaking to the computational linguistics community, and operationalize it through a set of annotation guidelines (available in the Appendix).
- We provide quantitative annotations of three stance dimensions on social media text, demonstrating moderate interrater agreement. These annotations will be made available upon publication.
- We analyze the thread structure properties of stancetaking in online forums, showing how stance utterances tend to pattern in coherent conversational threads.
- We explicate the linguistic features that enact various stances, using both computational and qualitative techniques.

2. Stancetaking, Sequence, and Community

2.1 Brief Theoretical Background on Stancetaking

Stance is used widely in sociolinguistics but not always with specificity, especially the specificity required of computational applications. Stancetaking (generally used as a more verbal synonym for stance) is always related to various kinds of relationships expressed in discourse. Du Bois (2007) argues that at its most basic level, stancetaking is about the evaluation of entities in the discourse by a speaker (or subject, as he refers to them). Alignments and disalignments are created between (canonically two) speakers as they display similarity and difference with respect to these evaluations. So in this "stance triangle," two sides refer to speakers' evaluation of the stance object, and the third side is the relative alignment of the speakers (Figure 1). The advantage of this stance model is that there is a specific basis on which to ground inter-utterance

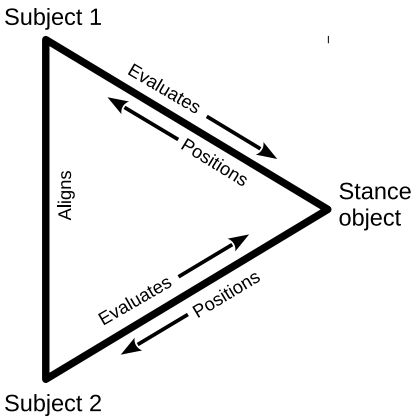

Figure 1
Stance triangle, adapted from Du Bois (2007), p. 163.

alignments, namely, the structure of individual utterances of evaluation. Note that this model is not one based on a single speaker, but is inherently dialogic in the sense that it requires more than one utterance to really know what is going on with respect to stancetaking.

Du Bois (2007) suggests that alignments can be discovered through an analysis of the poetic structure of two evaluations. For example, "I love that game!" might be followed by "I love it too," which has a similar structure and proposition as the first utterance:

(1) I love that game
I love it too

On the other hand, the second response might be "I hate that game," in which the contrast between love and hate shows the disalignment. Not all alignments and disalignments are so straightforward, but the comparison is useful.

(2) I love that game
I **hate** that game

One way that evaluations (and thus alignments) can be modified is by adjusting how much the "animator" (a more generic term for "speaker"; Goffman 1981) is invested in the talk (see also Kockelman 2004). Investment is essentially the strength of an utterance, although theoretically it is the alignment of the actual speaker with the speech uttered. Investment thus includes things like to what extent a speaker is likely to defend the claim subsequently, how epistemically certain they are, and so forth, although investment is likely represented through multiple linguistic cues. In the example, "love" shows high investment, as does "hate." But one can imagine an utterance in which the second utterance simply changed the investment of liking the game; something like "That game is all right." Here, the investment is lowered even though the animator is still technically aligning with the positive evaluation of the game provided by the first speaker.

(3) I love that game
That game is all right

In this example, investment is lowered both through the evaluative predicate adjective "all right" but also through the removal of the speaking subject from the sentence. That is, the latter evaluation frames the utterance as an objective description of the world: "That game is all right" has the same syntax as "The sky is blue." This syntactic change moves the evaluation away from being the responsibility of the animator (more technically, it separates the animator and principal). Investment is therefore an important dimension of stancetaking, and is manipulated by speakers as well as evaluation to do alignment work. Investment is related to, although not the same as, intensity (Zadeh et al. 2016) as it is used in sentiment analysis. Whereas sentiment analysis is usually coded as whether something is strongly positive and strongly negative, investment is a much broader and deeper concept that is rooted in discourse analytic theory. In short, investment is more than the strength of one adjective, but rather a holistic determination of the investment that a speaker has in their utterance. There may be instances when a highly positive affect is also a high investment (as in the "love" example), but because there are instances when they are separable, they should be theoretically considered, and practically annotated, separately.

We thus notice that there are three dimensions to stancetaking: evaluation (which we will refer to as AFFECT), ALIGNMENT, and INVESTMENT. It is important to note that these are *dimensions* of stancetaking, not different "stance types." We hold the view that in any interaction, a stance of some sort is always being taken (or at least bid for), and that these three dimensions are present in all of these unfolding stances; a neutral stance is still a stance. Given this model, we have explored to what extent online conversations can be annotated for the three dimensions (Kiesling et al. 2015).

Based on this discussion, we define **stance** as the discursive creation of a relationship between a language user and some discursive figure, and to other language users in relation to that figure. This discursive figure can be an interlocutor, a figure represented in the discourse, the animator, ideas represented in the discourse, or other texts. We will refer to the discursive figure of this definition as the **stance focus**. This term is based on Du Bois's (2007) notion of stance object, which is the entity that is being evaluated in an utterance; we use *focus* because it is less reifying than *object*. Stancetaking as we are using it, then, is related to argumentative stance, but differentiated from it in the sense that in our view a stance does not need an argument to be taken, and it is understood to arise primarily in multivocal interaction, with writing derivative of speech (a standard assumption in linguistics). Other concepts used in corpus and computational linguistics are also related but different. The notion of affect is used extremely variably, but often to refer to roughly the same thing as our affective dimension. One difference is perhaps that our dimension of affect is only related to claims in the talk/text, and is something created in the interaction, whether or not there is a "reality" in which a person actually feels or has felt that way toward a stance focus. Evaluation and assessment are synonyms for affect in this view. Sentiment is a terminological variant of this work.

A note on epistemicity: Epistemicity is often referred to in discussions, and juxtaposed with affect as a separate stance type. In this view, there are two types of stance, one affective and one epistemic. The affective stance in such approaches (see Lempert [2008], who expands on this division) is generally based on emotion or position with respect to others, and the epistemic is related to the expression of knowledge or certainty. Both of these terms are captured in different parts of the three-dimensional model we are using here. There is not a perfect mapping, but in general affective stance works out to AFFECT and ALIGNMENT, and epistemic stance falls under INVESTMENT, because the expression of certainty aligns the animator and the principal.

At the root of the theory is a model of interaction in which interaction is not something in which interactants frictionlessly represent or convey ideas in their minds, but something that is created collaboratively in interaction (see, e.g., Du Bois and Kärkkäinen [2012]), and stancetaking is contingent on the context and sequence in which any utterance falls. We cannot a priori say that a single linguistic form always indicates a specific stance, but we can find similarities among stances taken with a linguistic form containing it. In other words, stancetaking is a composite of all parts of an utterance in its sequential context, which combine in different ways each time they are used, even as they keep a consistency of flavor across these instances. This presents a challenge for annotation.

2.2 Annotating Stance

In annotating for stance, we focused on annotating the stance focus and a value for each of the three stance dimensions for each "utterance" in a Reddit comment thread. We chose to begin with each utterance because that is what the language producer has done.

Only if we could not find a focus for the entire utterance did we split the utterance (refer to Section 3.1). The instructions we gave to annotators perhaps explain stance focus the best:

> The stance focus is the thing that is made most relevant by an utterance. This can be an entity, for example if someone is talking about football and refers to the Steelers, the Steelers might be the focus, or one of the players. On the other hand, we do other things with language besides assert and evaluate things. We also do things, like ask, insult, compliment, suggest, etc. These can also have stance foci, usually on individuals or the talk itself. So the first step in the analysis is to determine the primary stance focus. One of the best ways to determine the focus is to look at things that are given information in the utterance, such as pronouns or things oriented to but not directly stated. Of course, things may actually be mentioned as well using NPs. If these are the focus, then they are more likely to have a definite article *the* or the proximal deictic *this* (and possibly the distal *that*). These all signal in various ways that the 'thing' is already in the discourse model and the listeners' attention is being focused on it.

To operationalize our definition of AFFECT, INVESTMENT, and ALIGNMENT, we use the following specific instructions for annotation of stance dimensions, each on a scale of 1 to 5, where 1 indicates very low, 3 indicates neutral, and 5 indicates very high. Detailed instructions for annotations are provided in the Appendix.

AFFECT. Affect is the polarity or quality of the stance to the stance focus. For example, if you are talking about food, and you say how yummy the french fries are, then the stance focus is the fries and they are evaluated positively. However, a focus can also be an act. In that case, affect has to do with whether the act itself is overtly positive or negative. So, a request done in an aggravated way ("Shut up!") is negative affect, but in a more mitigated way ("Could you please tone it down a bit?") is more positive. A score of 5 indicates highly positive affect, and 1 indicates highly negative affect. A score of 3 indicates neutral affect (i.e., the absence of a positive or negative affect toward the stance focus).

INVESTMENT. Investment is the dimension of how strongly invested in the talk the speaker is; how committed they signal their relationship to the stance focus. Would they defend their claims and opinions to the death? This dimension is about the talk itself. Again, a score of 5 indicates high investment, 1 indicates minimum investment, and 3 indicates neither high nor low investment.

ALIGNMENT. Alignment is how much a speaker/writer aligns (or not) to their interlocutor(s), real or imagined. Alignment is almost always present at a basic level in that the interlocutor must attend to the same discourse entities in order to hold a basic conversation. That is, speakers orient to the same things in talk. But people do not always align to the objects and figures in talk in the same way. Alignments and disalignments can occur in many ways, and we have to attend to all of them. Alignment will almost always be with respect to the person who just spoke, but alignment can be created prospectively or to a more general audience, especially in the case of Reddit, which is open for anyone on the Internet to read.

Speakers orient to the same things in talk, and speakers orient to each other via established discourse protocols. For example, a refusal to pick up the thread laid out in a first pair part or a changing of subject can be an indicator of an interlocutor's opposition

to a stance utterance already completed, which in turn is a lack of alignment. A score of 5 indicates high alignment of any type, and a score of 1 indicates strong disalignment.

An example of a high score for each dimension can be seen in the part of a thread from *r/Parenting* shown in Table 1. The initial post asks about a way to clean up marks on a wall done by a child. The first comment in 002 suggests using a product called Magic Eraser, and the following two posts support the product. These are high in AFFECT because they all evaluate the stance focus (the Magic Eraser) highly, sometimes in creative ways, such as saying they are "made from ground up fairies" (utterance 002). In terms of INVESTMENT, these are all high because they are so emphatic in their endorsement of the product (again, sometimes in very creative ways). Finally, because everyone agrees that the Magic Marker is magic, the ALIGNMENT is also high.

Low ALIGNMENT can be seen in the part of the thread in Table 2. This post is initiated by a question that basically asks why tipping for average service (rather than exceptional) is the practice in the United States. In 008-02, user A_4 suggests that "you can't fake it like an engineer." This comment sets off an argument about whether engineers can "fake" their job. It is this argument about engineering that garners a low ALIGNMENT score for all the utterances on the topic, as the speakers disagree about the topic, and more personally about how to "take" the criticism of engineers. This "double disagreement" about both the substance and the manner of the conversation leads to the lowest score for ALIGNMENT.

Speech Activity. Although we have not yet used them in an analysis, we also annotated what kind of speech activity the author was engaged in (e.g., lecturing, answering, challenging), and a name for the overall stance (acerbic, friendly, patronizing, etc.). These annotations were free text entries, with the annotators encouraged to use -ing verbs for the speech activity and adjectives for the overall stance. Note that speech activity is not the same thing as speech acts (Searle 1969). Speech acts are generally focused on a single act, whereas speech activity is the ongoing activity that the person can be identified as engaging in (hence the use of progressive aspect -ing morphology) (see also Levinson 1992).

Table 1
Thread with a maintenance of high AFFECT.

ID-rep	Content	Stance focus	User	Karma	Affect	Investment	Alignment
002	Magic Eraser. Those things are made from ground up fairies.	Magic Eraser	A_1	36	4	4	3
003	I wish I could up vote is more than once. I've got a 2 and 4 year old, I drop shipped an entire case of the generics from China last month. Best purchase decision ever.	Magic Eraser	A_2	6	5	5	5
004	Me too. Love that melamine foam.	Magic Eraser	A_3	1	4	4	4

Table 2
Thread example for the ALIGNMENT dimension.

ID-rep	Content	Stance focus	User	Karma	Affect	Investment	Alignment
008-02	People that are truly bad servers don't last very long at restaurants anyway. It is one of the jobs you actually need to be good at, you cant fake it like an engineer.	faking it like engineers	A_4	-2	3	3	3
009	You can't fake engineering	faking engineering	A_5	2	3	4	1
010	Lol... oh yes you can. You have to get the degree but that doesn't mean you are any use to anyone at your job.	faking engineering	A_4	2	3	3	1
011	Firstly, getting the degree is not such an easy task, especially from a top university. You then have to take the fundamentals of engineering exam, then get four years of real engineering experience, then take the difficult professional engineering exam. You can't fake all of that. You also can't be useless to an engineering company by that point.	how to be an engineer	A_5	3	2	4	1
012	god man I was just making a joke, I'm an engineer and I work with a lot of incompetent people. Live a little, don't take life so seriously.	joke	A_4	6	2	4	1
013	judging by barchueetadonai's track record on reddit and in life, he doesn't seem to take jokes well. Especially petty ones.	A_5's track record	A_6	1	2	3	4
014	Just because you don't value the work you have put in doesn't mean you can insult those of us who care about the *hard* work it takes to be an engineer.	on A_4's joke	A_5	-1	2	4	1

2.3 Sequence in Conversation

The context of sequence is a foundational concept in the approach to interactional analysis known as **conversation analysis** (Sacks 1995; Sidnell 2011). An extensive body of work in this field has shown that any current interactional utterance is already "pre-contextualized" to some extent by what has come before, and that a previous utterance can be "recontextualized" to be a different action than first assumed. For example, the statement "It's hot" (referring to air temperature) can count as a different utterance if it is uttered to a stranger at a bus stop and preceded by "Nice day" or "Summer's arrived," or if it is the first thing said. If it is the first thing said, and followed by something

like "Stop trying to flirt, dude," then it has been recontextualized into a particular kind of action and speech act. Conversation analysts have also noted that "first pair parts" structurally "prefer" certain second pair parts, and speakers will mark dispreferred second pair parts with things like pauses and discourse markers like *well*. For example, an invitation structurally "prefers" an acceptance, even if the invitee does not really want to go:

(4) Speaker A: Will you go to the show with me tomorrow night?
Speaker B: (pause) Well, I need to wash my hair so I can't.

In this case, the most direct answer is "no," (and likely the answer that the invitee actually wants to give, given the importance of the excuse), but "no" is structurally dispreferred and thus the form of the rejection is thus *pause + Well + account for not going*. Note that the positive answer would not require such elaboration; a simple "yes" would not be odd (although we might expect a little more investment with the addition of something like "that would be lovely"). So the sequence of utterances in interaction is more than important; it can determine the definition of the action of the utterance.

The sequences we investigate from a corpus perspective are not as detailed and richly contextualized as the qualitative analysis of conversation analysis approach, but patterns of relationship do appear in our data. Because we have annotated different stance dimensions, we extend the understanding of sequencing to try to understand the role of stancetaking in such sequences, and to what extent speakers try to match or contrast stances with others in a conversation, and even a community. We are thus extending a view of context not only to stancetaking, but the sequencing of that stancetaking and the context of specific communities (in this case, the form of subreddits).

2.4 Community and Reddit

Reddit[2] is a Web site that famously calls itself "the front page of the Internet." It was initially designed as a place to crowdsource (or crowdsort) the Internet; users would post interesting and important Web sites with the implication that they would be worth other users' time. It is organized into different topical subreddits, (or sometimes simply reddits). As of this writing there are over 1.1 million subreddits and nearly 250 million users, with new reddits being created at the rate of about 500 per day.[3] Given the huge number of subreddits, they represent a heterogeneous set of topics and create large amounts of texts and conversations.

Also important, however, is that subreddits tend to develop norms of interaction, beyond the topic of the subreddit, that are similar to a community of practice, a term coined by Lave and Wenger (1991) and first applied to language study by Eckert and McConnell-Ginet (1992). A community of practice is defined as a group of people who come together around a particular practice (Eckert and McConnell-Ginet 1992). Eckert (2000) suggests that one aspect of community of practices is indigenous norm creation. That is, that the norms (in this case, norms of interaction) are negotiated by members of the group coming together. In the case of Reddit, each subreddit can be seen as a community of practice coming together around a topic or activity (the reason for posting). Sometimes norms are dictated by moderators, but the moderators tend to be members of the group (they started a subreddit for the purpose stated in the subreddit

2 https://www.reddit.com/.
3 see http://redditmetrics.com/history for current statistics.

"community rules"). However, redditors often subscribe to specific subreddits and visit (and likely comment in) those more habitually, thus creating regular redditors who form a community. It is clear that norms develop in different subreddits, as redditors even discuss which subreddits are "toxic."[4]

The implications of this community building suggest that norms of stancetaking should accrue in these subreddit communities of practice. Indeed, the term "toxic" is a general description of subreddits with habitual and even emphasized disalignment. One of the questions for our project was whether different subreddits can be characterized by different stancetaking norms; in this article we partially address this question by investigating the patterns of stancetaking sequences in different subreddits, finding a slight difference between the subreddit *r/explainlikeimfive* (Explain Like I'm Five – ELI5) and *r/Parenting*.

3. Data Set and Annotation

Reddit posts are at minimum an original post with a link, question, or some other content. It is possible to have a Reddit thread with only such content. However, in most cases other redditors will comment on the post. Redditors can give original posts and comments positive and negative "karma," essentially evaluating any post and comment. The conversations we are focused on are in these comments. Comments are threaded and can develop into conversations. Figure 2 shows the anatomy of a thread as it appears on the Web interface.

Some posts garner very few comments, and others produce long comment threads. We controlled for length by sampling only posts with between 14 and 25 comments, and which were not active at the time when we gathered the data. These thresholds were chosen with several parameters in mind. The lower bound was motivated by the need for enough utterances to make meaningful comparisons of thread structure properties; furthermore, very short threads tended to be qualitatively different from longer ones, and so were left out for now. As for the upper bound, in the subreddits that we considered, there are relatively few threads with more than 25 comments. As these would necessarily take longer to annotate, we omitted them as well.

In the first training phase, we selected posts from the subreddits *r/Fitness*, *r/Parenting*, *r/Metal* (music), *r/Pittsburgh*, and *r/Atlanta*. We thought these might give us a range of interaction and stance types with which to test the annotation scheme. Once we determined that the annotation scheme was reliable enough, we decided to focus on *r/Parenting* and added the subreddit *r/explainlikeimfive* (ELI5). In this last subreddit, posters ask commenters to explain a complex topic "like I'm five" years old, adding another kind of explanatory stancetaking that we felt was likely to contrast with *r/Parenting*. Our subjective impression of the membership of these groups also suggested that *r/Parenting* might serve to balance the gender of the author pool, with *r/explainlikeimfive* contributors being mostly men and *r/Parenting* contributors being more women. Statistics about the data set are shown in Table 3.

Reddit Inc. claims that there were more than 100,000 active subreddits in August 2018.[5] Through manual annotation, it is possible to touch only a tiny fraction of this set of communities, and Reddit in turn is only one of several popular social media

4 https://www.reddit.com/r/AskReddit/comments/2v39v2/what_popular_subreddit_has_a_really_toxic/.

5 https://www.redditinc.com/, retrieved 20 August 2018.

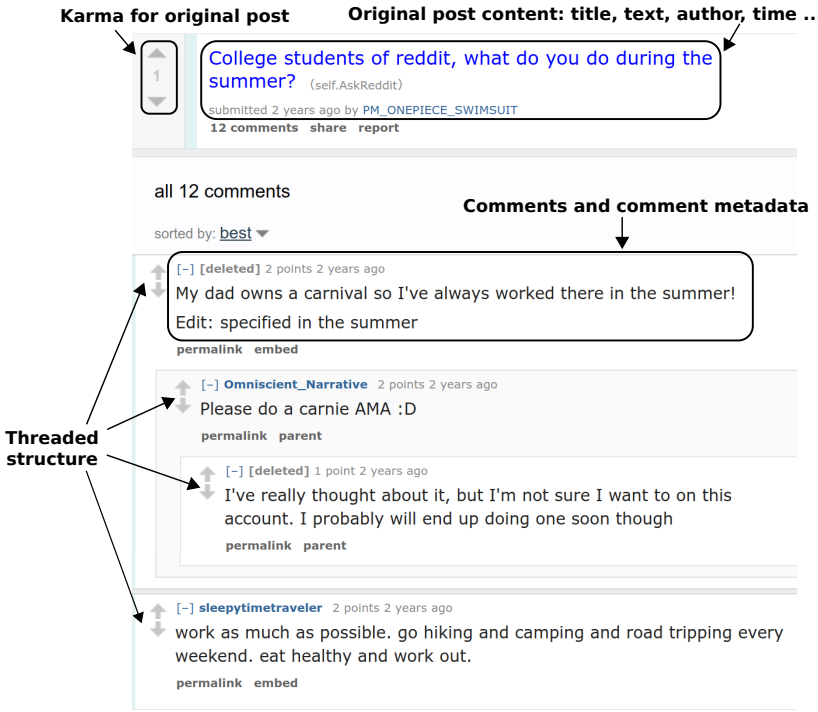

Figure 2
Anatomy of a Reddit thread.

platforms. Furthermore, bias may be introduced by our focus on subreddits whose interactional styles and stances were familiar, or at least comprehensible, to us. It is not possible to generalize from the small subset of communities considered here to the full range of interactive possibilities in online social media, or even to characterize the "typical" case. Nonetheless, this data set provides a starting point for the formal analysis of stancetaking, enabling us to quantify the extent to which annotators agree on stance in at least some cases, and exploring the possibility of linking stance to the structural analysis of online dialogues. Further work is necessary to determine the extent to which our findings generalize to other online communities.

3.1 Preprocessing: Segmentation

In order to discuss stance, we have as a central concept the **stance focus**. One of the problems we initially faced during the first training phase of annotation was that some comments contain more than one utterance or action with different stance foci.

Table 3
Data set statistics.

Total no. of threads	68
Total no. of utterances	1,265
Total no. of authors	616
Total no. of tokens	66,347

Therefore, in our preprocessing, we needed to segment the comments so that each annotation was associated with only one stance focus utterance. Once the sampled posts were identified, they were downloaded and reformatted so that the original post and comments could be data rows. We then inspected each post and, if necessary, resegmented comments if they contained comments with differing stance foci.

In the first of the following two examples, there is a single stance focus over the entire comment, so no segmentation is necessary. In the second example, the stance focus shifts, requiring segmentation:

- A laser powerful enough to burn paper, by itself, costs upwards of $150, and that doesn't include the rest of the printer mechanics. If it became a popular enough technology, that price could come down, though. Probably not enough to make it financially viable. *(Focus: the viability of making a thermal laser printer.)*
- It's not necessary. I never refrigerate mustard and it doesn't spoil. *(Focus: the idea of refrigerating mustard.)* However, it makes sense to keep mustard in the fridge for these reasons: it tastes better cold, it's a convenient place to keep it (near other sandwich ingredients), exercising an abundance of caution. *(Focus: hypothetical reasons to refrigerate mustard.)*

Segmentation was performed by one annotator and then checked by a second. If there was disagreement, a third annotator resolved each disagreement. We thus had a corpus of utterances, one for each stance focus. Stance foci are not always a locally mentioned object, but may rely on previous utterances. For example, in the first example, it may seem like the laser is the stance focus, but in fact because the discussion is about *making* thermal laser printers, and this utterance is mainly relevant to that focus, so that is the focus rather than the *cost* of a laser.

3.2 Annotation Process and Interrater Agreement

The annotation was conducted in three main phases:

- In the first phase, annotations were performed by four of the authors of this paper. Disagreements were discussed. These discussions helped to refine the annotation instructions, but conflicting annotations were not changed after discussion. Because it seemed that each thread introduced new challenges and causes for disagreement, we hypothesized a speech activity effect. For example, in the *r/Parenting* and local city subreddits, people tended to ask questions or give advice and help, whereas in the *r/Metal* (music) and *r/Fitness* subreddits, there was more argument and sarcasm. These subreddits also featured "inside knowledge," such as jokes and jargon, making annotation difficult.
- In the second phase, annotations were performed by the first author and a team of undergraduate research assistants. Because of the difficulties encountered in the first phase, we focused on two subreddits: *r/explainlikeimfive* (ELI5) and *r/Parenting*, where the speech activity is more focused.
- In the third phase, the student research assistants independently annotated 35 more threads, again focusing on *r/explainlikelimfive* and *r/Parenting*.

Interrater agreement was computed for the 33 threads annotated in phases 1 and 2, using Krippendorff's α (Hayes and Krippendorff 2007). To compute the chance

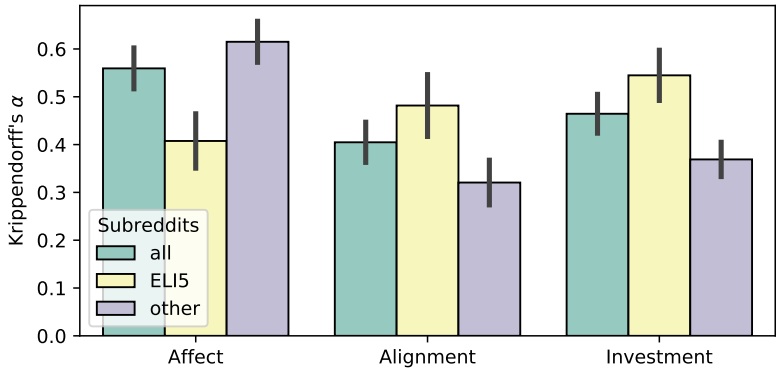

Figure 3
Interrater agreement, measured by Krippendorff's α. The error bars are standard deviations of 100 bootstrap samples.

level of agreement, the annotations were randomly permuted over stance utterances in the thread, thus preserving the distribution of ratings within the thread. Following Krippendorff (2007), we used the squared difference in ratings as the distance metric. As shown in Figure 3, moderate agreement was attained for all three stance dimensions, with a maximum of $\alpha = 0.57$ for AFFECT, and a minimum of $\alpha = 0.40$ for ALIGNMENT.

The level of agreement that we obtain for affect is nearly identical to the agreement reported by Thelwall et al. (2010) on the task of annotating sentiment on tweets. In both cases, the texts to annotate are short, increasing the sensitivity to specific linguistic features, which may be interpreted differently by annotators. Furthermore, as noted by Craggs and Wood (2004, page 97) Krippendorf's α does not have a cut-off score, but rather "should dictate the applications to which the resulting annotated data can be applied." Given that AFFECT is more dependent on decontextualized cues such as word meaning, and ALIGNMENT is the most dependent on sequential context, it makes sense that Krippendorff's α would be higher for the former than the latter.

There were substantial differences between the posts from *r/explainlikelimfive* (ELI5) and the other subreddits: for *r/explainlikelimfive*, agreement was lowest for AFFECT ($\alpha = 0.41$), reflecting the relatively muted affective stances taken in this community; agreement for INVESTMENT was $\alpha = 0.55$, reflecting the frequency of vehement disagreements. For the remaining subreddits, the agreement on AFFECT was $\alpha = 0.63$, with $\alpha < 0.4$ for INVESTMENT and ALIGNMENT. We also compared the agreements for various groups of annotators — linguists versus computer scientists, undergraduates versus more experienced researchers, and advisor–advisee pairs—but found no significant differences.

Over all three phases, we have annotated 68 Reddit threads.[6] The breakdown of various subreddits is shown in Table 4. The distribution of averaged annotation scores for each stance dimension is shown in Table 5.

6 Two additional threads were annotated at an early stage, with each annotator providing their own segmentation. To avoid the difficulty of aggregating across segmentations, we omit these from the data set.

Table 4
Annotated thread counts.

Subreddit	Multi-annotated threads	Single-annotated threads
r/explainlikeimfive	21	18
r/Parenting	8	17
r/Metal	2	
r/Atlanta	1	
r/Fitness	1	
Total	33	35

Table 5
Distribution of stance annotation scores.

		Stance Score Distribution %				
Stance Dimension	**Utterances**	**1 (low)**	**2**	**3 (neutral)**	**4**	**5 (high)**
AFFECT	1,258	1.03	18.68	67.97	05.56	06.76
INVESTMENT	1,252	0.08	07.91	48.80	18.05	25.16
ALIGNMENT	1,235	0.97	15.79	56.92	11.82	14.49

3.3 Research Questions

Using the corpora of Reddit threads annotated for three different stancetaking dimensions, we focus on the following research questions:

Thread Structure Analysis:

RQ1a: To what extent are different stance dimensions influenced by previous stances in the same thread?

RQ1b: How do these thread structure properties vary by subreddit?

RQ1c: How do these thread structure properties vary by stance dimension and level (e.g., high/low)?

RQ1d: What qualitative evidence suggests reasons for the patterns?

Keyword Analysis:

RQ2a: What are the keywords that are indicative of each of the stance dimensions?

RQ2b: How predictive are these keywords of stance annotations?

We study these research questions using the corpus of Reddit threads annotated for the three stance dimensions: AFFECT, INVESTMENT, and ALIGNMENT. For the purpose of subsequent analyses, we converted the annotation scores to three classes: high (scores 4 and 5), low (scores 1 and 2), and neutral (score 3). This is because of the sparseness of the very low (i.e., score of 1) and very high (i.e., score of 5) scores in the annotated data set (Table 5) and better interpretability when using three levels to describe the stance dimensions. When there are multiple annotations available for an utterance, we converted to the three-class scale as follows: an average score of 3.5 or

greater is high, an average score of 2.5 or less is low, otherwise neutral.[7] The class distribution for each stancetaking dimension is shown later in Table 9.

4. RQ1: Thread Structure Analysis

Our first research question focuses on the patterns of stancetaking in Reddit conversations and we use both quantitative and qualitative analyses to understand these patterns.

4.1 Quantitative Analysis

To measure the extent to which the current utterance's stance dimensions are influenced by the stance dimensions of the previous utterance (i.e., how "sticky" different stance dimensions are throughout a conversation), we compared the observed counts of adjacent utterances (P_O) with the same class of stance dimensions to the expectation (P_E) under a random reassignment of labels. For the random reassignment probability, we randomly shuffled the class labels for each utterance within each Reddit conversational thread 10,000 times and computed the proportion of adjacent utterances with the same class labels. This corresponds to a null model in which the overall distribution of stances is preserved, but there is no relationship between adjacent utterances.

These probabilities are shown in Table 6. A positive sign for the difference $P_O - P_E$ indicates that the observed probability of same-class stance dimensions in an utterance pair is higher than we would expect under the null model, indicating that the stance is "sticky." Statistical significance is computed using the empirical p-values across the 10,000 random reassignments. To correct for multiple comparisons, we use the Benjamini and Hochberg (1995) procedure to bound the overall false discovery rate at $p < 0.05$.

4.1.1 RQ1a: To what extent are different stance dimensions influenced by previous stances in the same thread? Results from the stickiness analysis when considering all the threads in our data set are shown in Table 6(a).[8] The results for ALIGNMENT suggest that ALIGNMENT is sticky throughout a conversation. For INVESTMENT, the situation is reversed, with adjacent utterances sharing the same INVESTMENT with lower probability than expected under the null model. We provide further explanations with a qualitative analysis in Section 4.2.

4.1.2 RQ1b: How do these thread structure properties vary by subreddit? We next address the differences in stancetaking stickiness across different subreddits, focusing on conversational threads from *r/explainlikeimfive* and *r/Parenting*. *r/explainlikeimfive* is a forum for people to request help understanding complex concepts and share friendly, simplified, and layman-accessible objective explanations.[9] *r/Parenting* is a community

7 Note that this method is more symmetric than rounding to an integer.

8 Note that the total number of utterance pairs slightly varies for each dimension because we removed missing or invalid annotations for each utterance at the stance dimension level. Further, the top level post of a thread may not be annotated for ALIGNMENT when it is unclear if the post aligns with the subreddit, and therefore the total number of ALIGNMENT utterance pairs are lower than that of the other two dimensions.

9 `https://www.reddit.com/r/explainlikeimfive/`.

Table 6
Results: Stance "stickiness." The asterisk* indicates statistical significance at $p < 0.05$, after Benjamini and Hochberg (1995) adjustment for false discovery rate.

	Same-class probability				
Stance dimension	**utterance**	**observed (P_O)**	**expected (P_E)**	$P_O - P_E$	
(a) All threads					
AFFECT	1,056	0.598	0.563	0.035	
INVESTMENT	1,046	0.459	0.523	−0.064	*
ALIGNMENT	947	0.548	0.493	0.055	*
(b) Only r/explainlikeimfive threads					
AFFECT	598	0.598	0.590	0.008	
INVESTMENT	588	0.463	0.524	−0.061	*
ALIGNMENT	542	0.542	0.477	0.065	*
(c) Only r/Parenting threads					
AFFECT	404	0.609	0.537	0.072	*
INVESTMENT	404	0.465	0.542	−0.077	*
ALIGNMENT	351	0.573	0.522	0.051	

dedicated to discussions related to parenting.[10] As shown in Table 6(b) and 6(c), ALIGNMENT is sticky in conversational utterances in *r/explainlikeimfive*; the results for *r/Parenting* are similar, but slightly below the level of statistical significance. INVESTMENT is "anti-sticky" for both subreddits—similar to the observations when considering all of the threads. A difference arises with respect to AFFECT, which is sticky in conversational threads in *r/Parenting*, but not in *r/explainlikeimfive*. These differences suggest that the two sub-communities have different conversational norms. In *r/explainlikeimfive*, there seems to be an emphasis on "objective" comments, and neutral affect is considerably more frequent in this subreddit than in *r/Parenting*.

4.1.3 RQ1c: How do these thread structure properties vary by stance dimension and level (e.g., high/low)? Next, we look at the patterns of transitions between different levels of stancetaking for AFFECT, INVESTMENT, and ALIGNMENT. First, we count the number of utterance pairs that transition from one level to another or remain at the same level (high (H) → low (L), high (H) → neutral (N), etc). To compare with the chance counts under the null hypothesis, we again perform 10,000 random reassignments of class levels within each thread, and compute empirical p-values under this sampling distribution. Transition counts and comparison with chance counts are shown in Figure 4, where a positive superscript indicates observed transition counts being significantly greater than chance counts and a negative superscript indicates the opposite.

As shown in Figure 4, the L → L transition is significantly higher than random chance for ALIGNMENT. L → L transitions for ALIGNMENT contribute to the stickiness observation in RQ1a. Further, the significantly higher number of N → H transitions indicate the possibility of moving upward from neutral ALIGNMENT to high ALIGNMENT (e.g., question/answer pairs where the answer utterances are annotated as high ALIGNMENT), but these are counterbalanced by significantly fewer transitions from

10 https://www.reddit.com/r/Parenting/.

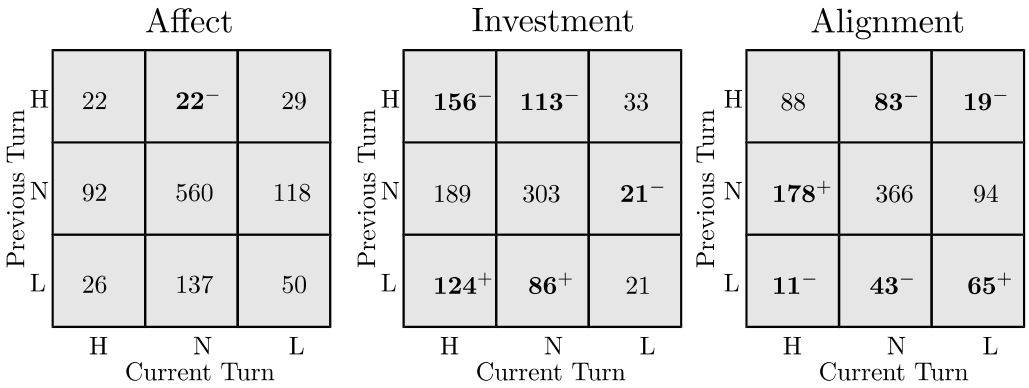

Figure 4
Transitions on stance dimensions for adjacent utterances. Observed transition counts that are significantly greater than chance ($p < 0.05$) are shown in **boldface** with a "+" superscript; observed counts which are significantly less than chance counts are shown in **boldface** with a "−" superscript. All p-values are adjusted for multiple comparisons using the Benjamini and Hochberg (1995) false discovery rate.

H → N ALIGNMENT. For INVESTMENT, the majority of the anti-stickiness observations are due to the H → H transitions. Many of the upward transitions (L → H and L → N) for INVESTMENT are again due to the question/answer utterance pairs, with questions often annotated as low investment.

4.2 Qualitative Analysis

What might the motivation be for the stickiness of ALIGNMENT and the anti-stickiness of INVESTMENT? In order to explore this, we perform qualitative analysis on selected segments that represent the significant patterns found. Recall that annotators were simply instructed to annotate based on how well they thought the speakers were aligning in a naive sense. That is, they were not instructed to annotate for ALIGNMENT based on a match of the evaluation (as for the Du Bois [2007] stance model), although that might have been something that speakers keyed into. There are thus a number of discourse structures that annotators could be keying on when they are evaluating ALIGNMENT (based in part on the discourse model of Schiffrin [1988], page 25):

Evaluation alignment A basic form of alignment is agreement on the evaluation of a stance focus, as outlined above in Du Bois's stance triangle model. To return to the example there, this kind of alignment is shown when both speakers evaluate the game highly ("I love that game! I do too!").

Propositional alignment A more general type of Evaluation alignment is a general agreement about the propositional content. For example, one person might remark that the sky is turning gray and the other speaker simply agrees. In such an exchange there is no overt evaluation of the color of the sky; it is simply gray and the speakers agree on that fact.

Action alignment Another way of viewing this kind of alignment is "cooperativeness," in the sense of participating in the activity faithfully. This means aligning with second pair parts to first (for example, answers for questions) or at least orienting to problematic or dispreferred second pair parts. For example, providing an answer

to a question that does not actually answer the questions is a disaligning utterance in the action structure. See the rejection of the invitation in Example (4).

Exchange alignment This type of alignment is not relevant for Reddit but included for completeness. Communication by speaking or signing, rather than through text, usually also exhibits a subtle utterance exchange system such that more than one speaker rarely speaks at a time (see Sacks, Schegloff, and Jefferson 1974; Sidnell 2011). When there is a disalignment on this discourse structure, we see interruptions or awkward pauses.

In the quantitative analysis, we have discovered several stickiness patterns and proposed a number of possible explanations for them. In what follows we provide some qualitative analyses to explore whether these explanations hold up in a more context-rich analysis. For example, one pattern that does occur is a shift from low to high in INVESTMENT, especially in the *r/explainlikeimfive* subreddit. A short inspection of the actual threads in which this happened revealed that these patterns arise from factual question–answer pairs: Questions of fact are low INVESTMENT because a question is by definition low INVESTMENT (the speaker is professing to not know the possible world in which something is true) and an answer, especially in *r/explainlikeimfive*, is high INVESTMENT. So this pattern was easily explained through qualitative inspection of some examples.

Once we discovered significant patterns of interest in the quantitative analysis, we looked for instances of the pattern in the annotated data, and pulled out the relevant sections of those threads. They were then inspected to understand more fully what stance utterances the commenters were making in order to have been annotated as they were. We turn to these in the next section.

4.2.1 RQ1d: What qualitative evidence suggests reasons for the thread structure patterns? Alignment is usually a backward-looking dimension: An utterance is always evaluated in the context of the previous one, unless we are beginning a conversation (and then in terms of what is expected of conversational openings; see Schegloff and Sacks [1973]). In light of this fact, it makes sense that ALIGNMENT would be relatively sticky: There is a built-in conversational connection between a previous ALIGNMENT and a new one, and if a Reddit post can be thought of as a small community, previous alignments produce further alignments, and previous disalignments do the same. This observation can be further motivated by Communication Accommodation Theory (Giles, Coupland, and Coupland 1991).

Du Bois (2007) suggests that alignment happens only in the case of evaluation, which predicts that the stickiness of AFFECT should be similar to the stickiness of ALIGNMENT. For example, suppose comment A has a low ALIGNMENT with the previous comment(er), P. The subsequent comment B will either align with A, or not. But comment B is not related to the A–P relationship, only the B–A relationship. Moreover, this ALIGNMENT must have a previous utterance with which to align or disalign. On the other hand, AFFECT is a comment-internal relationship. Recall that in our terms, AFFECT is defined as evaluation of a stance focus (this terminology may differ from other uses in computational linguistics). AFFECT is achieved utterance-internally, so it is more separable from previous utterances, at least in terms of its definition. The same is true for INVESTMENT.

Two important points arise from these observations. First, it is clear that annotators are keying on something more than evaluation alignment, because AFFECT stickiness is not correlated with ALIGNMENT stickiness. Herein we explore some of the ways

Table 7
Thread with a maintenance of high ALIGNMENT.

ID-rep	Content	Stance focus	User	Karma	Affect	Investment	Alignment
001	I know it's little late but everybody seems to know bits and pieces and I'd like a solid explanation.	Syrian conflict	A_7	3	3	2	4
010	this explains things very well, its a youtube video: <URL>	video explaining Syrian conflict	A_8	2	5	4	4
011	Really, really good video, thanks.	video explaining Syrian conflict	A_9	1	5	5	5

that ALIGNMENT stickiness works in our data. Second, the fact that ALIGNMENT would be more sticky is not surprising given that the definition of ALIGNMENT inherently looks to other utterances.

For an example from our data, consider the high alignments maintained in the thread in Table 7. This table is a view of the comment thread in the annotation environment. Utterance 010 is a response to utterance 001, which is the original post—in this case a request for information about the civil war in Syria. Utterances 002 through 009 were a separate comment thread not related to utterance 010. The rightmost column shows the levels for ALIGNMENT, all of which are high. Although the original post is usually not annotated for ALIGNMENT, in this case the author is aligning with the audience by taking the audience's perspective with "I know it's a little late," which anticipates objections that those contemplating answering might have. In a sense, the speaker is aligning with an imagined evaluation of their query. Note the low INVESTMENT because it is a query.

The response orients to the question, even though not directly answered, by recommending the link and explaining the target of the link. It is not uncommon for those answering such queries to simply post a link, so by actually explaining the content of the link target, the comment aligns with the request and "counts" as a valid second pair part to the first pair part in the original query (in this case, an explanation posted to a request for explanation). Finally, a third user aligns with the utterance in 010 by explicitly evaluating the action as a good example of an explanation, and providing thanks to the author A_2. In short, each contributor is cooperative in terms of the expected action in each slot, and overtly marks this cooperativeness. A more minimal and less overtly aligning version might read:

(5) Comment A: Please explain the Syrian civil war
Comment B: <link>

In this imagined exchange, even though there is a baseline of cooperativeness (there is a request for explanation and a link to an explanation), the speakers do not go out of their way to explain how this alignment is happening. So, in general, high levels of ALIGNMENT are those that are signaling this cooperation and agreement more overtly than not.

Why does the ALIGNMENT persist? Cooperativeness is a fundamental feature of human interaction, as pointed out by Grice et al. (1975). Thus, opposition tends to lead to more opposition from others, and alignment leads to more alignment. If that is the case, why then does ALIGNMENT persist more than other dimensions? ALIGNMENT is, again, the only dimension that inherently links with other utterances or an imagined audience, or both.

Consider an example of a shift upward in ALIGNMENT, shown in Table 8. This thread arises from a question about whether destroying a nuclear missile with conventional explosives will cause a nuclear explosion (it will not). The first comment shown is the ninth in the thread. It actually quotes the comment it is responding to ("Normal

Table 8
Thread with a shift from low to high ALIGNMENT.

ID-rep	Content	Stance focus	User	Karma	Affect	Investment	Alignment
009-1	"Normal explosives do not generate nearly enough energy to trigger fusion/fission." They can, in a carefully controlled and designed way.	Normal explosive	A_{10}	1	3	5	2
009-2	Nuclear weapons are usually triggered by a careful detonation of convention explosives in such a way as to compress the fissile (nuclear) material (URL). Compressing the material lowers the critical mass (mass necessary for a self-sustaining nuclear reaction), which allows the previously inert nuclear warhead to suddenly undergo nuclear fission and detonate. Obviously, the random/uncoordinated way that an intercepting missile would interact with the nuclear mass pretty much prevents it from the necessary compression for the nuclear weapon to detonate.	nuclear bombs	A_{10}	1	3	5	3
010	Interesting. I didn't really think about how they would detonate it if conventional explosives didn't work. But since it's required that the radioactive material be compressed for it to work, would it be fair to say that radioactive material under normal conditions couldn't be detonated by conventional explosives?	radioactive material	A_{11}	2	4	2	4

explosives do not generate nearly enough energy to trigger fusion/fission. They can, in a carefully controlled and designed way."). The second segment is still part of the same author's post, but the stance focus shifts from what a normal explosive can do (and the post it is responding to) to how nuclear weapons work more generally. Finally, a new author responds and then asks a question. The shift in ALIGNMENT goes from low (2) to high (4) over this stretch; because the first two segments are from the same comment, in the quantitative analysis these three segments would count as two pairs going from 2 to 4 and 3 to 4.

The "backwards-looking" property of ALIGNMENT is shown here, as is the accumulation of stance toward ALIGNMENT. In utterance 009-1, the author uses quotes to focus their response on a single portion of the previous comment. The disalignment comes from the correction of the previous claim (a propositional disalignment). The second part of the post (009-2) is simply some explanation of the mechanism of nuclear weapons, and the neutral score it gets for ALIGNMENT shows that it is neutrally aligning. That is, it is neither disaligning nor overtly marking some alignment.

In the next utterance (010), the author A_{11} offers the single comment "interesting," accepting the propositional content of utterance 009, and aligning propositionally. The next sentence aligns interactionally, by asking a question that accepts and builds on the facts offered in the previous utterance. This view suggests that there is a way ALIGNMENT can focus on larger norms of the speech event; in this example alignment does not rely on the previous alignment at all, and in fact is cooperative in the subreddit *r/explainlikeimfive*, since the entire point of that subreddit is to have things explained, and A_{11} is asking for a further explanation.

A look at this post suggests another reason why ALIGNMENT might persist. First, most of the ALIGNMENT persistence is on the neutral level. In this example we saw that neutral is really a lack of disalignment and lack of overt alignment; that is, it is a mundane and unremarkable conversational alignment that must exist for a coherent conversation to take place at all. In addition, even after disalignment there seems to be interactional pressure to revert to neutral ALIGNMENT.

This example thus shows some of the dynamics of how an upward shift in ALIGNMENT occurs (a significant shift in the quantitative data) and also how that might revert quickly back to neutral ALIGNMENT. It might be fruitful to contrast this with AFFECT, which seems to be more short-lived, most likely because it is a more utterance-internal dimension. In the nuclear weapons example, AFFECT is neutral during the explanations in utterances 009-1 and 009-2. The positive adjective "interesting" suggests a positive AFFECT, in that interesting is a positive attribute of an explanation. But that positive AFFECT does not persist, because the explanation that follows (not shown) is neutral for ALIGNMENT. Finally, the INVESTMENT pattern mentioned is clear in the *r/explainlikeimfive* nuclear weapons example, with explanations rated a score of 5 for INVESTMENT and the question rated a score of 2.

A further example is shown in Table 1, from *r/Parenting*. This post is started with a query about how to get rid of marks from toddlers on a wall. In utterance 002, the commenter A_1 suggests the Magic Eraser, and evaluates it highly by suggesting that it was created by magical beings (fairies), which is both a high AFFECT and high INVESTMENT. The interesting thing about this thread is the extreme INVESTMENT. Both A_1 and A_2 put in significant effort to come up with new (and very creative) superlatives for the positivity of their evaluations of the Magic Eraser. Although this sequence appears as sticky, and it does create evaluation ALIGNMENT among the speakers, each evaluation is clearly a separate contribution that the commenter would have made on their own. This is mainly an effect of the request in the post and how it was made;

that is, a request for a recommendation. Such a request has as its preferred second pair part a positive evaluation of something in the form of recommendation. So, these multiple high INVESTMENT and high AFFECT utterances are expected. This observation suggests that one reason for *r/Parenting* to have more of an effect for AFFECT than in other subreddits is if a higher proportion of initial posts are these kinds of requests for recommendations (either products or solutions to parenting problems). Although some of these might lead to disagreements (such as, for example, a query about sleep training), some, and possible more, are likely to lead to the kind of pattern seen in the Erasing example, in which both AFFECT and ALIGNMENT are successively ramped up.

Indeed, the nature of the individual post and its purpose may be the strongest factor separating out the subreddits. In *r/explainlikeimfive*, most of the initial posts are likely to be queries that have as the preferred next action a fairly certain explanation that is helpful. This means that there is a bias toward ALIGNMENT in the helpfulness of the explanations, and the practice of often thanking commenters for their explanation. For both kinds of initial posts in these subreddits—that is, both requests for explanations and requests for recommendations—the following actions are likely to lead to a stickiness of ALIGNMENT. For *r/explainlikeimfive*, there is less likely to be a high AFFECT, or INVESTMENT, given the kinds of dispassionate objective-sounding explanations that are preferred. In *r/Parenting*, however, there is more pressure for high AFFECT, given requests for recommendations. Another positive aspect of this analysis is that it lends itself to the statistical pattern; that is, it is not a totalizing explanation but one that suggests biases of the subreddits that lead to the stancetaking patterns we have found. A further analysis will investigate the patterning of the initial actions of each Reddit post for a correlation with the type of pattern that follows, and also whether certain initial actions are correlated with certain subreddits.

As we discuss in Section 5, this analysis also aligns with some of the keywords that populate the different stance dimensions. For example, the alignment dimension is evoked by *thank* and *thanks*, which are usually uttered after another action that has been fulfilled. So, these are, like ALIGNMENT, inherently backward-looking and create a positive relationship to the person being thanked. Note that *thank* is also the top word in AFFECT, which could be related to the fact that the action of thanking presupposes that the person thanking evaluates the action they are thanking positively.

A qualitative view of some of the threads that express the significant patterns in the quantitative analysis thus allows us to pull out interactional explanations for these patterns. The main finding is that much of the stickiness (or anti-stickiness) or the dimensions seems to rest on the type of action in the initial post and the preferred second pair parts to those actions, combined with the action goals and norms of a subreddit (for example, explanation for *r/explainlikeimfive* and advice for *r/Parenting*). Such preferred second actions are then biased in terms of the stance dimensions.

5. RQ2: Keyword Analysis

RQ2a: What are the keywords that are indicative of each of the stance dimensions?

RQ2b: How predictive are these keywords of stance annotations?

Next, we built a classifier to predict high, neutral, and low levels for each stance dimension. This enabled us to measure the extent to which stance dimensions can be

Table 9
Results: Stance dimension classifiers.

	Class Distribution %			Baseline	Model	
Stance Dimension	**H**	**N**	**L**	**Macro F1**	**Macro F1**	**ROC Area**
AFFECT	12.54	19.18	68.28	15.86	39.01	68.54
INVESTMENT	42.97	48.74	08.29	25.63	45.78	71.42
ALIGNMENT	28.08	56.61	15.31	21.69	34.92	66.80

predicted from lexical features, and, more importantly, to identify the keywords most associated with strong stances.

5.1 Classification Task

For each stance dimension, we built a three-class logistic regression classifier to label each utterance in a Reddit thread along three levels—high, neutral, low.[11] The feature set includes unigrams that appear in at least three conversational utterances, special tokens for quotes and URLs, and utterance length in percentiles.

We randomly split the data set of annotated conversational utterances into 70% training set, 10% development set, and 20% test set. We used the development set to tune the parameters of the logistic regression. Before using the annotated corpora for model building, we performed several preprocessing steps. Because of the non-standard nature of the language in Reddit conversations, we used NTLK's TweetTokenizer[12] to tokenize the text and downcased the tokens. This tokenization step preserves punctuation such as '?' and '!'. Some of the Reddit comments contain quotes from previous comments, which we replaced with a special token. We also replaced URLs with a special token.

The validity of the keyword analysis depends on the classifier achieving at least moderate predictive power. Because the classes are unbalanced, we evaluated each classifier using macro-F1 and average area under the ROC curve. In both cases, we average between the high and low classes. Table 9 compares the classifier against a random baseline. By construction, a random baseline will achieve 0.5 area under the ROC curve. These results show that keyword-based classification can attain moderate predictive power for these stance dimensions, even with relatively little training data. We therefore move to keyword analysis, using the most strongly weighted features for each stance dimension's classifier.

5.2 Top Keywords

Table 10 shows the top textual features that are predictive of high and low levels of AFFECT, INVESTMENT, and ALIGNMENT. In general, the keywords accord with our

11 The classifiers were implemented using scikit-learn (Pedregosa et al. 2011), with the option of `multinomial`, `saga` as the solver (Defazio, Bach, and Lacoste-Julien 2014), and maximum iterations of 3,000.

12 `http://www.nltk.org/api/nltk.tokenize.html`.

Table 10
Results: Top predictive terms for each stance dimensions.

AFFECT		INVESTMENT		ALIGNMENT	
HIGH	LOW	HIGH	LOW	HIGH	LOW
thank	please	!	little	thank	evidence
!	worse	tell	limit	limit	wrong
sing	everyone	hope	ink	other	able
noise	nothing	better	maybe	!	not
stop	entire	never	may	absolutely	opinion
friends	into	stick	wouldn't	thanks	worse
good	burn	parents	everyone	now	mom
fiber	no	kept	know	so	be
kindle	password	.	wants	point	has
love	effectively	carefully	actual	some	well

intuition about these stance dimensions. For example, "thank" is the top word in both AFFECT and ALIGNMENT. Thanking is normally an aligning utterance, and is usually done because another person has done something good for the person doing the thanking; in other words, thanking is assumed to be positively evaluated and hence high on AFFECT. Exclamation points are used to mark statements of high intensity, so seeing them at the top of INVESTMENT makes sense. The fact that they also appear high in AFFECT and ALIGNMENT suggests that such high INVESTMENT utterances are usually made to increase those dimensions as well (and maybe also in specific speech acts such as compliments).

Many of the words on the low side are understandable as well. For example, "nothing," "no," and "worse" are clearly relatively negative words. The appearance of "please" in the low for AFFECT seems a little surprising, but it may be that it is used frequently to mitigate the effects of a negative evaluation ("please don't take this the wrong way, but...") or a face-threatening command ("please read the guidelines before posting"). The words appearing in the low value for INVESTMENT are logical fits that seem to be epistemic mitigators of different types. "May" and "wouldn't" are modal auxiliaries, "maybe" is a clear hedge, and "little" and "limit" can be used to minimize as well. "Everyone" can lower the INVESTMENT by suggesting that whatever the statement is is unremarkable because "everyone believes (or does) it." (See Kiesling [2018] for an analysis of a conversation that includes this use of "everyone"). There are negative and likely disaligning words in the low value for ALIGNMENT as well: "wrong," "not," and "worse." Moreover, one can easily imagine contexts in which terms such as evidence and opinion are used in a disaligning sense ("You have no evidence" and "That's just your opinion and mine is different").

6. Related Work

6.1 Studies of Social Meaning in Sociolinguistics

The concept of *stance* or *stancetaking* has been long investigated by sociolinguists (e.g., Biber 2004; Du Bois 2007) and psychologists (Scherer 2005). Although there are slightly different notions of stance, all of them can be considered as different perspectives of

the same phenomena (Jaffe 2009). For our work we consider the social perspective of stance. Specifically, we follow the definition of Kiesling (2009), which is based on the *stance triangle* approach of Du Bois (2007). Kiesling defined stance as a person's expression of their relationship to their talk and to their interlocutors. This interactional or intersubjective aspect of stance has also been the focus of several others (Precht 2003; White 2003; Kärkkäinen 2006; Keisanen 2007). Thus the concept of stancetaking provides a unified framework for analyzing the different forms of *interactional styles*. For a survey of stance-related literature in linguistics, refer to Chindamo, Allwood, and Ahlsen (2012). In general, most of the work on stancetaking has been qualitative and context-bound. One of the advantages of our project is to test whether an analysis of stancetaking and the annotations can be performed in a replicable way that leads to a deeper understanding of how such concepts can be used to explain sociolinguistic patterns. Our results suggest that this is a profitable research pursuit, that such annotations and definitions could provide insight into some other patterns in sociolinguistics.

In terms of sociolinguistic theory, there are two ways this work is important. The first is as a model of stancetaking. With the exception of Du Bois (2007) and Kockelman (2004), most such models are relatively ad hoc and arise from specific conversations and contexts being analyzed. For example, Goodwin (2007) provides an insightful analysis of how conversations are embodied, and uses terms such as embodied stance, instrumental stance, epistemic stance, cooperative stance, moral stance, and affective stance, all of which are useful and relatively understandable, but not defined nor explicated as to what unifies them under the heading of stance. This style of analysis is typical, and leads to a wide range of uses and understandings. The definition on which our analysis rests (focused on relationships of animator to figures in talk), is relatively general but allows for a more consistent model. This definition is what results in the three dimensions of stance, which we have shown can be reliably encoded and used for an analysis of conversations.

Second, this work shows that we can find stancetaking patterns in conversation and across communities by using the stance model posited here. Ochs (1992) has argued that social identity categories are mediated by stance; she argues that stances (along with acts and activities) are the notions more likely to be what speakers are orienting to in an actual interaction. In other words, rather than choosing a linguistic form because it is masculine, a speaker might choose it because, for example, it is low investment, and such low investment stances are constitutive of masculinity. Similar semiotic processes could apply by implication to other social categories in addition to gender, such as class and ethnicity. Given this view in which stancetaking underlies sociolinguistic patterns, it is important to have a robust model of stancetaking that can be reliably applied in sociolinguistic studies. The results in this article suggest that such a project is entirely feasible, and that given enough data such identity patterns such as gender should become clear. Such data would also need identity information about the interactants, data that is not available for Reddit users.

From a qualitative perspective, Kiesling (2018; in press) has argued that masculine styles of speaking are connected to low investment stances. This suggests a connection to our results: Although we have no data on the gender of the forum participants, one can imagine that *r/Parenting* is likely to have more women than *r/explainlikeimfive*, and *r/Parenting* is the subreddit that has more range for AFFECT. In addition, our analyses suggest that some of the differences in stance are related to differences in the kinds of acts and activities that take place in the different subreddits, further supporting Ochs's (1992) model that these three notions can help explain patterns of language use by social category. Eckert's (2008) notion of indexical field is another sociolinguistic area in which

this model of stancetaking could be used. In her model, linguistic variants have a field of potential indexical meanings (the indexical field), including stance meanings. Our work shows that patterns of stancetaking might be associated with different variables that are coterminous with different communities such as subreddits. Eckert's (2008) way of thinking about style could also be profitably explored through this stance model, such that individual patterns of stancetaking develop into styles, much as Bucholtz and Hall (2005) argue for in their influential model of language and identity. Our results for the different subreddits suggest how such *stance accretion* (in Bucholtz and Hall's [2005] terminology) could take place. Overall, our analyses are a vindication of using stance as an explanatory concept in sociolinguistics generally, and specifically a systematic model of stancetaking that makes the distinction we suggest in the three dimensions.

6.2 Studies of Social Meaning in Computational Linguistics

Social and interactional meaning has garnered increasing interest in the computational linguistics community in recent years. Linguistic and social constructs such as sentiment (Wiebe, Wilson, and Cardie 2005), subjectivity (Riloff and Wiebe 2003), opinion mining (Pang, Lee et al. 2008), factuality (Saurí and Pustejovsky 2009), belief (Prabhakaran, Rambow, and Diab 2010; Werner et al. 2015), politeness (Danescu-Niculescu-Mizil et al. 2013), respect (Voigt et al. 2017), formality (Pavlick and Tetreault 2016), and power differences (Prabhakaran, Rambow, and Diab 2012) have been operationalized for computational investigation. Our operationalization of stancetaking using the dimensions of AFFECT, ALIGNMENT, and INVESTMENT is related to, although not the same as, constructs such as sentiment, subjectivity, opinion mining, and argumentation. One important way in which the notion of stancetaking differs from constructs such as sentiment, subjectivity, and opinion is that these constructs are studied in the context of a single utterance by a speaker/writer, whereas our model of stancetaking is not based on single speaker/writer, but is inherently dialogic and interactional in nature. Stancetaking is about the evaluation of entities in the discourse by a speaker and alignments/disalignments are created between speakers as they display similarity and difference with respect to these evaluations. Another relevant effort is the TAC 2017 Source-and-Target Belief and Sentiment Evaluation,[13] which unifies annotation for belief and sentiment toward entities within the text. However, the notion of stancetaking is different, again due to its inherent dialogic and interactional nature.

Stance and Stancetaking. Stancetaking is also distinct from other notions of *stance* in computational linguistics. One such notion is argumentative stance, which is about the position or stance (pro vs. con) that a speaker/writer takes on an issue in a debate (Walker et al. 2012). A slightly different notion of stance is presented in the SemEval-2016 Task 6 (Mohammad et al. 2016) on detecting stance from tweets. This task defined stance detection as the task of automatically determining whether the author of the text is in favor of, against, or neutral toward a (pre-chosen) proposition or target entity. However, our notion of stancetaking is different from these. We consider stancetaking as a multidimensional construct indicating the relationship between the audience, topic, and talk itself; and we capture it through the dimensions of AFFECT, ALIGNMENT, and INVESTMENT.

13 https://tac.nist.gov/2017/KBP/.

Crowdsourcing and Expert Annotations. As noted before, computational researchers have investigated a range of constructs related to social meaning. Many of these investigations have used crowdsourcing to build large data sets of annotations (e.g., Danescu-Niculescu-Mizil et al. 2013; Pavlick and Tetrault 2016; Voigt et al. 2017), shedding light on how these interpersonal constructs are generally understood. Many of these constructs can be explained through the unifying framework of interactional stancetaking, but the terms of this framework are not likely to be known to the typical crowdworker. For this reason, we have focused on "expert" annotations, mainly from students trained by the authors, who performed the task over the course of a semester or more. The applicability of crowdsourcing to the annotation of interactional stancetaking is a question for future work.

Lexicons. Another expert-driven perspective on social meaning is the use of carefully curated lexicons. Although there are many examples of this approach (e.g., Stone 1966; Biber and Finegan 1989; Taboada et al. 2011), the dominant example today is the Linguistic Inquiry and Word Count (LIWC) set of lexicons (Tausczik and Pennebaker 2010), designed by social psychologists to capture a broad range of social and cognitive phenomena. Several of LIWC's word lists touch on the stancetaking dimensions identified in this article: affect and positive and negative emotion, certainty and tentativeness, and inclusion and social phenomena. It is an open question as to whether these phenomena are best understood by annotating individual examples, or by listing words and phrases in the abstract. Researcher intuitions about a word's stancetaking properties may not always match reality: A classic counter-intuitive finding is that sentences starting with "please" are judged to be less polite on average (Danescu-Niculescu-Mizil et al. 2013); we observe a similar result with "please" indicating negative affect. Furthermore, the stancetaking properties of individual words and phrases are always shaped by the pragmatic context, including discourse and (especially in social media) extralinguistic factors (Benamara, Taboada, and Mathieu 2017). A further concern for lexicon-based methods is whether predefined word lists can possibly keep up with the variety and rapid change that predominate in online social contexts (Eisenstein 2013).

Unsupervised Learning. A third computational approach to characterizing interpersonal meaning in language has focused on the use of unsupervised techniques such as clustering, topic modeling, and matrix factorization (e.g., Schwartz et al. 2013). To ensure that the resulting latent dimensions are focused on interpersonal meaning rather than topic or genre differences, these approaches are often applied to restricted vocabularies, such as address terms (Krishnan and Eisenstein 2015) or stance markers (Pavalanathan et al. 2017). An advantage of unsupervised methods is that they can be applied to very large data sets; for example, Pavalanathan et al. (2017) analyze the use of stance markers in more than 50 million Reddit threads. But while the resulting dimensions represent words and phrases that pattern together, these groupings may reflect stylistic differences in the ways stances are enacted, rather than fundamental differences in the stances themselves. Manual annotations of stancetaking are therefore crucial to help ground and validate such unsupervised approaches.

6.3 Social Dynamics of Online Discussions

Prior investigations on the social dynamics of interactions in online discussions focus on the structural properties of individual conversations (Whittaker et al. 1998; Harper,

Moy, and Konstan 2009; Kumar, Mahdian, and McGlohon 2010; Backstrom et al. 2013; kooti et al. 2015) as well as the properties of the online community as a whole (Preece 2001; Lampe and Johnston 2005; Maloney-Krichmar and Preece 2005). Our focus on the thread structure properties of stancetaking utterances is closely related to the line of research studying the thread structure properties of individual conversations and how those properties are connected with various social phenomena.

The structural properties of discussion threads reveal the dynamics of social interaction in an online community. For example, properties of the conversational thread, such as the sequence of participant arrival, links among initial participants, and temporal comment arrival pattern, are found to determine member interactions such as re-entry to previously contributed discussions (Backstrom et al. 2013); length of the conversational thread is found to reveal interactivity among members (Whittaker et al. 1998); and depth and breadth of the threads are found to be the characteristics of the topic and authors of the discussions (Kumar, Mahdian, and McGlohon 2010). Conversational structure and meta-thread features are found to significantly improve agreement/disagreement detection in online debate forums, compared with using lexical features alone (Rosenthal and McKeown 2015). The conversational context in Twitter discussions is also found to be useful in sentiment classification (Ren et al. 2016). Recently, there has been work in modeling threaded conversations using neural networks, considering both the hierarchical structure and timing of comment arrival (Zayats and Ostendorf 2018) to predict the popularity of comments. In this work, we focus on another social phenomena, interactional stancetaking, and investigate how the thread structure interacts with various patterns of stancetaking utterances throughout threaded conversations.

7. Conclusion

One of the biggest difficulties in studying aspects of interaction such as stancetaking is subjective variability: Speakers and listeners often disagree about whether a single utterance is rude, offensive, or any other interpretation. Nevertheless, discourse analysts have been able to find ways of explicating stance, as described in this article. We have demonstrated a model of dividing up stancetaking such that it becomes more manageable for corpus annotation and analysis. By dividing stance into three dimensions, the relational aspects and meanings of interaction can be annotated more fruitfully. Our analysis of the words frequently associated with these dimensions provides further evidence that these dimensions are useful analytically. More importantly, as qualitative discourse analysts have long shown, where an utterance is in the unfolding sequence of discourse is crucial for its interpretation. This observation is crucial for stance given its relational definition. Our project has shown that this sequential embeddedness can be explored through the concept of stancetaking and its dimensions.

A key question for future work is the connection between interactional stancetaking and other phenomena. We hypothesize that stancetaking is connected to constructs such as politeness, formality, and affect. One way to test this is to obtain annotations for these constructs on the same Reddit data; alternatively, we can apply our computational models of stancetaking to data that has already been annotated for these other facets of social meaning. We are also interested in linking the stancetaking dimensions to social metadata, such as the community reception of comments containing various stance utterances, and the impact of these utterances on the trajectory of conversational threads.

Our hope is that the stance dimensions elaborated in this paper can be a useful tool for analyzing communication across a range of settings, both online and offline.

Appendix A. Reddit Stance Coding Guidelines

The following are the verbatim guidelines given to annotators of all threads.

These guidelines are descriptive and give some made up examples. Real examples are worth a thousand words, so please make sure to have a look at the two sample coded files as well.

A.1 Background

Stance creates relationships of speaker to some discursive figure, which is the focus of the stance and to other interactants. This discursive figure can be an interlocutor, a figure represented in the discourse, the speaker/author, ideas represented in the discourse, or other texts. If this list sounds a lot like "anything and everything," then you get the idea. In general, though, the stance focus is the thing that is made most relevant by an utterance.

A.2 Reddit Thread Structure

Reddit is an Internet discussion forum mega-site. It is divided by topics in many subreddits. In a particular subreddit, someone will post a comment, link, or question, and then anyone can comment on that initial post. Exchanges generally ensue, as other commenters comment in response to earlier ones. These exchanges resemble conversations. On the Reddit page, this structure is clear through indentation and other markers. In our coding sheets, the structure can be recovered although not through indentation but through cross-reference of "in response to." Comments are therefore not strictly chronologically organized. In addition, threads and comments can move up and down, depending on how many "upvotes" and "downvotes" the comment receives from other redditers.

A.3 Splitting Posts

Longer posts often have too many "moves or utterances" to characterize as having a single focus and stance, and they will have to be split up. Most posts will not need to be split. The length of the post is not necessarily an indication that it will need to be split. The test is mainly whether the primary stance focus is different and if there is a shift in stance during the post. If there is a post in which it is not possible to assign a single focus or stance to the whole post, then look for whether parts of the post could take a single focus or stance fairly easily, and split it so that each of those are in a different row. In order to do this, you need to insert a row and then copy the entire row to the new row, then edit the original and copy(s) so that only the coded text is present. The second column in the coding sheet is for keeping track of split posts. There are two numbers separated by a dash. The first number is the consecutive number of the original post using three digits. The number after the post is the order in which the text originally appeared.

For example, consider this (truncated) post:

Delanakatrella	003	t2_a02js	It can be really tough. I knew people at Columbine (which dates me since I was in 8th grade when Columbine happened). It still doesn't feel real to me. If you need someone to talk to feel free to PM - it really helps to talk to people.	cddwkq5

We divided like so:

Delanakatrella	003-01	t2_a02js	It can be really tough.	cddwkq5
Delanakatrella	003-02	t2_a02js	I knew people at Columbine (which dates me since I was in 8th grade when Columbine happened). It still doesn't feel real to me.	cddwkq6
Delanakatrella	003-03	t2_a02js	If you need someone to talk to feel free to PM - it really helps to talk to people.	cddwkq7

A.4 What to Code

First determine the stance focus. We are giving each comment a numerical stance code for each of the three dimensions (affect, investment, and alignment). In addition, we are giving each comment two more codes: one that describes the "activity" and one general descriptive stance adjective.

In general, try not to worry about what numbers you have already given, or whether or not you have a lot of variety throughout the thread. There might be some threads that end up being all threes or all fives. As tempting as it is, try not to go too quickly—try to have an explicit rationale for each number, even a neutral 3. Remember that the stance focus is what you should be worried about for the three stance dimensions. You may want to revisit what is outlined below before coding each thread!

A.5 Quick Reminders

Affect: Relationship to the stance focus. How does the author indicate they feel about the stance focus? This is a dimension of evaluation.
Investment: Relationship to the talk itself. How strongly does the author feel about their claim, view, etc.?
Alignment: Relationship to the previous author. Does the post show sympathy, agreement, mutual knowledge, common identity?

A.6 Stance Focus

The stance focus is the thing that is made most relevant by an utterance. This can be an entity—for example, if someone is talking about football and refers to the Steelers, the Steelers might be the focus, or one of the players. On the other hand, we do other things with language besides assert and evaluate things. We also do things like ask, insult, compliment, suggest, etc. These can also have stance foci, usually on individuals or the talk itself. So the first step in the analysis is to determine the primary stance focus. One of the best ways to determine the focus is to look at things that are given information in

the utterance, such as pronouns or things oriented to but not directly stated. Of course, things may actually be mentioned as well using NPs. If these are the focus then they are more likely to have a definite article *the* or the proximal deictic *this* (and possibly the distal *that*). These all signal in various ways that the "thing" is already in the discourse model and the reader's attention is being focused on it.

A.7 Affect

Affect is the polarity or quality of the stance to the stance focus. For example, if you are talking about food, and you say how yummy the French fries are, then the stance focus is the fries and they are evaluated positively. However, a focus can also be an act. In that case, affect has to do with whether the act itself is overtly positive or negative. So, a request done in an aggravated way ("Shut up!") is negative affect, but in a more mitigated way ("Could you please tone it down a bit?") is more positive.

Score	What it means
5	Most positive affect. Commenter expresses great admiration, appreciation, approval, etc., for the stance focus.
3	Neutral. Neither positive nor negative affect expressed.
1	Most negative affect. Commenter expresses derision, dislike, disgust for what they are talking about.

A.8 Investment

How strongly invested in the talk the speaker is; how committed they signal their relationship to the stance focus. Would they defend their claims and opinions to the death? This dimension is about the talk itself.

It is true that by posting at all there is a certain amount of investment displayed, but since we do not know how many people have viewed each post but not commented, we really cannot use "not commenting" as the lowest level of investment.

Questions: It might seem that questions are by default low investment, and they often are. But they can also be challenging of previous comments and thus show a higher investment in assumptions underlying the question. The upshot is that do not code questions with low investment automatically without considering how they are fitting into the overall conversation.

Some things that are important to look for are hedges, which definitely reduce the investment. So things like "I think" and "maybe" and "just" are signs of a lower investment. But use your judgment—these can be used ironically as well and end up with the opposite effect!

Score	What it means
5	Highest investment. Commenter seems absolutely certain of claims and would defend them to eternity
3	Neutral. Assertion is neither strong nor weak.
1	Lowest investment. Commenter expresses uncertainty.

A.9 Alignment

How a speaker/writer aligns (or not) to their interlocutor(s), real or imagined, vis-vis the stance focus. Alignment is almost always present at a basic level in that the interlocutor must attend to the same discourse entities in order to hold a basic conversation. That is, speakers orient to the same things in talk. But people do not always align to the objects and figures in talk in the same way. Alignments and disalignments can occur in many ways, and we have to attend to all of them.

Alignment will almost always be with respect to the person who just talked, but alignment can be created prospectively or to a more general audience, especially in the case of Reddit, which is open for anyone on the Internet to read. If in doubt, evaluate alignment based on alignment to the author of the post to which the current post is a response. Do not try to imagine all of the audiences on the Internet!

Score	What it means
5	Highly aligned. Expressing agreement, sympathy, elaborating on what someone else has said or is figured as saying, etc.
3	Average alignment. Unmarked but not disaligning either.
1	Highly disaligned. Resisting doing a second pair part (like not answering a question), disagreeing, criticizing, being dismissive, etc.

A.10 Activities

Eventually we will have a list of these. For now, put in what you think the person is doing with their speech (for example informing, criticizing, asking, etc.). Your verb should be in the X+ing form.

A.11 Stance Adjectives

Eventually we will have a list of these. For now, put the general way you would describe the person's stance (for example helpful, critical, skeptical, etc.).

A.12 Notes

Put any comments at all in here, especially about how you chose your codes or questions about them or how difficult they were.

A.13 Tricky Things to Watch Out For

A.13.1 Sarcasm. Sarcasm is one of the hardest things we have to deal with, since it is generally intended to have the opposite effects than what is obvious. If you detect it, you should lower the investment because it is a way of distancing oneself from the claims made. Affect is usually lower too, but it depends on the statement (that is, if the statement is a negative evaluation and the opposite is meant, then the affect is actually higher). Alignment is the trickiest, but in general if the sarcasm is meant to support or align with the previous or original post, then raise the alignment. Also raising alignment

for a sarcastic post is when it draws the reader in to make shared assumptions in order to recognize the sarcasm.

A.13.2 In Jokes. An in-joke is one in which the reader needs some community knowledge in order to get the joke. These are one thing to consider when a post does not seem to make sense. Sometimes one can tell based on a further post's reaction to it, but one might need to do some research on terms, etc., to find out. If you are not sure please make a note, because an in-joke significantly changes some of the stance scores.

Namely, the investment goes down (it is a joke after all), and the alignment goes up. The latter is because the in-joke virtually identifies the assumed reader as a member of the community with common knowledge, thus implicitly aligning the author and reader.

Acknowledgments

We thank the editors and the anonymous reviewers for their helpful and constructive feedback. This research was supported by Air Force Office of Scientific Research award FA9550-14-1-0379, by National Institutes of Health award R01-GM112697, and by National Science Foundation awards 1452443, 1111142, and 1110904. An early version of this research was presented at the Southeastern Conference on Linguistics (SECOL 82). We especially thank the students who helped annotate the data: Sarah Hochrein, Erica Hom, Joni Keating, Kaylen Sanders, and Charlene Shin.

References

Anand, Pranav, Marilyn Walker, Rob Abbott, Jean E. Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 1–9, Portland, OR.

Backstrom, Lars, Jon Kleinberg, Lillian Lee, and Cristian Danescu-Niculescu-Mizil. 2013. Characterizing and curating conversation threads: Expansion, focus, volume, re-entry. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, pages 13–22, Rome.

Benamara, Farah, Maite Taboada, and Yannick Mathieu. 2017. Evaluative language beyond bags of words: Linguistic insights and computational applications. *Computational Linguistics*, 43:201–264.

Benjamini, Yoav and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57:289–300.

Biber, Douglas. 2004. Historical patterns for the grammatical marking of stance: A cross-register comparison. *Journal of Historical Pragmatics*, 5(1):107–136.

Biber, Douglas and Edward Finegan. 1989. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text*, 9(1):93–124.

Bucholtz, Mary and Kira Hall. 2005. Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7(4-5): 585–614.

Chindamo, Massimo, Jens Allwood, and Elisabeth Ahlsen. 2012. Some suggestions for the study of stance in communication. In *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*, SOCIALCOM-PASSAT '12, pages 617–622, Amsterdam.

Craggs, Richard and Mary McGee Wood. 2004. A categorical annotation scheme for emotion in the linguistic content of dialogue. In Andre, L. Dybkjær, W. Minker, and Heisterkamp, editors, *Tutorial and Research Workshop on Affective Dialogue Systems*, pages 89–100, Springer.

Danescu-Niculescu-Mizil, Cristian, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 250–259, Sophia.

Defazio, Aaron, Francis Bach, and Simon Lacoste-Julien. 2014. Saga: A fast incremental gradient method with support for non-strongly convex composite objectives. In *Advances in Neural Information Processing Systems*, pages 1646–1654, Montreal.

Du Bois, John W. 2007. The stance triangle. In Robert Engelbretson, editor, *Stancetaking in Discourse*. John Benjamins Publishing Company, pages 139–182.
Du Bois, John W. and Elise Kärkkäinen. 2012. Taking a stance on emotion: Affect, sequence, and intersubjectivity in dialogic interaction. *Text and Talk*, 32(4):433–451.
Eckert, Penelope. 2000. *Language Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Wiley-Blackwell.
Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics*, 12(4):453–476.
Eckert, Penelope and Sally McConnell-Ginet. 1992. Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology*, 21(1):461–488.
Eisenstein, Jacob. 2013. What to do about bad language on the Internet. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NACCL)*, pages 359–369, Atlanta, GA.
Giles, Howard, Justine Coupland, and Nikolas Coupland. 1991. *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge University Press.
Goffman, Erving. 1981. *Forms of Talk*. University of Pennsylvania Press.
Goodwin, C. 2007. Participation, stance and affect in the organization of activities. *Discourse & Society*, 18(1):53–73.
Grice, H. Paul. 1975. Logic and Conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics, Vol. 3: Speech Acts*. Academic Press, New York, pages 41–58.
Harper, F. Maxwell, Daniel Moy, and Joseph A. Konstan. 2009. Facts or friends?: Distinguishing informational and conversational questions in social Q&A sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 759–768, Boston, MA.
Hayes, Andrew F. and Klaus Krippendorff. 2007. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1):77–89.
Jaffe, Alexandra. 2009. *Stance: Sociolinguistic Perspectives*. Oxford University Press.
Kärkkäinen, Elise. 2006. Stancetaking in conversation: From subjectivity to intersubjectivity. *Text & Talk—An Interdisciplinary Journal of Language, Discourse & Communication Studies*, 26(6):699–731.
Keisanen, Tiina. 2007. Stancetaking as an interactional activity: Challenging the prior speaker. In R. Engelbretsen, editor, *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, pages 253–281.
Kiesling, Scott, Jacob Eisenstein, Jim Fitzpatrick, and Umashanthi Pavalanathan. 2015. The development of a stance annotation scheme: Lessons for computational linguistics and sociolinguistic theory. In *82nd Meeting of the Southeastern Conference on Linguistics (SECOL 82)*, Raleigh, NC.
Kiesling, Scott F. 2018. Masculine stances and the linguistics of affect: On masculine ease. *NORMA: International Journal for Masculinity Studies*. doi:10.1080/18902138.2018.1431756.
Kiesling, Scott F. in press. Stances of the 'gay voice' and 'brospeak': Towards a systematic model of stancetaking. In R. Barrett and K. Hall, editors, *Oxford Handbook of Language and Sexuality*. Oxford University Press, New York.
Kiesling, Scott Fabius. 2009. Style as stance. In A. Jaffe, editor, *Stance: Sociolinguistic Perspectives*. Oxford University Press, page 171–194.
Kockelman, Paul. 2004. Stance and subjectivity. *Journal of Linguistic Anthropology*, 14(2):127–150.
Kooti, Farshad, Luca Maria Aiello, Mihajlo Grbovic, Kristina Lerman, and Amin Mantrach. 2015. Evolution of conversations in the age of email overload. In *Proceedings of the 24th International Conference on World Wide Web*, pages 603–613, Florence, NY.
Krippendorff, Klaus. 2007. Computing krippendorff's alpha reliability. *Departmental Papers (ASC)*, page 43.
Krishnan, Vinodh and Jacob Eisenstein. 2015. "You're Mr. Lebowski, I'm The Dude": Inducing address term formality in signed social networks. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. pages 1616–1626, Denver, CO.
Kumar, Ravi, Mohammad Mahdian, and Mary McGlohon. 2010. Dynamics of conversations. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 553–562, Washington, DC.
Lampe, Cliff and Erik Johnston. 2005. Follow the (slash) dot: Effects of feedback on new members in an online community. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group*

Work, GROUP '05, pages 11–20, Sanibel Island, FL.
Lave, Jean and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
Lempert, Michael. 2008. The poetics of stance: Text-metricality, epistemicity, interaction. *Language in Society*, 37(04):569–592.
Levinson, Stephen C. 1992. Activity types and language. In J. Heritage and P. Drew, editors, *Talk at Work: Interaction in Institutional Settings*. Cambridge University Press, pages 66–100.
Maloney-Krichmar, Diane and Jenny Preece. 2005. A multilevel analysis of sociability, usability, and community dynamics in an online health community. *ACM Transactions on Computer-Human Interaction*, 12(2):201–232.
Mohammad, Saif, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, CA.
Ochs, Elinor. 1992. Indexing gender. *Rethinking Context: Language as an Interactive Phenomenon*, 11:335.
Ochs, Elinor. 1993. Constructing social identity: A language socialization perspective. *Research on Language and Social Interaction*, 26(3):287–306.
Pang, Bo, and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.
Pavalanathan, Umashanthi, Jim Fitzpatrick, Scott Kiesling, and Jacob Eisenstein. 2017. A multidimensional lexicon for interpersonal stancetaking. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 884–895, Vancouver.
Pavlick, Ellie and Joel Tetreault. 2016. An empirical analysis of formality in online communication. *Transactions of the Association for Computational Linguistics*, 4:61–74.
Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
Prabhakaran, Vinodkumar, Owen Rambow, and Mona Diab. 2010. Automatic committed belief tagging. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 1014–1022, Beijing.
Prabhakaran, Vinodkumar, Owen Rambow, and Mona Diab. 2012. Predicting overt display of power in written dialogs. In *Proceedings of the North American Chapter of Association for Computational Linguistics (NAACL)*, pages 518–522, Montreal.
Precht, Kristen. 2003. Stance moods in spoken English: Evidentiality and affect in British and American conversation. *Text—Interdisciplinary Journal for the Study of Discourse*, 23(2):239–258.
Preece, Jenny. 2001. Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology*, 20(5):347–356.
Ren, Y., Y. Zhang, M. Zhang, and D. Ji. 2016. Context-sensitive Twitter sentiment classification using neural network. *Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016)*, pages 215–221, Phoenix, AZ.
Riloff, Ellen and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, pages 105–112, Sapporo.
Rosenthal, Sara and Kathy McKeown. 2015. I couldn't agree more: The role of conversational structure in agreement and disagreement detection in online discussions. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Prague.
Sacks, Harvey. 1995. *Lectures on Conversation*. Wiley-Blackwell.
Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696.
Saurí, Roser and James Pustejovsky. 2009. Factbank: A corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3):227.
Schegloff, Emanuel A. and Harvey Sacks. 1973. Opening up closings. *Semiotica*, 8(4):289–327.
Scherer, Klaus R. 2005. What are emotions? And how can they be measured? *Social Science Information*, 44(4):695–729.
Schiffrin, Deborah. 1988. *Discourse Markers*. Studies in Interactional Sociolinguistics. Cambridge University Press.

Schwartz, H. Andrew, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E. P. Seligman, and Lyle H. Ungar. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, 8(9):e73791.

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*, volume 626. Cambridge University Press.

Sidnell, Jack. 2011. *Conversation Analysis: An Introduction*. John Wiley & Sons.

Stone, Philip J. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.

Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.

Tausczik, Yla R. and James W. Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.

Thelwall, Mike, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the Association for Information Science and Technology*, 61(12):2544–2558.

Voigt, Rob, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt. 2017. Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences, U.S.A.*, 114(25):6521–6526.

Walker, Marilyn A, Pranav Anand, Robert Abbott, and Ricky Grant. 2012. Stance classification using dialogic properties of persuasion. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 592–596, Montreal.

Werner, Gregory, Vinodkumar Prabhakaran, Mona Diab, and Owen Rambow. 2015. Committed belief tagging on the factbank and lu corpora: A comparative study. In *Proceedings of the Second Workshop on Extra-Propositional Aspects of Meaning in Computational Semantics (ExProM 2015)*, pages 32–40, Denver, CO.

White, Peter R. R. 2003. Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text—Interdisciplinary Journal for the Study of Discourse*, 23(2):259–284.

Whittaker, Steve, Loren Terveen, Will Hill, and Lynn Cherny. 1998. The dynamics of mass interaction. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, CSCW '98, pages 257–264, Seattle, WA.

Wiebe, Janyce, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2):165–210.

Zadeh, Amir, Rowan Zellers, Eli Pincus, and Louis-Philippe Morency. 2016. Mosi: Multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos. *arXiv preprint arXiv:1606.06259*.

Zayats, Victoria and Mari Ostendorf. 2018. Conversation modeling on Reddit using a graph-structured LSTM. *Transactions of the Association of Computational Linguistics*, 6:121–132.