Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2012 Proceedings

Proceedings

Certainty Affect Detection in Informal Text

Jeffrey Kim
Information School, University of Washington, Seattle, WA, United States., jykim@uw.edu

Sonia Savelli
Information School, University of Washington, Seattle, WA, United States., ssavelli@uw.edu

Caroline Dombrowski
Information School, University of Washington, Seattle, WA, United States., cd1@uw.edu

Ali Fathalian
Information School, University of Washington, Seattle, WA, United States., fatalian@uw.edu

Follow this and additional works at: http://aisel.aisnet.org/amcis2012

Certainty Affect Detection in Informal Text

Jeffrey Kim
University of Washington
jykim@uw.edu

Sonia Savelli
University of Washington
ssavelli@uw.edu

Caroline Dombrowksi
University of Washington
cd1@uw.edu

Ali Fathalian
University of Washington
fatalian@uw.edu

ABSTRACT

Uncertain and varied contexts have been shown to change individual skills and perceptions, particularly of leaders (Rast, 2012; Hutchison, Jetten & Gutierrez, 2011; Smith, Hogg, Martin & Terry, 2007). For this reason, we investigate the certain and uncertain affect of active players in a voluntary, virtual organization. Uncertainty is a constant in human interaction, and we investigate active players' text for patterns of certainty and uncertainty affect that can assist in understanding social interactions in online communities. In this forum data, the most active players expressed significantly less uncertain or mixed certainty affect. We discuss how certainty and uncertainty affect detection provides a window into teaming and community dynamics.

Keywords

Affect detection, social influence, certainty, uncertainty, virtual teams, informal language

INTRODUCTION

In a networked world, virtual organizations have different characteristics and patterns of affect expression than traditional organizations. In addition, volunteer-driven teams coalesce around particular problems, and dissipate when the problem is resolved—for instance, creating a particular Wikipedia article together, during crowdsourcing (i.e., Dubach, Muhdi, Stöcklin & Michahelles, 2011) or the DARPA Red Balloon challenge (Smith, 2010). All three kinds of uncertainty (environmental, response and cause-effect [Millikin, 1987]) are present in voluntary virtual teams, because the ground rules are emergent and the goals ever-shifting. Affect computing can offer a window into the machinations and processes of these virtual organizations, or collaborative networks, particularly around motivation and participation (Luneski & Moore, 2008). However, at present there exists no substantiated, well-validated dictionary for automated analysis of certainty and uncertainty in informal text. This research advances us a step closer to developing an informal language dictionary that can be used for text analysis software programs like Linguistic Inquiry and Word Count (LIWC) (Tausczik & Pennebaker, 2010) or other automated software.

In ad hoc, virtual teams, communication becomes the critical component for action, even as social patterns emerge in the midst of that communication. This paper considers the influence of certainty and uncertainty in a voluntary, virtual organization: a digital game forum. Text and word choices have been linked to authority, social status and interaction (Gilbert, 2012; Hosman & Siltanen, 2011). Because these word choices are indicative of affect and feeling, we consider communication styles of participants in a deliberately uncertain, ambiguous situation as is present in digital game forums.

Uncertainty has also been linked to conformity and lack of conformity to a group's social norms (Smith, Hogg, Martin & Terry, 2007), suggesting that influential individuals may have unique certainty or uncertainty expression patterns. Conveying certainty and uncertainty alters perception of information, and we theorize that especially certain and uncertain individuals will be important in the organization. In uncertain conditions, it has been shown that people are influenced by presentation of certain or uncertain information, which changes the amount of attention given to information (Mills, 1965). When individuals are highly certain of information and have high attitude certainty, they are also less likely to seek feedback or doubt themselves (Clarkson, Tormala, DeSensi & Wheeler 2009). Conversely, when individuals are highly uncertain they tend to identify even with low status groups (Grieve & Hogg, 1999; Mullin & Hogg, 1998; Reid & Hogg, 2005), which may help to explain zealotry or radicalism (Hogg et al., 2010).

Group leaders' social ties and their network positions in communication have been a subject of many studies (Mehra et al., 2006). In an analysis of the personal friendship ties of group leaders in the sales division of a financial services firm, Mehra et al. (2006) found that social networks within the group were related to group performance and suggested that informal social ties of leaders would enhance their group performance. Johnson et al. (2002) studied the network structure of scientists in Antarctic research stations and found that informal factors in scientists' network dynamics had effects on the patterns of interaction as well as group's performance, productivity and morale. In informal organizations without hierarchies, activity level (number of posts) is one way to identify leaders (ala Huffaker, 2010). In this paper, we use activity level as the main indicator of leadership and augment that with an analysis of the betweenness/centrality and authority from social network analysis to investigate the connection between certainty and uncertainty affect and activity level in the game.

LITERATURE REVIEW

Voluntary Virtual Teams

Wide adoption of Information and Communications Technology (ICT) applications made collaboration across geographical and organizational boundaries faster and cheaper. Collaboration and group interaction are reaching into social media space, and in some online communities, spontaneous or voluntary collaboration is expected for short, task-focused projects. Individuals participate in activities that will serve perceived shared goals without explicitly written guidelines about how to participate. With aggregated information and emergent team dynamics, increases in collective intelligence can take place, giving teams advantages in knowledge (Bettencourt, 2009). ICT applications made this type of collaboration possible, but do not necessarily make distributed collaboration easy to establish and manage (Berry, 2011).

Virtual teams suffer from lack of trust, difficulty defining tasks and variations in experience levels (Munkvold & Zigurs, 2007). Levels of trust and diversity of background are known to influence social processes even in short-term, one-off virtual organizations (Paul & He, 2011). Complicating interaction among geographically dispersed teams, diurnal, temporal effects are known to change virtual affective states and expression of those states (Golder & Macy, 2011). The affective social processes of the group are often only available in text form, which provides a key analytic tool for researchers, but also makes it harder for team members to detect, handle, and resolve affective problems like frustration, mismatch of goals and relational conflict (i.e., Hinds & Bailey, 2003).

Multiple researchers have considered the correlation between betweenness/centrality of a person in a social network and the degree that he is perceived as a leader or an emergent leader in both formal and informal communities (Balkundi & Kilduff, 2006 ; Mullen & Salas, 1991; Brass, 1984; Lawrence & Lorsch, 1967; Freeman, Roeder & Mullholland, 1979). Freeman et al. concluded that "betweenness, then, seems to be the key to understanding choice as leader." Generally research has indicated that an individual with high betweenness plays a coordination role within a community because he/she links distant parts of a community together and brokers the flow of often crucial information.

Wasserman and Faust (1994) review existing research on the relationship between the authority role in a social network and the influence on dissemination of information with focus in the context of citation networks. They conclude that the information distributed or created from an individual with an authority role, gets disseminated with a higher rate and more importantly uses a substantially more effective dissemination path. More recently, Marlow (2004) has studied the authority role in the weblog community stating that people with a higher structural authority degree are more influential and followed in that community.

Subgroups are inevitable when knowledge, geography and time are all differently shared among team members, and so the individuals who span those groups, conveying information across boundaries, are important (i.e., Ahmad & Lutters, 2011). Barley (1986) found that organizational members' actions follow structured patterns, like behavioral grammars. Through the process of individuals' recognizing, accepting and internalizing organizational structure (e.g. policies, implicit rules), they develop and follow organizational routines. The strength of ties and the number of cooperative roles in a community also affect knowledge production (Kim & Yang, 2011). These voluntary virtual organizations also may be composed of unique participants—people who seek, contribute to and relish ambiguous and developing situations. Voluntary virtual organizations may have different kinds of individuals who participate, and thus those individuals may be interesting to profile. This paper characterizes some of those individuals and lays the groundwork for later analysis of different kinds of leaders.

Uncertainty affect

How do virtual team members manage ambiguity caused by the very promise of virtual teaming? Bechky (2006) observed that roles are enacted in particular situations and in the absence of formal rules and permanent organizational structures, role structure and negotiation create the order needed for work coordination. Much psychological research has focused on how

certainty (or confidence) and uncertainty influence social interactions. That research has taken into account verbal, postural, gestural and delivery cues. Experimental studies have found that linguistic delivery and particular phrases influence perceptions of information (e.g., Wesson & Pulford, 2009). Furthermore, a review of the psychological literature concluded that individuals are very consistent in their personal use of uncertainty expressions (Clark, 1990) suggesting that dimensions of leadership, influence and persuasion are related to certainty and uncertainty affect.

However, many social interactions now take place online, solely mediated by text. Can we take the lessons from psychological research, detect certainty and uncertainty in written text and then assess the characteristics of those individuals? Luneski and Moore (2008) suggest that considering emotions as part of collaborative networks is key to understanding their dynamics. In addition, uncertainty in text has been analyzed to detect expertise and the process of argumentation (Campbell, 1985).

Current research often focuses on social influence, or the ability to cause responses and references to online content. For instance, Huffaker (2010) reveals how being active over time and posting frequently, both original content and replies to other people's content, is a key part of social influence. In addition, according to Tauszik & Pennebaker (2010), highly active players who post frequently have a kind of dominance. While higher levels of activity do not necessarily indicate leadership, in this paper we make the assumption that the most frequent posters are a kind of leader—a thought leader, even if not a catalyst for group formation. For these reasons, in the analysis below, we specifically consider the players who posted the most.

We also consider whether the most active posters, who shape the content of the conversation, are prototypical or unusual leaders (ala Knippenberg, 2011; Rast, Gaffney, Hogg & Crisp, 2012). Given that the game is a situation of high uncertainty that may produce atypical leaders (Rast et al., 2012). Highly certain players may also be more likely to remain certain, if personality traits are responsible for their certainty even in a very ambiguous game context. Repeated attitude certainty— posters who frequently post in a consistently certain way—may be less prone to doubt themselves and more likely to be certain all the time (Clarkson, et al., 2009). Certainty affect has potential for understanding individual characteristics and group dynamics.

In the tradition of Levinson (1983) and Michael (1994), we see uncertainty as a social construct, not just an individual cognitive construct or motivating force. Uncertainty is part of normal social discourse emerging from the process of negotiation. This conceptualization of the certainty and uncertainty affect follows the "emotions as social constructs model" as laid out in Calvo & D'Mello (2010). This paper explores how certainty and uncertainty affect of leaders within a virtual organization differ from non-leaders and suggest ways in which this style of communication may be important to the dynamics of the group.

Research Questions

We suggest that certainty and uncertainty affect will differ across individuals, and that those differences are important to group dynamics. The framing question for this project was: How do people in a spontaneous organization express their opinions? Our specific research questions are as follows:

1. Can we assign a meaningful uncertainty/certainty score to informal language communications?

2. Does certainty affect differ across social roles?

3. Are the most active players/group leaders prototypical or unusual?

METHODS

This qualitative analysis focuses on creating groundwork for future automated analysis. In particular, this work seeks to provide baseline scores for certainty and uncertainty in extremely informal language (forum posts). While we believe that this research is a stepping stone for the development of an informal language dictionary to be used with text analysis software programs such as LIWC (Tausczik & Pennebaker, 2010), initial pilot tests with the existing LIWC certainty and tentative dictionaries showed very poor match to human coding of the same data. Therefore, this work focuses on using human coding of certainty and uncertainty, while future work will focus on the development of a baseline for testing and augmenting the existing LIWC dictionaries for informal language.

Dataset

Five unique threads in the "I Love Bees" (ILB) dataset were randomly selected for analysis. This discussion forum ("The Haunted Apiary" is publicly available online. The threads were on a variety of topics, ranging from identifying physical

locations where clues would be revealed to discussing the storyline of the game. These threads represented 279 unique posts created by 125 unique authors. Posts ranged from a minimum of 2 words to a maximum of 724 words (M=86.43, SE=6.00), with a minimum of 1 sentence to a maximum of 50 sentences per post (M=5.57, SE=0.98). Each of the 125 unique authors contributed anywhere from one to 14 unique posts (M=2.21, SE=0.18).

Preprocessing

The following pre-processing was conducted on each post prior to analysis:

- Signatures, extra line breaks, metadata and images were removed.
- Quotes by other players were removed from each post.
- Individual author was linked to the post, as well as the person to whom they were replying (if any).

Human Coding: The Gold Standard

Six coders worked in pairs with predefined certainty and uncertainty codes. Each pair was responsible for coding two threads (one pair only coded one thread as it contained over 26% of the posts) and working together to achieve inter-rater reliability of over 85%. In addition they were told, "not to stress out about the coding, but to go with their gut instinct, and not second-guess their codes." The definitions were as follows:

- Certain: The author is sure of the information provided.
  - *e.g., Definitely, 100%, absolutely, for sure.*
- Uncertain: The author lacks confidence about the provided information.
  - *e.g., Perhaps, maybe, not exactly, probably.*
- Mixed: The author expresses both certainty and uncertainty.

Coders used a scale from 0-5, with strong encouragement to only assign a zero when the post was truly lacking in certainty or uncertainty. See Table 1 for a breakdown of the Gold Standard and sample posts.

Gold Standard	Gold Standard Description	Sample Post Text
0	Neither certain nor uncertain	what would be a good page name for Ariel it has to fit the road/portal/way theam?
1	Uncertain	Couldn't the glass coffin just be a stasis chamber like Master Chief was in at the beginning of Halo 1?
2	Somewhat uncertain	I'm not sure this is a puzzle solve. I still think that we're missing files, especially between our hoodlums and much lifting. That could simply be a cut though. Not to mention the remaining 567 axons. That means, with 7 axons a file (an assumption, I know), that we have a story spanning 111 files. We onlt have 30 files. Should we change the tag to partial puzzle solve?
3	Neutral/Mixed (both certainty and uncertainty expressed)	What time on the 24th? I was thinking - since I have one of the locations near me - that IF we were to show up to the locations on Aug 24, what time would we show up? It is 6am LOCAL time? Texas time? Pacific time? I would guess that UNLESS we come up with another solution to the coordinates by then that visiting the sites at that time would be a good strategy. It would only cost some sleep...
4	Somewhat certain	I think the link between the Kamal and Jersey stories is the restaurant. Why yes, I am clinging tenuously to the fact that both storylines make a point of saying the restaurant in question serves tuna, but I can see Jersey asking Durga to spy on his cheating mother, and picking up on the communications between Sophia and Kamal by accident.
5	Certain	Jeanie is most certainly not a smart AI, however. SP has shown she is capable of adapting and changing, something dumb AI's can't do.

Table 1: Sample of posts and the Gold Standard assigned by human coders. Note the informal language, including misspellings, game jargon, erratic punctuation and unusual capitalization.

As we were interested in the certainty and uncertainty affect of individuals within this community, all posts with no level of certainty/uncertainty (i.e., those coded 0) were removed. This resulted in the elimination of 69 posts from the original 279

posts, as well as the elimination of 25 authors (who had posted very few times). Therefore, all subsequent analyses included 210 posts from 100 unique individuals.

Social Network Analysis

We used the Pajek social network analysis software (de Nooy, Mrvar & Batageli, 2005) to analyze all interactions of the 100 game players in the five discussion threads. In particular, we considered the following brokerage roles: coordinator, gatekeeper, representative and liaison. These roles indicate information moving between subgroups. Further, centrality, hub and authority scores were calculated as these have been identified with leadership (see Literature Review, above).

RESULTS

We measured the certainty affect for each of the 100 authors in two different ways. For the 51 authors who posted more than once, the resulting human-coded posts were aggregated by individual author and averaged to find the individual's certainty score. For those authors who only posted one time we simply used the Gold Standard of their one post. This resulted in one Gold Standard per individual. The second way we measured certainty affect was to categorize the 0 – 5 Gold Standard rating for each post into one of three categories: 1 and 2 were categorized as uncertain, 3 was categorized as mixed, and 4 and 5 were categorized as certain. Posts coded as 0 were discarded, as discussed above. For each author with more than one post, we computed the percentage uncertain, percentage mixed, and percentage certain for that author by dividing the number of posts in the uncertain category by the total number of posts for that author and similarly for mixed and uncertain posts. For authors with only one post we simply translated the Gold Standard for that post into one of the three categories above. Again, each author had one Uncertain Percentage, one Mixed Percentage, and/or one Certain Percentage.

Research Question 1: Can we assign a meaningful uncertainty/certainty score to informal language communications?

Coders were easily able to reach the desired inter-rater reliability of 85%. As is typical in qualitative coding, low initial levels of disagreement were quickly resolved through discussion of code definitions and iterative coding. With posts where coders did not agree, one of the codes from the coding pair was randomly assigned. See Table 2 for initial and final inter-rater reliability agreement, calculated as number of posts coded identically by the coding pair divided by total number of posts.

The ease with which the six human coders, three of whom had little prior experience with coding for certain/uncertain affect, arrived at an inter-rater reliability of 95% (see Table 2) suggests that humans are able to easily detect certainty and uncertainty in online informal text. In general, the disagreement was at the level of one point on the scale away, rather than the difference between a zero and a three, or a one and a four. Calibrating through discussion and iterative coding easily resolved this, suggesting that like Armstrong et al. (1997), coders found the same things in the text.

Thread Name	Initial Inter-rater Reliability	Final Inter-rater Reliability
Behind the Kitchen	56%	94%
Military	54%	96%
What are Axons	27%	93%
Sleeping Princess	28%	95%
WAV Assembly	56%	95%
Average Inter-Rater Reliability	**44.2%**	**94.6%**

Table 2: Initial and Final Inter-rater Reliability for all 5 Threads, calculated as simple percentage of agreement (i.e., # of posts in agreement / total # of posts).

Research Question 2: Does certainty affect differ across social roles?

Of the 100 authors, 14 posted four or more times compared with the other 86 authors, who posted 3 times or less. These 14 authors represent 35% of the posts in our dataset (74 out of 210 posts). These Top 14 most active authors were significantly higher in authority (M=.14, SE=.06) than the least active authors (M=.02, SE=.004), $t(13)=2.21$, $p=.045$. With respect to betweenness/centrality, these Top 14 most active posters had a significantly higher vector score (M=.03, SE-.008) than the least active authors (M=.003, SE=.0009), $t(13)=3.90$, $p=.002$ (see Figure 1 for a Pajek Social Network Analysis of betweenness/centrality for all 100 authors). There were no significant differences with respect to the Hub Vector.
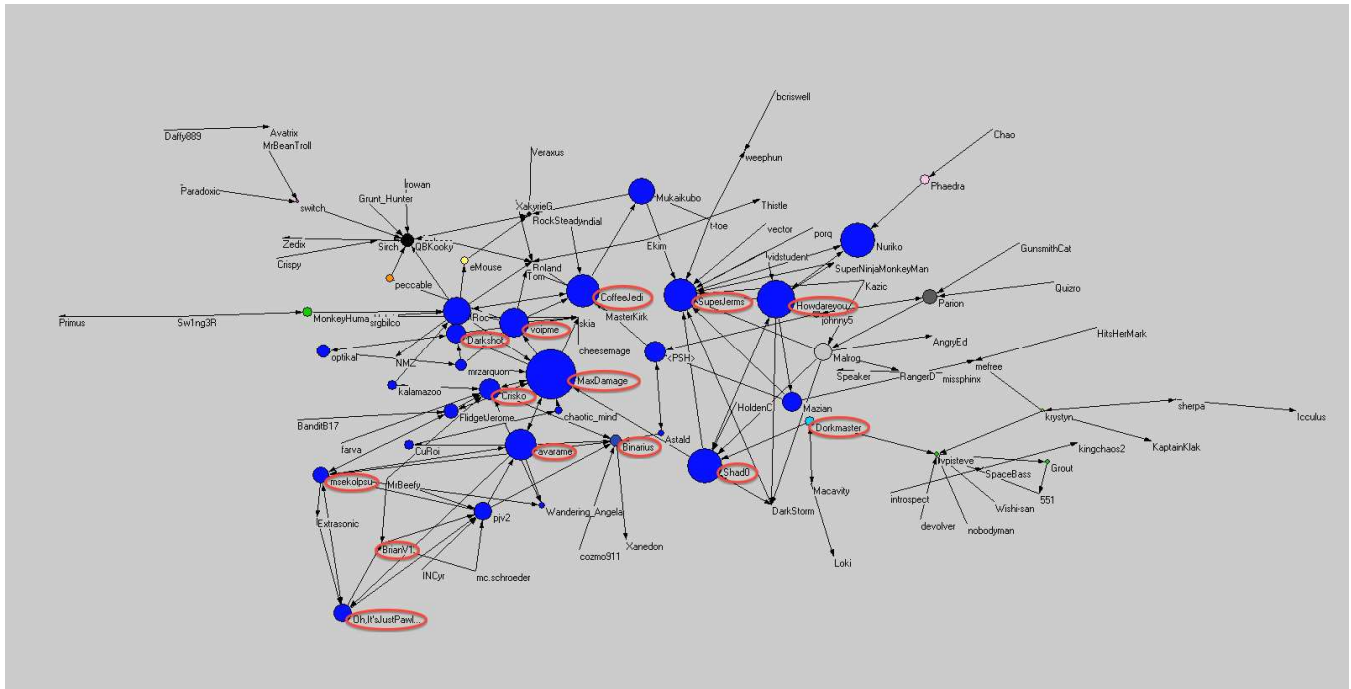
Figure 1: Social network analysis of betweenness for all authors in the five discussion threads. The red boxes indicate players in the Top 14, the most active players. A larger circle indicates a larger betweenness/centrality vector.

Furthermore, of the Top 14 authors, all authors except BrianV1 held at least one brokerage role (Gatekeeper, Representative, Coordinator, or Liaison) in the network (see Table 3). These results suggest that the most active authors may indeed be leaders in this virtual teaming environment.

Author Name	# of Posts	Betweenness/ Centrality Vector	Hub Vector	Authority Vector	Brokerage Role (Yes/No)
msekolpsu	10	0.011538	0.084216	0.081603	Yes
Howdareyou	8	0.061660	0.428439	0.199459	Yes
CoffeeJedi	6	0.049104	0.019995	0.025778	Yes
Max Damage	6	0.104885	0.058925	0.207746	Yes
SuperJerms	6	0.046770	0.027885	0.793614	Yes
avarame	5	0.043881	0.134310	0.047965	Yes
Crisko	5	0.019601	0.099345	0.053869	Yes
Binarius	4	0.005041	0.000000	0.106025	Yes
BrianV1	4	0.000000	0.004058	0.014961	No
Darkshot	4	0.017566	0.006552	0.052152	Yes
Dorkmaster	4	0.004014	0.083570	0.000000	Yes
Oh, It's Just Pawl...	4	0.013552	0.034050	0.055695	Yes
Shad0	4	0.053351	0.354520	0.290595	Yes
voipme	4	0.038402	0.019425	0.030911	Yes
Mean	5.27	0.033526	0.096806	0.140027	
Std. Dev.	1.75	0.029040	0.131931	0.206337	

Table 3: Pajek Social Network Analysis Results for the Top 14 most active authors. The mean and standard deviation in the final two rows are only for this group of 14 authors.

Research Question 3: Are the most active players/group leaders prototypical or unusual?

While the community as a whole tended toward certainty, individuals, with the exception of MaxDamage and Crisko who were always certain and Voipme who was always uncertain, tended to post in several categories, and that included the top posters (see Figure 2). Overall, authors were more certain (54%) than uncertain (32%) or mixed certainty (14%).
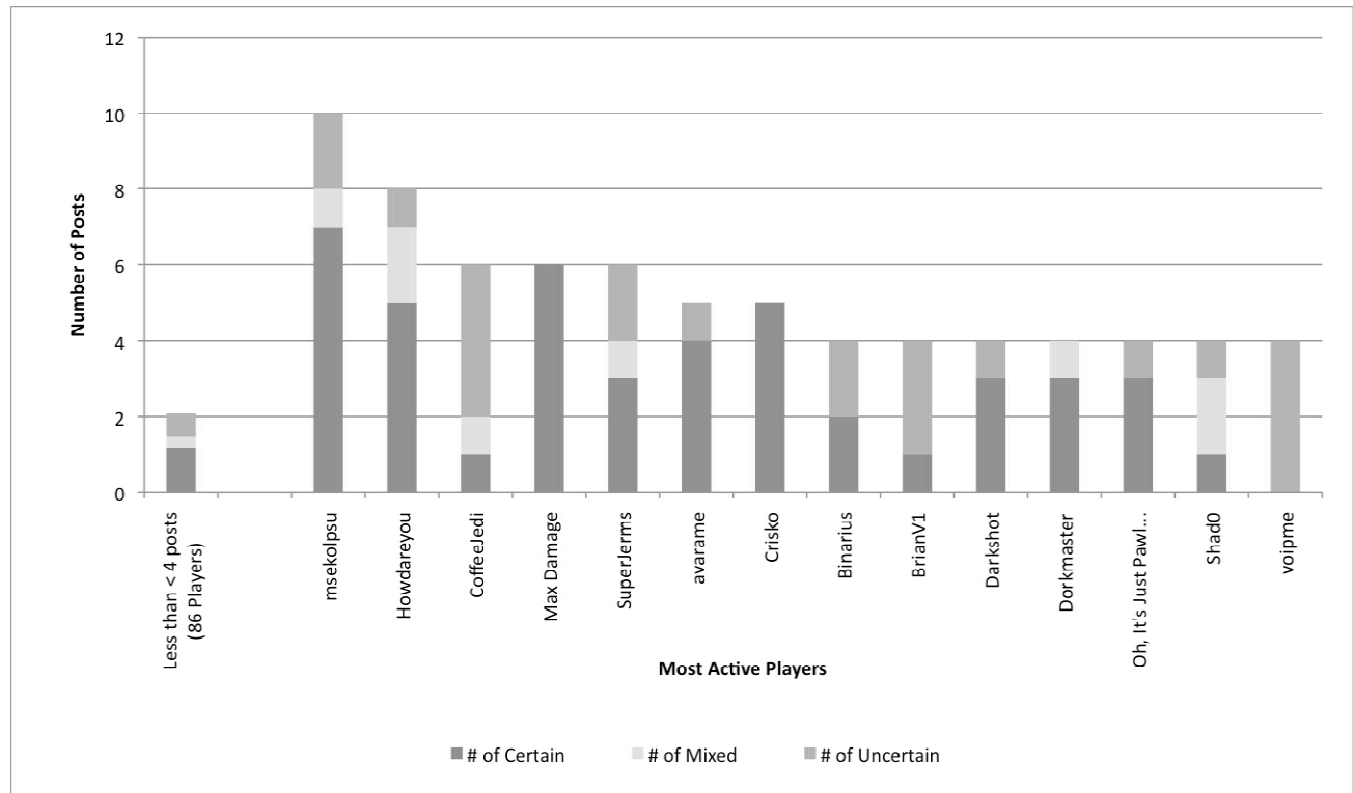

Figure 2: Mean number of certain, mixed and uncertain posts for players with less than 4 posts and total certain, mixed and uncertain posts for the Top 14 Most Active Players

Although there were no significant differences between the Mean Gold Standard of Top 14 most active authors (M=3.45, SE=.23) and the least active authors (M=3.37, SE=.14), $t(24)$=.29, p=.776, there was a significant difference in the percent of uncertain and mixed posts. Specifically, the Top 14 authors posted significantly fewer percentage of uncertain posts (M=12%, SE=3%) than did the least active authors (M=46%, SE=8%), $t(98)$=-3.82, p<.001 and they posted significantly fewer mixed posts (M=3%, SE=2%) compared with the least active authors (M=23%, SE=5%), $t(98)$=-3.84, p<.001. There was no significant difference between the percent of certain posts for the Top 14 compared with all others. These results suggest that the entire community uses a similar degree of certain affect in their written text, but the leaders use less uncertain and mixed affect than non-leaders.

DISCUSSION

Overall, the leaders, defined as the most active contributors were significantly less likely to post mixed or uncertain posts and had significantly higher authority and betweenness/centrality scores than non-leaders. These results, along with the high prevalence of brokerage roles in the top posters, also suggest that they are key figures in information exchange. This finding sets the groundwork for future analysis of different definitions of leadership.

In this voluntary virtual organization, members face ambiguity caused by the lack of explicit formal rules and organizational structures. Yet, they organize and sustain their collective effort. Considering affect in communication is one way to understand the interactions and community norms. This study first set out to establish if the large psychological body of literature on gestures, word choices and certainty, in combination with the research on uncertainty in the environment and its effects on perceptions of others, could inform useful human coding of affects of certainty and uncertainty in informal written

communication. The ease with which this coding was conducted (Research Question 1) opens the door to affect detection of certainty and uncertainty in many texts. In particular, it upholds the notion that certainty and uncertainty affect can be found in virtual organizations' written documentation. For that reason, we suggest that affect computing can and should consider certainty and uncertainty in communication, broadening the emotions considered in this kind of analysis as Calvo & D'Mello (2010) and Luneski and Moore (2008) suggest.

Given the extremely uncertain nature of the digital game, and the networked, voluntary, virtual team that had never met, the environmental uncertainty was very high. Yet we found that the players tended towards certainty (54% of the posts). Less than one-third of the posts were uncertain; this suggests a community where certainty was valued. Considering individuals, that emphasis on certainty became even more pronounced. Uncertain and mixed affect differed for the most active contributors, as compared to the rest of the players (Research Question 3). Active posters in the online forum had a far fewer percentage of posts expressing uncertain or mixed affect. This finding suggests that the influence of individuals in online community can be linked to the affect in their communication, in addition to their position in the information flow and social ties. For example, uncertainty scores (level of uncertainty affect) may help to predict who will influence or shape the group's norms (ala Barsade, 2002), particularly since individuals tend to be consistent in their use of certain and uncertain terms (Clark, 1990). As our analysis revealed that players express certainty most of the time, we suspect this might be the case for all individuals who are constructing an argument or building opinions. In this case, the most active players could be expressing less uncertainty because they are more adept at argumentation. This suggests that perhaps certainty and uncertainty affect is a *trait*, not a *state*, because only some individuals expressed less uncertainty (Barsade & Gibson, 2007).

We found that the people who are most socially active in the digital game are non-prototypical leaders—not like the other players in their distribution of mixed and uncertain posts (Research Question 3). This fits with Rast et al.'s (2012) analysis where environmental uncertainty increased acceptance of non-typical leaders. Acceptance of atypical group members suggests that the group is heterogeneous and welcomes different opinions (Hutchison, Jetten & Gutierrez, 2011). Social network analysis showed how these top authors are positioned in the community network structure (Research Question 2): all but one has at least one brokerage role. This means the top (most active) authors make an impact on the flow of information within the community. This topical or opinion dominance through frequent posting is also correlated with higher authority vectors, demonstrating their leadership status (similar to Huffaker's "high communication levels," 2010, p. 593). The top authors' higher betweenness/centrality value also indicates that their posting activities are more critically positioned in information exchanges. Because these most active players express less uncertainty, and uncertainty has been linked to more debate and discussion (Campbell, 1985), these players may be reducing the possibility of people arguing with their opinions, which could give them more authority. If true, this expression of uncertainty affect (or lack of expression) would have an effect on players' social influence in the community.

We suspect the absence of uncertain or mixed posts authored by the most active players' reveals a challenge of collaboration in voluntary virtual teams. Voluntary virtual team members come to make a collective effort to solve problems (or puzzles in our case) with other people they never met before. Perhaps they do not have time to establish trust or ways to define tasks (Munkvold & Zigurs, 2007). Lack of time for socialization does not mean that members skip this socialization part for their collaboration. Linguistic cues may assist in establishing "acceptable" social norms in their social interaction. Individuals with fewer posts may not see themselves as comfortably a part of the community as the top authors, thus tend to rely on "tentative" affect (i.e. uncertain and mixed certainty posts) far more than the top 14 authors. However, this is speculation, as our data is based on a subset of the entire forum.

On the other hand, leaders of the community may feel less compelled to position themselves in order to be accepted. They don't have to make an extra effort to conform. Lack of formal or explicit social rules in voluntary virtual organizations does not mean that the individuals operate without structure. They observe and internalize norms, and possibly their affective traits shape who is likely to become a leader.

Future Work

We suggest that affect computing researchers can and should consider certainty and uncertainty in communication, broadening the emotions considered. This would allow for linking the presence and absence of certainty and uncertainty to social influence and leadership. Using the affect level of communication content as a measure, this study lays the groundwork to help analyze interaction patterns and social norms in online text. Certainty and uncertainty are omnipresent in human communication. For that reason, considering the connection between social interactions and individuals' certainty levels sets the stage for the next round of work on social structure, wherein we will be building two sentiment analysis dictionaries, one for certainty and one for uncertainty. Certainty and uncertainty are related to linguistically unique terms and, based on this

study, to different behaviors (activity levels). Analyzing the effect of topic, different definitions of leadership and other variables may lead to greater understanding of group formation, argumentation and affect.

However, there is another reason why we intend to separate these two affective states. We characterize uncertainty as something constructed, elaborated, and evaluated during social interactions (Michael, 1994). We intend to use the certainty and uncertainty affect as a measure of changing social and organizational structures. Also, understanding the relation between the feeling state (internal) and the informal text (external) could be a key way to understand if certainty and uncertainty affect is a feeling state or a feeling trait (Barsade & Gibson, 2007). In the future, analysis of the communication content could study the changing structure of social interaction, including social roles. For example, Moon and Han (2011) observed that interpersonal similarity and information flow help measure the level of influence for community members in blogosphere. We hope to shift more attention to the content of communication and develop automated detection of affect to enable rapid analysis of affect in virtual organizations. Augmenting LIWC and other automated techniques by creating and augmenting informal language analysis could vastly expand the capabilities of affect detection. This study takes a step in that direction by creating a gold standard and analyzing individuals for their affect of certainty and uncertainty.

ACKNOWLEDGMENTS

REFERENCES

1. Ahmad, Rahayu and Lutters, Wayne G. (2011) Examining ICT-Mediated Cultural Factors for Subgroup Impact on Virtual Team Dynamics. *AMCIS 2011 Proceedings,* Paper 229.

2. Armstrong, D., Gosling, A., Weinman, J., & Marteau, T. (August 01, 1997). The Place of Inter-Rater Reliability in Qualitative Research: An Empirical Study. *Sociology, 31,* 3, 597-606.

3. Balkundi, P., & Harrison, D. (2006). Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. Academy of Management Journal.

4. Barley, Stephen R. (1986) Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments. *Administrative Science Quarterly*, 31: 78-108.

5. Barsade, S. G. (2002) The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*: 644-675.

6. Barsade, S.G. and Gibson, Donald E. (2007) Why does affect matter in organizations? *Academy of Management Perspectives*, 21(1): 36-59.

7. Bechky, B.A. (2006). Gaffers, gofers, and grips: role-based coordination in temporary organizations. *Organization Science* 17 (1): 3–21.

8. Berry, G.R. (2011) Enhancing Effectiveness on Virtual Teams: Understanding Why Traditional Team Skills are Insufficient. *Journal of Business Communication* 48: 186-206.

9. Bettencourt, Louis M.A. (2009) The Rules of Information Aggregation and Emergence of Collective Intelligent Behavior. *Topics in Cognitive Science* 1: 598-620.

10. Brass, D. J. (1984). Being in the right place: A structural analysis of individual influence in an organization. Administrative Science Quarterly, 29, 518−539.

11. Calvo, Rafael A. and D'Mello, Sidney. (2010) Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE* 1(1): 18-37.

12. Campbell, Brian (1985). Uncertainty as Symbolic Action in Disputes Among Experts, *Social Studies of Science*, 15: 429-453.

13. Clark, Dominic A. (1990) Verbal uncertainty expressions: A critical review of two decades of research. *Current Psychology*; Fall, 9(3): 203.

14. Clarkson, Joshua L., Tormala, Zakary, L., DeSensi, Victoria L. and Wheeler, S.C. (2009) "Does attitude certainty beget self-certainty?" *Journal of Experimental Social Psychology*, 45: 436-439.

15. de Nooy, Wouter; Mrvar, Andrej and Batagelj, Vladimir (2005). *Exploratory Social Network Analysis with Pajek*, Cambridge University Press, 2005.

16. Dubach, Erica; Muhdi, Louise; Stöcklin, Dominic; and Michahelles, Florian. (2011) Crowdsourcing for "Kiosk of the Future" – A Retail Store Case Study. *AMCIS 2011 Proceedings*. Paper 324.

17. Freeman, L. C., Roeder, D., & Mulholland, R.R. (1979/1980). Centrality in social networks—ii. Experimental results. Social Networks, 2, 119-141.

18. Gilbert, Eric. (2012) Phrases that Signal Workplace Hierarchy. *CSCW* Feb 11-15.

19. Golder, Scott A. and Macy, Michael W. (2010) Diurnal and Seasonal Mood Varywith Work, Sleep, and Daylength Across Diverse Cultures. *Science*, 333: 1878-1881.

20. Hinds, P. J., and Bailey, D. E. 2003. Out of sight, out of sync: Understanding conflict in distributed teams. *Organization science*: 615-632.

21. Hosman, Lawrence A. and Siltanen, Susan A. (2011) Hedges, Tag Questions, Message Processing and Persuasion. *Journal of Language and Social Psychology*, 30: 341.

22. Huffaker, D. (2010). Dimensions of leadership and social influence in online communities. *Human Communications Research, 36:* 593-617.

23. Hutchison, Paul, Jetten, Jolanda and Gutierrez, Roberto. (2011) Deviant but desirable: Group variability and evaluation of atypical group members. *Journal of Experimental Social Psychology*, 47: 1155-1161.

24. Johnson, Jeffrey C., Palinkas, Lawrence A. and Boster, James S. (2002) Informal Social Roles and the Evolution and Stability of Social Networks, Dynamic Social Network Modeling and Analysis, National Research Council, 2002

25. Kim, Jae Kyung and Yang, Kyung Hoon. (2011) Promoting Knowledge Sharing in Virtual Communities of Practice: Effects of the Cooperative Type and Tie Strength. *AMCIS 2011 Proceedings*. Paper 348.

26. Knippenberg, Daan van. (2011) Embodying who we are: Leader group prototypicality and leadership effectiveness. *The Leadership Quarterly* 22: 1078-1091.

27. Lawrence, P. R., & Lorsch, J. W. (1967). Organization and environment: Managing differentiation and integration. Boston: Harvard University Press.

28. Levinson, S. C. (1983). *Pragmatics*. Cambridge, England: Cambridge University Press.

29. Luneski, Andrej and Moore, Roger K. (2008) Affective computing and Collaborative Networks: Towards emotion-aware interaction. *Pervasive Collaborative Computing*, 283: 315-322.

30. Marlow C (2004), Audience, structure and authority in the weblog community. In: International communication association conference, New Orleans, LA

31. Mehra, Ajay, Dixon, Andrea L., Brass, Daniel J., and Robertson, Bruce (2006) The Social Network Ties of Group Leaders: Implications for Group Performance and Leader Reputation. Organization Science, 17(1): 64-79

32. Michael, M. (1994) Discourse and Uncertainty: Postmodern Variations. *Theory and Psychology*. 4(3): 383-404.

33. Milliken, F.J. (1987) Three Types of Perceived Uncertainty about the Environment: State, Effect and Response Certainty. *Academy of Management Review*, 12(1): 133-143.

34. Mills, Judson. (1965) Effect of Certainty on Exposure to Information Prior to Commitment. *Journal of Experimental Psychology*, 1: 348-355.

35. Moon, Eunyoung and Han, Stave. (2011) A qualitative method to find influencers using similarity-based approach in the blogosphere. *International Journal of Social Computing and Cyber-Physical Systems,* 1:1, 56-78.

36. Mullen, B., & Salas, C. J. (1991). Effects of communication network structure: Components of positional centrality. Social Networks, 13, 169−185.

37. Munkvold, B.E. and Zigurs, Ilze. (2007) Process and technology challenges in swift-starting virtual teams. *Information and Management,* 44: 287-299.

38. Paul, Souren and He, Fang. (2011) Diversity, Group Behavior, and Performance of Short Duration Ad Hoc Virtual Teams: Findings from a Laboratory Experiment. *AMCIS 2011 Proceedings - All Submissions*. Paper 172.

39. Rast, David E., Gaffney, Amber M., Hogg, Michael E. and Crisp, Richard J. (2012) Leadership under uncertainty: When leaders who are non-prototypical group members can gain support. *In Press, Journal of Experimental Psychology*.

40. Smith, John R. (2010) The Red Balloon. *Multimedia, IEEE* 17(2): 2-3.

41. Smith, Joanne R., Hogg, Michael A., Martin, Robin and Terry, Deborah J. (2007) Uncertainty and the influence of group norms in the attitude–behaviour relationship. *British Journal of Social Psychology*, 46: 769–792

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

43. Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

44. Wesson, C.J. and Pulford, B.D. (2009) Verbal Expressions of Confidence and Doubt. *Psychological Reports*, 105(1): 151-160.