Discovering Habits of Effective Online Support Group Chatrooms

Elijah Mayfield and Miaomiao Wen
Carnegie Mellon University
Pittsburgh, PA 15213
{emayfiel, mwen}@cs.cmu.edu

Mitch Golant
The Cancer Support Community
Washington, DC
mitch@cancersupportcommunity.org

Carolyn Penstein Rosé
Carnegie Mellon University
Pittsburgh, PA 15213
cprose@cs.cmu.edu

ABSTRACT

For users of online support groups, prior research has suggested that a positive social environment is a key enabler of coping. Typically, demonstrating such claims about social interaction would be approached through the lens of sentiment analysis. In this work, we argue instead for a multifaceted view of emotional state, which incorporates both a static view of emotion (sentiment) with a dynamic view based on the behaviors present in a text. We codify this dynamic view through data annotations marking information sharing, sentiment, and coping efficacy. Through machine learning analysis of these annotations, we demonstrate that while sentiment predicts a user's stress at the beginning of a chat, dynamic views of efficacy are stronger indicators of stress reduction.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Discourse*;

Keywords

discourse analysis; efficacy; sentiment analysis; social media; synchronous chat; group dynamics; information exchange

1. INTRODUCTION

The Internet provides invaluable resources for adults who suffer from chronic or life threatening diseases. Beyond the physical hardships associated with many of these conditions, patients suffer from intense emotions like fear, anger, and sadness, which lead to the experience of stress. Online support groups have the tremendous potential to make social support experiences accessible to patients with limited social circles or even physical mobility [41]. Nevertheless, while the positive effects of face-to-face support groups have been studied extensively, the inner workings of online support groups are much less well worked out. For example, facilitation in face-to-face support groups is well studied [1], but online, the role of a facilitator is still unclear. Additionally, most prior work studying online support groups has focused on asynchronous communication, for example on maintaining group membership [52] or meeting needs of caregivers [49]. Our work focuses instead on sychronous chat.

In this paper we present operationalizations that are valuable for monitoring and understanding the stress of each participant in an online support group context. Improving automated analysis in this context has potential benefits for facilitators, both in real-time, as interfaces can be designed to support facilitation through this style of analysis, and after the fact, as a source of insight for training, reanalysis, and summarization. The computer-supported collaborative work (CSCW) community has provided inroads to both of these applications. In real-time, triggered interventions in an online conversational setting can greatly benefit from a deeper understanding of communication processes [23, 24]. Interface designs supporting group discussions, especially for facilitators, have also been shown to be effective in chatroom contexts for learning [47] and project meetings [11]. After the fact, social network analysis has used graphs between participants to study group problem solving dynamics [46] or for performing structural analysis of group social networks [14, 42]

We focus in this work on post hoc analysis of chat, in particular on two aspects of group interactions. From a linguistic standpoint, the text of a chatroom discussion can be viewed through the lens of sequences of interaction; a higher-level view of the group dynamics of a chat may focus on the linking of participants together through network analysis of this linguistic connection between users. Second, from a psychological standpoint, we study the emotions and coping strategies of participants in online chat, at a level that allows us to understand the potential shortfalls of coarse-grained sentiment analysis. In order to make the best use of the data that is available to developers of groupware systems, a deep understanding of the processes behind both linguistic and psychological aspects of chat is required.

Chatrooms elicit a complex structure of multiparty conversation, and understanding this structure is of particular importance for sociolinguistic research. In particular, the disentangling of threads of conversation, and the flow of information and emotional support between speakers, lends insight into how groups communicate. Understanding how to represent threading has been a problem in the CSCW

GROUP'12, October 27–31, 2012, Sanibel Island, Florida, USA.

community for many years [45]. Often, CSCW researchers have followed the conversation analysis framework and the notion of adjacency pairs from Schegloff [43]. One study has directly applied this framework to the context of online support groups [35], identifying the structure of interactions between facilitators and patients.

In prior work attempting to understand the emotion of speakers, identifying the coping strategies and the social connections being made between speakers has typically been formulated as a sentiment analysis problem, where speakers' opinions are labelled as either positive or negative. This formulation has been used to understand group discussions in numerous contexts, such as political blog posts [39] or visualizations of computer-mediated communication [32]. In these works, emotion or sentiment is treated as a property of the language employed in the text. Language then reflects the emotional state of the one who has uttered that text.

In contrast, our work aligns with a more complex view of emotion, in which text reflects not just the bias of the one who has uttered the text, but also of their assumed audience. This view has been largely explored in sociolinguistic literature, both in the context of fiction writing [7] and publication within academic communities [19]. Work in this tradition expects a text not only to display the projected image of the author but also that of the assumed audience. In our work, we explore an additional factor, namely, the contrast between the reflection of the participant's emotional state and the reflection of how that emotional state is changing through participation in the interaction of a chatroom. Automating this style of analysis has led to more complex modelling of bias detection [33] or contentiousness of viewpoints in news articles [40].

To further researchers' understanding of these processes, we demonstrate this contrast in the context of online support groups through an empirical analysis of chatroom behavior, both at the level of basic information flow, and at the level of framing the coping experience itself. We demonstrate that the features that predict emotional state are different from the features that predict changes in emotional state over the duration of the conversation. This analysis argues against the idea that emotion is a property of text, and argues instead for a slightly more nuanced view. By this understanding, conversational strategies operate on emotions, and thus emotion and processing emotion may be operationalized as distinct social processes that are encoded in language in distinct ways. To motivate the annotation of our transcripts used by the machine learning models, and explanatory theory from the social cognitive processing model [25] as well as social cognitive theory [3].

In the remainder of the paper we review related work related to theories of coping and work related to analysis of emotion in text. We then describe the data used in our analysis, how the data provides opportunities to contribute to the literature on support groups as well as basic research on conversation. Next we discuss the operationalization of the reliable annotation schemes we have developed, and our annotation effort. Then we present regression models that show the impact of these annotations both on the level of language use, as well as higher-level annotations, for analyzing both entrance stress and shift in stress of a participant over the course of a chat. We conclude with a discussion of the limitations of this work and plans for continued research.

2. PRIOR WORK

Our work draws from and contributes to several bodies of literature. First, it draws from the extensive literature on coping, from which we primarily draw from the Social Cognitive Processing model and Social Cognitive Theory. Among other things, this literature promotes disclosure of negative affect in support group interactions. We also examine our data from a linguistic point of view, understanding the nature of discourse and discursive processes for sharing information. Finally, the language technologies community has shown great promise in simpler views of sentiment and emotion recognition, which contributes to the notions of stress and coping that we are interested in. Because we wish to facilitate automated analysis in later stages of our research, we also take into account work from this field.

2.1 Social Psychology Theories of Coping

The coping literature provides us with lenses through which we can view conversational behavior in our chat corpus from the perspective of pscyhological processes that have already been identified in the social psychology literature. One such lens is the Social Cognitive Processing model [25], which places a great emphasis on disclosure, positive social settings, and disinhibition. A model that provides a complementary perspective, investigating the constructs of coping efficacy and empowerment, is that of Social Cognitive Theory [3]. We review both here.

The social cognitive processing model defines the coping process as one of intense emotional encounters that provide opportunities for confronting and then integrating traumatic experiences in ones psyche, which is referred to as "adaptation" [25]. In this paradigm, cancer is a traumatic life event that shifts a person's core beliefs about themselves. In this model, stress is caused by a discrepancy between a person's mental model of themself, formed over a lifetime, and the "new normal" that emerges after being diagnosed. Observed events do not match a person's preconceptions. Reducing stress therefore requires either *assimilation*, where events are reappraised to fit preconceptions, or *accommodation*, changing preconceptions to fit new information.

These notions can be easily associated with conversational practices. For example, patients who share a great deal of information about themselves, especially through narrative and contemplation, should be expected to reduce stress. A major factor impeding this information sharing, however, is social constraint. Unsupportive social environments, where sharing information is met with unexpected responses or suppression of discussion, are likely to impede the cognitive processing that leads to stress reduction. On the other hand, positive responses and empathetic responses are more likely to foster an environment that encourages emotional expression and validation. This is particularly important in expressing primary negative affect– fear, anger and sadness. Social constraint on disclosure of these emotions leads to increased anxiety, suppressing anger increases the likelihood of hostility, and suppression sadness increasing the likelihood of depression [12].

The issue of social constraint becomes even more important in a chatroom-based online setting, where the only form of support is via a text-based exchange. Walther and others have underscored the "self-disclosure miracle" associated with online support groups: the only method of connection is by sharing about yourself [50]. However, social cues for cohesively handling multiple "threads" of conversation are more limited in a text-only context [6]. Tone is also difficult due to the lack of both verbal and non-verbal cues, such as body language, and participants overestimate their ability to convey tone [22]. Therefore, informational and emotional exchanges with others who understand your experiences is believed to enhance social cognitive processing.

In our qualitative explorations of the chat data, we noticed that one distinguishing characteristic of participants was the extent to which they framed their hardships as fixable or unfixable. In some cases, this characteristic distinguished individuals who participated in the same interaction, but came out with differential effects on reported distress. In particular, we observed participants who engaged in discussion about their issues framed as fixable and came out reporting a reduction in distress, and in the same interaction observed participants who engaged in discussions in which issues were framed as unfixable, and who came out reporting increased levels of distress. One goal of the analysis we report in this paper was to investigate whether this pattern is stable across the corpus, or only idiosyncratic. In our investigation of the coping literature, we found evidence from questionnaire based research that beliefs in ones ability to cope with hardship is associated with positive tangible effects on well-being [37]. The general concept of Self-efficacy, which is defined as people's belief in their ability to effect change in a particular context, has its roots in Social Cognitive Theory [3]. The more specific relevant construct within the coping literature is the concept of coping efficacy, which can be defined as a person's belief that they have the ability to impact their own well-being, as it pertains to their experience with their illness, either physically or emotionally. Drawing from our observations related to framing, we expect to find evidence of coping efficacy in an operationalization of the framing of hardships in the online discussions [48].

2.2 Information Authority in Discourse

The Negotiation framework, as formulated by the systemic functional linguistics (SFL) community, places a special emphasis on how speakers function in a discourse as sources or recipients of information or action. We use a formulation rooted in the sociolinguistic literature [26]. This work highlights the moves that are made in a dialogue as they reflect the authoritativeness with which those moves were made, and gives structure to exchanges back and forth between participants. We are interested in this framework because of its descriptiveness for social interactions, which makes it easy to track shifts in positioning and the sources of information over time.

We use a formulation of Negotiation which has previously shown to be reliable and automatable for coding using machine learning [31]. The key insight from these codes (defined in section 4.1) is a notion of the source of information, resulting in a metric of authority and ownership over information. This definition of authority, based in the Negotiation framework, has been shown in prior work to have meaningful relationships with constructs that are important to analysis of group interactions. In a learning setting, measurements of authoritativeness using this definition have correlated with learning gains [18] and self efficacy [17]. In task-based dialogue with two speakers, per-line annotation of authority has assisted in predicting task success [29] and in identifying differences in practices across cultures [28]. To our knowledge, however, the work we present here is the first study of authoritativeness's impact in a purely social setting.

For our purposes, this framework also provides an important advantage over analysis using adjacency pairs from conversation analysis. While initiation-response pairs give structure to a dialogue, they do not give a clearly-defined unit of analysis for emotion or higher-level concepts, such as efficacy. Indeed, for long sequences of information exchange, the complex structure of adjacency pairs rapidly becomes unreliable between human annotators. The Negotiation framework, by contrast, focuses heavily on a relatively flat notion of sequences, emphasizing the exchange of information between speakers. This makes this framework ideal for defining a unit of analysis above an individual utterance.

2.3 Automated Analysis of Emotion in Text

In the past decade, there has been a consistent stream of work in the language technologies community related to the analysis of emotions in text using machine learning and text mining techniques. A thorough review of this literature is beyond the scope of this paper, thus we restrict ourselves to mentioning a few representative publications. Some of the most famous early work demonstrated that machine learning models for classifying sentiment of product reviews and movie reviews outperformed rule based models [38]. Other work has focused on how transitory emotions are expressed within an ongoing text or discourse [20]. A series of investigations have sought features that are effective predictors of sentiment. The greatest improvements in performance have been achieved with features that insightfully capture the essence of the linguistic constructions used to express sentiment [53, 21]. Other recent work investigates how to achieve greater generality in trained models for sentiment analysis [8].

Our current work is not concerned with automatic detection of sentiment. Rather, what we explore is the conceptualization of the relationship between emotion and text. In particular, below we discuss how we distinguish expression of general outlook from the more specific expressions of ones conception of their own coping efficacy or that of others. This paper demonstrates the value of these more nuanced operationalizations of emotion in text, which provides motivation for our planned future work in which we adapt and extent techniques from the area of sentiment analysis to the problem of predicting these three more specific constructs.

3. DATA COLLECTION

Our data comes from the Cancer Support Community[1]. Participants were recruited and provided with a wide range of support services such as educational services, physician lectures, and exercise programs. Each conversation took place in the context of a weekly meeting in a real-time chat an online chatroom over the course of one hour, with up to 6 participants in addition to a professional therapist facilitating the discussion. Facilitators were explicitly asked to avoid developing 'therapeutic relationships' with online group members and to instead use their comments to promote the development of social support among members and to focus group discussion. Group members were also en-

[1]www.cancersupportcommunity.org

K2				C	[M], fast question, did your son have a biopsy?
K2				C	or does that happen when he comes home
	K1			V	i have 3 dogs.
	K1			V	2 are new puppies so they keep me happy.
	K1			V	man's best friend
	f			S	:-D
	o			C	and women
	K2			J	what kind of dogs????
		K1		C	[D], I keep seeing that you are typing and then it stops
		K2		C	how are you doing this week
	K1			V	i feel ery surrounded by women lol.
	K1			V	the puppies are a maltese/yorkie mix and the full grown is a pomaranian/yorkie.
K1				M	No, he did not have a biopsy.
K1				M	The surgeon that he saw examined him and said that by feel, he did not think the lump was cancerous, and he should just wait until he got home and saw a surgeon to see what he suggested.
f				C	that has to be very hard
			o	M	A question, however– [J], you would probably know.
			K2	M	He was told that they could not just do a needle biopsy, that he would have to remove the whole lump in order to tell if it was malignant.
		o		D	Yes.
		K1		D	I was waiting for [M] to answer.
			K1	J	That sounds odd to me

Table 1: An example excerpt with Negotiation labels and threading structure annotated.

couraged to interact using an asynchronous discussion board during the week; those posts are not studied in this paper.

This corpus provides the opportunity to study support groups from a fresh angle. Studies of online support groups have mostly focused on bulletin boards or forums [4]. In our work, on the other hand, we study chatrooms, which allow for a synchronous discussion that is not supported by forum discussions. Previous work has described facilitators' perceptions of differences between F2F and online support groups [36]. Limitations of the medium are the primary concern, including difficulty pacing the discussion and creating cohesion within the group.

In addition to contributing to the literature on online support, our collected corpus allows us to investigate issues related to a distinct flavor of group discussions than what has been the object of study within the language technologies community. Freely available benchmark corpora are more limited in scope than our corpus along a variety of dimensions. For example, the well known Switchboard corpus contains telephone conversations, but is limited to two speakers, and each conversation averages only six minutes in length [13]. Aoki et al. [2] studied floor-taking in small-group sociable talk, and had large groups of up to ten speakers, but studied only seven one-hour conversations. These studies involved spoken language interactions, and therefore do not directly address the question of group participation in a chatroom setting. In online synchronous interactions, [10], studying IRC chatroom thread disentanglement, performed their experiments on a corpus of one conversation held over two hours, totalling 1,500 utterances, while [44] collected approximately 1,600 utterances in the context of classroom-based text message threads.

By contrast, our conversations are held between many speakers, in a chatroom setting, and persist over a total of one hour each. Our corpus consists of 21 such conversations totalling 9,365 lines of chat. In addition, our unlabelled data, which may be useful for future work, has over 2,000 such conversations spanning hundreds of users over the course of five years. Therefore our study, though larger than any known previous work on chatroom dynamics at this fine-grained level of analysis, has massive room for growth in our future work.

What makes our corpus particularly valuable for addressing our question about how language processes encode evidence of reflecting stress and processing stress is the inclusion of a per-user, per-conversation self-reported measure of distress. Upon entering and exiting the chatroom, participants were asked to mark their level of distress on a scale from 0 to 10, with non-response left available as an option. In assembling our corpus of 21 conversations, we selected only those with a high response rate on distress indicators. Thus, in no conversations that we sampled did more than one speaker provide a non-response to the distress measure. To the extent that we are successful in learning which features of conversational interactions encode the process of reflecting stress or processing stress, these features will have the potential to offer insight into these processes even in the absence of these external metrics.

4. ANNOTATION

Our analysis uses two levels of data annotation. These conversations were annotated for information sharing and authority with the Negotiation framework [26], a sociolinguistic framework descring these factors in conversation. These annotations are made per-line. Next, we group the turns into threads, using the structure from the turn-level Negotiation coding as an intermediary. Finally, we annotate coping efficacy and sentiment at the level of threads. In this section we describe each of these annotation schemes in detail.

4.1 Information Sharing

The Negotiation coding scheme attempts to code information exchange on a turn-by turn basis. To this end, we use the following annotations:

- Utterances containing new information are marked as **K1**, as the speaker is the "primary knower." These utterances can be either opinions, retelling of narrative or other contextualized information, or general knowledge about the world, for instance about medicine or hospital procedures.
- Utterances requesting information are marked **K2**, or "secondary knower," when the speaker is signalling that they want information from another speaker in the dialogue.
- Social feedback moves are labelled **f**, when there is clear sentiment attached but no new content. This can include emoticons, as well as fixed expressions such as "good luck" or social niceties, such as thanking or friendly banter.
- Moves which directly subvert or contradict a previous move are labelled **ch**, for challenge. These lines are more than just disagreement, which is labelled a K1 - instead, a challenge move rejects a premise from a previous turn.
- All other moves are labelled **o**. This includes for example typo correction, delaying or attention grabbing moves, or moves for tracking of the conversation (such as naming the target of an upcoming move).

There are multiple advantages to using the Negotiation framework to annotate information exchange. First, it has well-defined notions of what does or does not count as information exchange, meaning that inter-annotator reliability is achievable at a high level. Second, there is a notion of *sequences*, with internal constraints on ordering of labels based on observed attributes of language use. For instance, a **K2** move should not follow a **K1** move in the same sequence, and a speaker should not give a **K1** move in response to their own **K2**. A full treatment of these constraints is given in [31]. In that work, this layer of structure was built into machine learning models to enhance the accuracy of those systems for automated coding.

One final attribute that we can define based on Negotiation labels is the *authoritativeness* of a speaker. This construct refers not to the control over a discourse directly, but rather to the source of information in a discourse. Using Negotiation labels, we can define authoritativeness as a continuous variable, assigned per speaker, based on their ratio of information sharing to requesting:

$$Auth(S) = \frac{\#\ K1}{\#\ K1\ +\ \#\ K2} \tag{1}$$

This framework is not specifically dedicated to the domain of chatrooms nor does it have any codes related specifically to cancer ,support groups, or even the medical domain. This is advantageous for the prospect of future work using these annotations. In the language technologies community, automatic coding of information exchange, including the Negotiation framework, has rapidly been advancing. This is likely to benefit our work more quickly if the models built for other domains can be easily transferred to our data. Therefore, we do not attempt to refine these categories for a topic-specific set of codes.

4.2 Defining Threads

Earlier in this paper we discussed the notion of threads as a series of sequences, potentially entangled, that are based on a shared topic. In this section, we elaborate on their definition and describe ways in which we have annotated thread-level behavior.

It is easy to observe that social talk, especially that with multiple participants, often diverges into separate simultaneous conversations, which we term threads. These behaviors have been studied quantitatively before through the lens of conversation analysis [2], inspired by social science, or through discourse coherence [10], a more information theoretic approach. Others have framed the problem as a topic detection [5] or information retrieval [44] task. We attempt to capture aspects of each of these approaches in our annotation. Humans can navigate these conversations with ease, even in unfamiliar environments. Threading is occasionally discussed explicitly in our chatroom data, e.g.:

> **S:** I know you said you have some trouble with all the jumping around that happens online.
> **S:** Are you okay with it now? Have you gotten used to it? I know it's a little more so with more folks?
> **M:** Yes, it's better now.
> **M:** I just post a response and don't worry so much about being out of sequence.
> **S:** [M], good!
> **B:** that's the only way to do it [M] we all just try the best we can.
> **B:** it helps to direct your answer to the person who made a statement that you are replying to.

For our annotation, we begin with Negotiation sequences as our unit of analysis. These sequences, based on constraints, are useful as a stepping stone to defining threads. We group sequences based on their topic using intuitive judgements, along with heuristic rules - for instance, connecting sequences based on anaphora or coreference to entities in a previous sequence.

4.3 Thread-Level Annotations

For each of these threads, we wish to assign each user participating in those threads an annotation describing their attitude, especially as it relates to coping efficacy. We use the following annotation schemes:

4.3.1 Sentiment

We mark each user as expressing positive (1), neutral (2), or negative (3) sentiment. Positive sentiment includes affect as well as many social supportive moves or expressions of hope, while negative sentiment can include uncertainty or worry in addition to affect. All annotation of sentiment was performed manually and holistically based on the entire content of a user's contribution to a thread, rather than relying on keyword-based or machine learning approaches, which are vulnerable to sparsity of evidence in the highly contextual chat domain our data is drawn from.

4.3.2 Coping Efficacy

In this work we define efficacy specifically as it relates to coping with illness, and perform annotation on each thread in two passes. First we identify whether a person expressed any content related to their belief in the ability to effect change or well-being. This is the *efficacy identification* task. Then, for those cases where expressions of efficacy were identified, we assign a grade for positive (1), conflicted (2), or negative (3) efficacy; this is the *rating* task. Because of the nature of these groups, we assume that mentioning efficacy without an explicit marker of affect is positive. Therefore, our middle variable represents cases where explicit indicators of both positive and negative efficacy were present in the same span. Importantly, we measure both *self-efficacy*, a person's belief about their own ability to effect change, and *other-efficacy*, the encouragement or support given to others that they can effect change in their own lives.

To illustrate by example, negative efficacy often takes the form of doubt or uncertainty, in both other- and self-efficacy:

> **B:** how the hell can I say yes when I don't know how I'll be in April or even if I'll be in April

Positive efficacy, on the other hand, is often characterized by certainty, confidence, or, especially in the case of other-efficacy, advocacy for plans of action:

> **I:** It's about taking control of your life and not allowing cancer to take control of you.
> **I:** And enjoying life, too!

While the examples so far have matched positive efficacy with positive sentiment, the two annotations are not always correlated. In the case of negative efficacy, we see that the feeling is often softened with positive sentiment, while positive efficacy is often tempered or restrained by the real problems faced in difficult situations:

> **D:** just an incurable cancer eating away at my bones, it could be worse!

> **A:** I have your encouragement, which after a day like today, is priceless.
> **A:** even if my attitude is in the cellar at the moment.

4.3.3 Thread Ownership

To tie the notion of information sharing into our annotation of threads, we define thread *ownership*. This term is meant to define the user who holds the floor in a thread, sharing the most information on the topic of that thread. We assign ownership based on the user that gives the most **K1** moves in a thread. All lines and sequences in that thread are designated as "owned" by the user, as they are related to the topic they are sharing information on.

This construct allows us to not only measure amount of information shared, but also the response of other group members to that information. For instance, without this notion, it would be difficult to assign **f** moves to a particular user for calculating positive social feedback. With the notion of thread ownership we make the simplifying assumption that all **f** moves in a thread are directed at the thread owner. Conceptually, we use this as a proxy measure of the amount of social feedback an individual member receives over the course of a chat.

Annotation	Reliability
Negotiation	$\kappa = 0.70$
Thread Segmentation	$f = 0.75$
Self-Efficacy Identification	$\kappa = 0.73$
Self-Efficacy Grading	$r = 0.72$
Other-Efficacy Identification	$\kappa = 0.77$
Other-Efficacy Grading	$r = 0.75$
Sentiment	$r = 0.68$

Table 2: Inter-annotator agreement across multiple layers of annotation.

4.4 Inter-annotator Reliability

Fundamentally, the analysis we present here relies on hand-coded data. Our annotations span multiple levels and we ensured that each level was reliable. Therefore, for each annotation scheme, we iteratively developed a coding manual while testing the inter-rater reliability between two annotators. During these evaluations, neither annotator was aware of the others' annotations. Once we established a coding manual that achieved high reliability, all data for a particular scheme was annotated by a single annotator. For efficacy, because many threads contain no efficacy indicators, each annotator completed two conversations, instead of one, to gather enough data points for a meaningful kappa between annotators.

Our annotations are not all performed at the same level and cannot be evaluated identically. For Negotiation coding, each possible label per line is independent, so we can use the standard kappa metric. For graded thread-level annotations, we wish to account for "near-miss" labels. Efficacy must first be annotated as either present or not present; for this evaluation we use kappa agreement. Sentiment and efficacy (when identified as present) both can be annotated in a range from 1-3. To measure the agreement between annotators we use correlation coefficient. For thread disentanglement within a chat, the standard evaluation metric is the micro-averaged f-score [10], and we use the measures from prior work to compute this f-score. All evaluations of agreement are presented in Table 2 and all are high.

5. PREDICTION EXPERIMENTS

The previous section defined two levels of annotations. The first (the Negotiation framework) is primarily linguistic in nature, and is annotated on a turn-by-turn level. The second (sentiment and efficacy) is primarily social and based on a holistic evaluation of behavior, rather than particular linguistic indicators. In this section, we turn to the question of how these dimensions interact with self-reported stress levels.

We wish to test two aspects of stress prediction. First, how well can a model discern a person's overall stress level, based on their behavior in a chat? Second, can a model detect the impact that the chat will have on stress level? What elements of chat behavior result in lowered stress self-reporting on exit? This value ranges from 0 to 10, though in the sampling of the data set used in this work, no user recorded a value higher than 8.

5.1 Prediction with Thread-Level Annotations

To refine our understanding of social cognitive processing, coping efficacy, and their relation, we test the amount of variance explained with our thread level annotations. We first calculate the average self-efficacy, average other-efficacy, and average sentiment of each participant. For a given speaker S, we calculate their sentiment, self-, and other-efficacy scores $Eff(S)$ as the weighted average over the threads in which they expressed a non-zero score.

We evaluate the impact of our measurements using two metrics. The first is the amount of variance in stress (either entrance or shift) that is explained by this variable in a regression. For this we measure both the impact of a feature alone (a regression with one factor) and the impact of incrementally adding that feature to a multiple regression. In each table listed below, we list features in the order they were added to this multiple regression, so that we can list a third value, showing the improvement in the model's r^2 by adding that feature. The second measurement that we use to evaluate a model is quadratic loss, also known as mean squared error, based on the model's fitted value for a data point and that data point's actual value.

$$Eff(S) = \sum_{t \in T} \frac{\#\ S\ lines\ in\ t}{\#\ S\ lines\ total} \times Eff(S,t) \qquad (2)$$

This provides three variables representing the sentiment, self-efficacy, and other-efficacy of each speaker. We demonstrate their explanatory power in two models: entrance stress and shift in stress over the course of a conversation. For these analyses, we exclude speakers who were facilitators; we also do not include speakers who did not express any positive or negative self- or other-efficacy over the course of an entire conversation. This results in $n = 68$ total data points across 21 conversations.

First, we perform a multiple regression predicting the entrance stress level of users. The results of this regression are given in table 3. This essentially means we are measuring the impact that incoming stress has on the behavior that is subsequently observed. We see that sentiment is predictive of entrance stress levels, while efficacy measures have essentially no impact. This is consistent with the commonly-assumed static view of sentiment.

Annotation Variable	Variance Explained **Alone**	**Cum.**	**Gain**	**MSE**
Sentiment	.2763	.2763	-	2.436
Self-Efficacy	.0517	.2774	.0011	2.432
Other-Efficacy	.0000	.2886	.0112	2.394

Table 3: Variance in entrance stress explained by each thread-level measure of language affect.

Next, we perform a multiple regression predicting exit stress. Entrance stress is included as a variable in the regression, essentially meaning that our variables are predicting the residual impact of the conversation. The results of this multiple regression are shown in table 4. This regression indicates that entrance stress is the most important factor in determining exit stress.

The key finding from this regression is that while sentiment is useful in predicting the incoming stress of a user, the predictive power of sentiment analysis for shifts in stress due to chat is then limited. Instead, indicators of efficacy show the strongest predictiveness. Between other- and self-efficacy, over 7% of variance is explained over the baseline. Interestingly, self-efficacy and other-efficacy are not significantly correlated. This suggests that there are distinct motivations behind expressing high coping self-efficacy and encouraging high efficacy among others.

Annotation Variable	Variance Explained **Alone**	**Cum.**	**Gain**	**MSE**
Entrance Stress	.5101	.5101	-	1.166
Self-Efficacy	.1193	.5454	.0353	1.082
Other-Efficacy	.0479	.5836	.0382	0.991
Sentiment	.1955	.5854	.0018	0.987

Table 4: Variance in residual exit stress explained by each measure, indicating the impact of efficacy on coping during a conversation.

5.2 Prediction from Language Behaviors

At a high level, we have now shown the impact of sentiment and efficacy behaviors. Sentiment is highly correlated with entrance stress levels, but is not predictive of shift in stress over the course of a conversation. Contrastingly, entrance stress has no bearing on efficacy, but the levels of efficacy expressed in a conversation indicate the degree of stress reduction upon exiting. We now reanalyze those behaviors through the lens of language behaviors at the turn level. Again, we study both entrance stress and shift in stress. Our goal is to test whether convergent evidence for these behaviors exists on the level of information sharing.

We perform a machine learning regression experiment. We developed multiple indicators of linguistic behavior and use them as a feature space to predict the entrance stress of participants. Feature selection and regression were performed using the LightSIDE machine learning toolkit [30]. From our large set of linguistic features, we perform correlation-based feature subset selection [16] to find a small number of predictive, uncorrelated features. Linear regression is performed using the M5P algorithm [51]. Experiments were performed using leave-one-out cross-validation, where for each fold a regression model is trained on 20 conversations and evaluated on the final, held-out conversation.

To evaluate performance of our model, rather than discussing specific weights on the final regression model, we instead show coverage of different features across folds. Because each fold is trained separately, feature selection will give different results. In the tables below we list only the features that occurred in at least 4 of the 21 folds. Again, we first attempt to predict entrance stress; we then attempt to predict the residual shift in exit stress after taking entrance stress into account. For the first task of predicting entrance stress, the most frequently selected features are shown in table 5.

Our model for predicting change in stress must be conditioned on the entrance stress of a speaker. Therefore, to perform machine learning on shift, we attempt to predict the residual change in stress after performing a regression based on entrance stress. Because users with an entrance stress level of 0 essentially always remain at 0, we exclude those cases from our training set. This results in a total of

Feature	Folds	High Stress
% K2 Lines	20	Fewer K2 lines
# Owned Sequences	20	More ownership
Authoritativeness	4	Less authority
# Seqs with **f** Responses	4	More feedback
Average r^2 (cross-validation): 0.217		

Table 5: Variance in entrance stress explained by utterance-level language annotations.

Feature	Folds	Stress Reduction
# Owned Sequences	20	More ownership
% K2 Lines	11	Fewer K2 lines
Authoritativeness	10	Less authority
% ch lines	9	Fewer ch lines
Average r^2 (cross-validation): 0.138		

Table 6: Variance in exit stress residual explained by utterance-level language annotations.

$n = 49$ users for this experiment. The results of this machine learning experiment are given in table 6.

The results of these models are in line with the higher-level constructs from our first set of experiments. Language behaviors account for over 20% of variance in entrance stress, similar in scale to the performance of sentiment annotation. Meanwhile, after accounting for entrance stress, language behaviors as we have represented them account for an additional 13.8% of variance in the residual shift.

Qualitatively, we find that the features which are used by the model align with our findings from section 5.1. The most important indicators of an individual's stress are related to information sharing. In most folds, the feature indicating the percent of K2 moves, along with sequence ownership, are the strongest (or only) predictors. Users with high incoming stress tend to request less information from others, as a percentage of their time, and share much more information, in absolute terms. Both insights are valuable. The first suggests that users experiencing high stress do not take on the facilitative, conversation-leading role that would be indicated by a high percentage of K2 moves. Instead, their contributions are more focused on either sharing their own information, or taking on a less engaged role in the discussion, primarily providing feedback to others. Next, the inclusion of ownership rather than K1 count alone suggests the group process behavior of "rallying around" stressed individuals. On a further group process level, we see empirical evidence of feedback moves being used more frequently for stressed individuals. More social feedback from others is indicative that the user came in with a high stress level.

The Authoritativeness feature plays a curious role, as it at first glance is weighted counter to our hypotheses on information sharing. However, this is a peculiarity in the regression - when authoritativeness is selected as a feature, the initial intercept of the regression is much higher, indicating very high stress; but the weighting for the % K2 lines feature is also higher. This suggests two alternative models. In one, predicted stress is assumed to be very high, and is drastically lowered by behaviors eliciting information from others. In the other, predicted stress is affected primarily through thread ownership, which indicates information sharing, and the impact of information requests is less drastic. Both models fit the social cognitive processing hypotheses emphasizing information sharing.

We observe similar features selected by the stress reduction model compared to the entrance model. This is expected, as users entering with high stress inherently have more room for stress reduction. The two features used most frequently for predicting entrance stress continue to appear in this context. The weighting, however, tends to give less weight towards information requests and more towards ownership. An interpretation of this model is that while speakers with high entry stress do not tend to ask for information from others, it is not this lack of questioning that leads to stress reduction. Rather, the lack of questioning is an artifact of the high entry stress, and it is the process of sharing information that is most highly indicative of stress reduction.

Social feedback from other speakers is not included in the stress reduction model, while it is occasionally included in the entry stress prediction model. This suggests that while users may respond strongly to negative sentiment and high entrance stress by giving support, it is less necessary to have explicit support, and more important merely to have an environment conducive to sharing and disclosure. In contrast, a factor that appears in the stress reduction model that does not appear in entrance stress involves **ch** moves. These moves are rare in our corpus. When they do appear, it is usually on topics unrelated to their illness. This suggests that these support groups do avoid direct challenges in supportive contexts; such negative moves occur only outside of the context of illness, aligning with theory from social cognitive processing on reducing social constraints and allowing disclosure.

6. CONCLUSIONS

This work presents an empirical study of the emotional and social functions of online support group chatrooms. Emotional states are a fluid attribute and do not fit well into a single numeric value, especially when discussing coping with a serious illness. To this end, this work highlights multiple views of emotional affect. We measure both self-reported entrance and exit scores, as well as annotating at a thread level language use and behavior that indicates authority, information exchange, sentiment, coping self-efficacy, and belief in the coping efficacy of others.

The picture that emerges illustrates the complexity of language. The stress of a user upon entering a chatroom can assuredly be predicted using sentiment analysis techniques on their language. However, we find that there is little added value from sentiment analysis when predicting improvement in stress over the course of a conversation.

Instead, we find utility in constructs from the social sciences related to information sharing and efficacy. In a machine learning paradigm we used regression models to demonstrate that coping efficacy is more explanatory than sentiment. We also find that information sharing, which is closely backed by the social cognitive processing model of coping, is predictive of stress reduction. In particular, after disentangling conversation into distinct threads, we find that "ownership" of a topic - sharing your own narrative or beliefs - is linked to stress reduction. These findings have important implications on the future direction of automated analysis of text, where sentiment classification continues to improve in performance at breakneck speed, with less attention being paid to more complex models of human emotion.

6.1 Future Work

These experiments were conducted on 21 conversations and nearly 10,000 lines of chat, one of the larger corpora to have ever been analyzed at this granularity. However, this is a fraction of the data available to us. Thousands of conversations within the Cancer Support Community have not been annotated or analyzed. Furthermore, our most recent work has shown that the information exchange annotations and thread structure within this dataset can be reliably annotated [27]. Automation of thread-level sentiment or efficacy annotation presents a promising direction for future work; results in related areas such as email conversations [15] and dialogue systems [9] have been effective. Thus it is likely that an ensemble of methods will allow us to annotate our entire corpus along all dimensions described in this paper.

This large-scale annotation will allow us to study many questions about the interactions of users in online support groups. Longitudinal studies of support groups have been performed in the context of discussion boards, e.g. for language adoption [34] and membership attrition [52], but studying the change in language use in synchronous communication is much sparser. In this work we have also limited ourselves to the text contributions of members; however, we have valuable information in the form of social network data, and work in social network analysis between group members may lend further insight into the group practices which are most effective for stress reduction. Automatic annotation of very large data sources, augmented by the insights gained in this paper and proven methods for group interaction analysis, will lead to a much deeper understanding of support group communication habits.

Acknowledgements

The research reported here was supported by National Science Foundation grant IIS-0968485.

7. REFERENCES

[1] Cancer support groups: A guide for facilitators. *American Cancer Society*, 2005.

[2] P. M. Aoki, M. H. Szymanski, L. Plurkowski, J. D. Thornton, A. Woodruff, and W. Yi. Where is the "party" in "multi-party"? analyzing the structure of small-group sociable talk. In *Proceedings of the Internation Conference on Computer Supported Collaborative Work*, 2006.

[3] A. Bandura. *Self-Efficacy: The Exercise of Control*. 1997.

[4] A. Barak, M. Boniel-Nissim, and J. Suler. Fostering empowerment in online support groups. In *Computers in Human Behavior*, 2008.

[5] E. Bingham, A. Kaban, and M. Girolami. Topic identification in dynamical text by complexity pursuit. In *Neural Processing Letters*, 2003.

[6] S. D. Black, J. A. Levin, H. Mehan, and C. N. Quinn. Real and non-real time interaction: unraveling multiple threads of discourse. In *Discourse Process*, 1983.

[7] W. C. Booth. *The Rhetoric of Fiction*. 1983.

[8] H. I. Daumé. Frustratingly easy domain adaptation. In *Proceedings of the Association for Computational Linguistics*, 2007.

[9] J. Drummond and D. Litman. Examining the impacts of dialogue content and system automation on affect models in a spoken tutorial dialogue system. In *Proceedings of the ACL Special Interest Group on Discourse and Dialogue*, 2011.

[10] M. Elsner and E. Charniak. Disentangling chat. *Computational Linguistics*, 2010.

[11] W. Geyer, H. Richter, L. Fuchs, T. Frauenhofer, S. Daijavad, and S. Poltrock. A team collaboration space supporting capture and access of virtual meetings. In *Proceedings of the International Conference on Supporting Group Work*, 2001.

[12] J. Giese-Davis, A. Conrad, B. Nouriani, and D. Spiegel. Exploring emotion-regulation and autonomic physiology in metastatic breast cancer patients: Repression, suppression, and restraint of hostility. *Personality and Individual Differences*, 2008.

[13] J. Godfrey, E. Holliman, and J. McDaniel. Switchboard: telephone speech corpus for research and development. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 1992.

[14] S. P. Goggins, K. Galyen, and J. Laffey. Network analysis of trace data for the support of group work: Activity patterns in a completely online course. In *Proceedings of the ACM Conference on Supporting Group Work*, 2010.

[15] G. Groh and J. Hauffa. Characterizing social relations via nlp-based sentiment analysis. In *Proceedings of the International Conference on Weblogs and Social Media*, 2011.

[16] M. Hall and L. A. Smith. Practical feature subset selection for machine learning. In *Australian Computer Science Conference*, 1998.

[17] I. Howley, D. Adamson, G. Dyke, E. Mayfield, J. Beuth, and C. P. Rosé. Group composition and intelligent dialogue tutors for impacting students academic self-efficacy. In *Proceedings of Intelligent Tutoring Systems*, 2012.

[18] I. Howley, E. Mayfield, and C. P. Rosé. Missing something? authority in collaborative learning. In *Proceedings of Computer Supported Collaborative Learning*, 2011.

[19] K. Hyland. *Disciplinary Discourses: Social Interactions in Academic Writing*. 2004.

[20] D. Inkpen, F. Keshtkar, and D. Ghazi. Analysis and generation of emotion in texts, in knowledge engineering: Principle and technique. In *Proceedings of the International Conference on Knowledge Engineering Principles and Techniques*, 2000.

[21] M. Joshi and C. P. Rosé. Generalizing dependency features for opinion mining. In *Proceedings of the Association for Computational Linguistics*, 2009.

[22] J. Kruger, N. Epley, J. Parker, and Z. Ng. Egocentrism over e-mail: Can we communicate as well as we think? In *Journal of Personal and Social Psychology*, 2005.

[23] R. Kumar and C. P. Rosé. Architecture for building conversational agents that support collaborative learning. In *IEEE Transactions on Learning*, 2011.

[24] R. Kumar, C. P. Rosé, Y.-C. Wang, M. Joshi, and A. Robinson. Tutorial dialogue as adaptive collaborative learning support. In *Proceedings of Artificial Intelligence in Education*, 2007.

[25] S. J. Lepore. A social-cognitive processing model of emotional adjustment to cancer. In *Psychosocial intervention and cancer: An introduction*, 2001.

[26] J. Martin and D. Rose. *Working with Discourse: Meaning Beyond the Clause*. Continuum, 2003.

[27] E. Mayfield, D. Adamson, and C. P. Rosé. Hierarchical conversation structure prediction in multi-party chat. In *Proceedings of SIGDIAL Meeting on Discourse and Dialogue*, 2012.

[28] E. Mayfield, D. Adamson, A. I. Rudnicky, and C. P. Rosé. Computational representations of discourse practices across populations in task-based dialogue. In *Proceedings of the International Conference on Intercultural Collaboration*, 2012.

[29] E. Mayfield, M. Garbus, D. Adamson, and C. P. Rosé. Data-driven interaction patterns: Authority and information sharing in dialogue. In *Proceedings of AAAI Fall Symposium on Building Common Ground with Intelligent Agents*, 2011.

[30] E. Mayfield and C. P. Rosé. An interactive tool for supporting error analysis for text mining. In *NAACL Demonstration Session*, 2010.

[31] E. Mayfield and C. P. Rosé. Recognizing authority in dialogue with an integer linear programming constrained model. In *Proceedings of Association for Computational Linguistics*, 2011.

[32] J. McIntire, O. I. Osesina, and M. Craft. Development of visualizations for social network analysis of chatroom text. In *Proceedings of the International Conference on Collaboration Technologies and Systems*, 2011.

[33] D. Nguyen, E. Mayfield, and C. P. Rosé. An analysis of perspective in interactive settings. In *Workshop on Social Media Analysis at the ACM Conference on Knowledge Discovery & Data Mining*, 2010.

[34] D. Nguyen and C. P. Rosé. Language use as a reflection of socialization in online communities. In *Workshop on Language in Social Media at ACL*, 2011.

[35] K. Ogura, T. Kusumi, and A. Miura. Analysis of community development using chat logs: A virtual support group of cancer patients. In *Proceedings of the IEEE Symposium on Universal Communication*, 2008.

[36] J. E. Owen, E. O. Bantum, and M. Golant. Benefits and challenges experienced by professional facilitators of online support groups for cancer survivors. In *Psycho-Oncology*, 2008.

[37] E. M. Ozer and A. Bandura. Mechanisms governing empowerment effects: A self-efficacy analysis. In *Journal of Personality and Social Psychology*, 1990.

[38] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the Association for Computational Linguistics*, 2004.

[39] S. Park, M. Ko, J. Kim, Y. Liu, and J. Song. The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns. In *Proceedings of the International Conference on Computer-Supported Collaborative Work*, 2011.

[40] S. Park, K. S. Lee, and J. Song. Contrasting opposing views of news articles on contentious issues. In *Proceedings of the Association for Computational Linguistics*, 2011.

[41] S. Rodgers and Q. Chen. Internet community group participation: Psychosocial benefits for women with breast cancer. In *Journal of Computer Mediated Communication*, 2005.

[42] D. Rosen, V. Miagkikh, and D. Suthers. Social and semantic network analysis of chat logs. In *In Proceedings of the International Conference on Learning Analytics and Knowledge*, 2011.

[43] E. Schegloff. *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press, 2007.

[44] D. Shen, Q. Yang, J.-T. Sun, and Z. Chen. Thread detection in dynamic text message streams. In *Proceedings of SIGIR*, 2006.

[45] M. Smith, J. J. Cadiz, and B. Burkhalter. Conversation trees and threaded chats. In *Proceedings of the International Conference on Computer-Supported Collaborative Work*, 2000.

[46] G. Stahl. How a virtual math team structured its problem solving. In *In Proceedings of the Conference on Computer-Supported Collaborative Learning*, 2011.

[47] G. Stahl, J. X. Ou, M. Cakir, S. Weimar, and S. Goggins. Multi-user support for virtual geogebra teams. In *Proceedings of the North American Conference on Geogebra*, 2010.

[48] D. Tannen. *Framing in Discourse*. 1993.

[49] M. Tixier, G. Gaglio, and M. Lewkowicz. Translating social support practices into online services for family caregivers. In *Proceedings of the ACM Conference on Supporting Group Work*, 2009.

[50] J. Walther and S. Boyd. Attraction to computer-mediated social support. 2002.

[51] Y. Wang and I. H. Witten. Induction of model trees for predicting continuous classes. In *Proceedings of the European Conference on Machine Learning*, 1997.

[52] Y.-C. Wang, R. Kraut, and J. M. Levine. To stay or leave? the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the International Conference on Computer-Supported Collaborative Work*, 2012.

[53] J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin. Learning subjective language. In *Computational Linguistics*, 2004.