

# Automatic Recognition of Conversational Strategies in the Service of a Socially-Aware Dialog System

Ran Zhao, Tanmay Sinha, Alan W Black, Justine Cassell

Language Technologies Institute, School of Computer Science  
Carnegie Mellon University, Pittsburgh, PA 15213 USA  
{rzhaol, tanmays, awb, justine}@cs.cmu.edu

## Abstract

In this work, we focus on automatically recognizing social conversational strategies that in human conversation contribute to building, maintaining or sometimes destroying a budding relationship. These conversational strategies include self-disclosure, reference to shared experience, praise and violation of social norms. By including rich contextual features drawn from verbal, visual and vocal modalities of the speaker and interlocutor in the current and previous turn, we can successfully recognize these dialog phenomena with an accuracy of over 80% and kappa ranging from 60-80%. Our findings have been successfully integrated into an end-to-end socially aware dialog system, with implications for virtual agents that can use rapport between user and system to improve task-oriented assistance.

## 1 Introduction and Motivation

People pursue multiple conversational goals in dialog (Tracy and Coupland, 1990). Contributions to a conversation can be divided into those that fulfill *propositional* functions, contributing informational content to the dialog; those that fulfill *interactional* functions, managing the conversational interaction; and those that fulfill *interpersonal* functions, managing the relationship between the interlocutors (Cassell and Bickmore, 2003; Fetzer, 2013). In the category of talk that fulfills interpersonal goals are conversational strategies - units of discourse that are larger than speech acts (in fact, a single conversational strategy can span more than one turn in conversation), and that can achieve social goals.

In this paper, we propose a technique to automatically recognize conversational strategies. We demonstrate that these conversational strategies are most effectively recognized when verbal (linguistic), visual (nonverbal) and vocal (acoustic) features are all taken into account (and, in a demo paper published in this volume, we demonstrate that the results here can be effectively integrated into an end-to-end socially-aware dialog system).

As naturalistic interactions with dialog systems increasingly become a part of people's daily lives, it is important for these systems to advance their capabilities of not only conveying information and achieving smooth interaction, but also managing long-term relationships with people by building intimacy (Pecune et al., 2013) and rapport (Zhao et al., 2014), not just for the sake of companionship, but as an intrinsic part of successfully fulfilling collaborative tasks.

Rapport, or the feeling of harmony and connection with another, is an important aspect of human interaction, with powerful effects in domains such as education (Ogan et al., 2012; Sinha and Cassell, 2015a; Sinha and Cassell, 2015b) and negotiation (Drolet and Morris, 2000). The central theme of our work is to develop a dialog system that can facilitate such interpersonal rapport with users over interactions in time. Taking a step towards this goal, our prior work (Zhao et al., 2014) has developed a dyadic computational model that explains how interlocutors manage rapport through use of specific conversational strategies to fulfill the intermediate goals that lead to rapport - face management, mutual attentiveness, and coordination.

Foundational work by (Spencer-Oatey, 2008) conceptualizes the interpersonal nature of *face* as a desire to be recognized for one's social value and individual positive traits. Face-boosting strategies such as *praise* serve to create increased self-esteem in the individual and increased interper-

sonal cohesiveness or rapport in the dyad (Zhao et al., 2014). (Spencer-Oatey, 2008) also posits that over time, interlocutors intend to increase *coordination* by adhering to behavior expectations, which are guided by sociocultural norms in the initial stages of interaction and by interpersonally determined norms afterwards. In these later stages, general norms may be purposely violated to accommodate the other’s behavioral expectations.

Meanwhile, in the increasing trajectory of interpersonal closeness, *referring to shared experience* allows interlocutors to increase coordination by indexing common history and differentiating in-group and out-group individuals (Tajfel and Turner, 1979) (cementing the sense that the two are part of a group in ways that similar phenomena such as “referring to shared interests” do not appear to). To better learn about the other person *mutual attentiveness* plays an important role (Tickle-Degnen and Rosenthal, 1990). We have seen in our own corpora that mutual attentiveness is fulfilled by leading one’s interlocutors to provide information about themselves through the strategy of *eliciting self-disclosure*. As the relationship proceeds and social distance decreases, these self-disclosures become more intimate in nature.

Motivated by this theoretical rationale and our prior empirical findings concerning the relationship between these conversational strategies and rapport (Sinha et al., 2015), in the current work, our goals are twofold: Our theoretical question is to understand the nature of conversational strategies in greater detail, by correlating them with associated observable verbal, vocal and visual cues (section 5). Our methodological question is then to use this understanding to automatically recognize these conversational strategies by leveraging statistical machine learning techniques (section 6).

We believe that the answers to these questions can contribute important insights into the nature of human dialog. By the same token, we believe this work to be crucial if we wish to develop a socially-aware dialog system that can identify conversational strategy usage in real-time, assess its impact on rapport, and then produce an appropriate next conversational strategy as a follow-up to maintain or increase rapport in the service of improving the system’s ability to support the user’s goals. (Papangelis et al., 2014).

## 2 Related Work

Below we describe related work that focuses on computational modeling of social conversational phenomena. For instance, (Wang et al., 2016) developed a model to measure self-disclosure in social networking sites by deploying emotional valence, social distance between the poster and other people and linguistic features such as those identified by the Linguistic Inquiry and Word Count program (LIWC) etc. While the features used here are quite interesting, this study relied only on the verbal aspects of talk, while we also include vocal and visual features.

Interesting prior work on quantifying social norm violation has taken a heavily data-driven focus (Danescu-Niculescu-Mizil et al., 2013b; Wang et al., 2016). For instance, (Danescu-Niculescu-Mizil et al., 2013b) trained a series of bigram language models to quantify the violation of social norms in users’ posts on an online community by leveraging cross-entropy value, or the deviation of word sequences predicted by the language model and their usage by the user. Another kind of social norm violation was examined by (Riloff et al., 2013), who developed a classifier to identify a specific type of sarcasm in tweets. They utilized a bootstrapping algorithm to automatically extract lists of positive sentiment phrases and negative situation phrases from given sarcastic tweets, which were in turn leveraged to recognize sarcasm in an SVM classifier. Experimental results showed the adequacy of their approach.

(Wang et al., 2012) investigated the different social functions of language as used by friends or strangers in teen peer-tutoring dialogs. This work was able to successfully predict impoliteness and positivity in the next turn of the dialog. Their success with both annotated and automatically extracted features suggests that a dialog system will be able to employ similar analyses to signal relationships with users. Other work, such as (Danescu-Niculescu-Mizil et al., 2013a) has developed computational frameworks to automatically classify requests along a scale of politeness. Politeness strategies such as requests, gratitude and greetings, as well as their specialized lexicons, were used as features to train a classifier.

In terms of hedges or indirect language, (Prokofieva and Hirschberg, 2014) proposed a preliminary approach to automatic detection, relying on a simple lexical-based search. Machine learn-

ing methods that go beyond keyword searches are a promising extension, as they may be able to better capture language used to hedge as a function of contextual usage.

However, a common limitation of the above work is its focus on only the verbal modality, while studies have shown conversational strategies to be associated with specific kinds of nonverbal behaviors. For instance, (Kang et al., 2012) discovered that head tilts and pauses were the strongest nonverbal cues to interpersonal intimacy. Unfortunately, here too only one modality was examined. While nonverbal behavioral correlates to intimacy in self-disclosure were modeled, the verbal and vocal modalities of the conversation was ignored. Computational work has also modeled rapport using only nonverbal information (Huang et al., 2011). In what follows we describe our approach to modeling social conversational phenomena, which relies on verbal, visual and vocal content to automatically recognize conversational strategies. Our models are trained on a peer tutoring corpus, which gives us the opportunity to look at conversational strategies as they are used in both a task and social context.

### 3 Study Context

Reciprocal peer tutoring data was collected from 12 American English-speaking dyads (6 friends and 6 strangers; 6 boys and 6 girls), with a mean age of 13 years, who interacted for 5 hourly sessions over as many weeks (a total of 60 sessions, and 5400 minutes of data), tutoring one another in algebra (Yu et al., 2013). Each session began with a period of getting to know one another, after which the first tutoring period started, followed by another small social interlude, a second tutoring period with role reversal between the tutor and tutee, and then the final social time.

Prior work demonstrates that peer tutoring is an effective paradigm that results in student learning (Sharpley et al., 1983), making this an effective context to study dyadic interaction with a concrete task outcome. Our student-student data, in addition, demonstrates that a tremendous amount of rapport-building takes place during the task of reciprocal tutoring (Sinha and Cassell, 2015b).

### 4 Ground Truth

We assessed our automatic recognition of conversational strategies against this corpus annotated

for those strategies (as well as other educational tutoring phenomena not discussed here). Interrater reliability (IRR) for the conversational strategy annotations, computed via Krippendorff's alpha, was 0.75 for self-disclosure, 0.79 for reference to shared experience, 1.0 for praise and 0.75 for social norm violation. IRR for visual behavior was 0.89 for eye gaze, 0.75 for smile count (how many smiles occur), 0.64 for smile duration and 0.99 for head nod. Below we discuss the definitions of each conversational strategy and nonverbal behavior that was annotated.

#### 4.1 Coding Conversational Strategies

**Self-Disclosure (SD):** Self-disclosure refers to the conversational act of revealing aspects of oneself (personal private information) that otherwise would not be seen or known by the person being disclosed to (or would be difficult to see or know). A lot of psychological literature talks about the ways people reveal facts about themselves as ways of building relationships, but we are the first to look at the role of self-disclosure during social and task interactions by the same dyad, particularly for adolescents engaged in reciprocal peer tutoring. We coded for two sub-categories: (1) revealing the long-term aspects of oneself that one may feel are deep and true (e.g, "I love my pets"), (2) revealing one's transgressive (forbidden or socially-unacceptable) behaviors or actions, which may be a way of attempting to make the interlocutor feel better by disclosing one's flaws (e.g, "I suck at linear equations").

**Referring to Shared Experience (SE):** We differentiate between shared experience - an experience that the two interlocutors engage in or share with one another at the same time (such as "that facebook post Cecily posted last week was wild!") - from shared interests (such as "you like Xbox games too?"). Shared experiences may index a shared community membership (even if a community of two), which can in turn build rapport. We coded for shared experiences (e.g, going to the mall together last week).

**Praise (PR):** We annotated both labeled praise (an expression of a positive evaluation of a specific attribute, behavior or product of the other; e.g, "great job with those negative numbers"), and unlabeled praise (a generic expression of positive evaluation, without a specific target; e.g, "Perfect").

**Violation of Social Norms (VSN):** Social norm violations are behaviors or actions that go against general socially acceptable and stereotypical behaviors. In a first pass, we coded whether an utterance was a social norm violation. In a second pass, if a social norm violation, we differentiated: (1) breaking the conversational rules of the experiment (e.g. off-task talk during tutoring session, insulting the experimenter or the experiment, etc); (2) face threatening acts (e.g. criticizing, teasing, or insulting, etc); (3) referring to one's own or the other person's social norm violations or general social norm violations (e.g. referring to the need to get back to focusing on work, or to the other person being verbally annoying etc). Social norms are culturally-specific, and so we judged a social norm violation by the impact it had on the listener (e.g. shock, specific reference to the behavior as a violation, etc.). Social norm violations may signal that a dyad is becoming closer, and no longer feels the need to adhere to the norms of the larger community.

## 4.2 Coding Visual Behaviors

**Eye Gaze:** Gaze for each participant was annotated individually. Front facing video for the individual participant was supplemented with a side camera view when needed. Audio was turned off so that words didn't influence the annotation. We coded (1) Gaze at the partner (gP), (2) Gaze at one's own worksheet (gO), (3) Gaze at partner's worksheet (gN), (4) Gaze elsewhere (gE).

**Smile:** A smile is defined by the elongation of the participant's lips and rising of their cheeks (smiles will often be asymmetric). It is often accompanied by creases at the corner of the eyes. Smiles have three parameters: rise, sustain, and decay (Hoque et al., 2011). We annotated a smile from the beginning of the rise to the end of the decay.

**Head Nod:** We coded temporal intervals of head nod rather than individual nod - the beginning of the head moving up and down until the moment the head came to rest.

## 5 Understanding Conversational Strategies

Our first objective, then, was to understand the nature of different conversational strategies (discussed in section 4) in greater detail. Towards this end, we first under-sampled the non-annotated

examples of self disclosure, shared experience, praise and social norm violation in order to create a balanced dataset of utterances. The utterances chosen to reflect the non-annotated cases were randomly selected. We made sure to have a similar average utterance length for all annotated and non-annotated cases, to prevent conflation of results due to lower or higher opportunities for detection of multimodal features. The final corpus (selected from 60 interaction sessions) comprised of 1014 self disclosure and 1014 non-self disclosure, 184 shared experience and 184 non-shared experience, 167 praise and 167 non-praise, 7470 social norm violation and 7470 non-social norm violation.

Second, we explored observable verbal and vocal behaviors of interest that could potentially be associated with different conversational strategies, assessing whether the mean value of these features were significantly higher in utterances with a particular conversational strategy label than in ones with no label (two-tailed correlated samples t-test). Bonferroni correction was used to correct the p-values with respect to the number of features, because of multiple comparisons involved. Finally, for all significant results ( $p < 0.05$ ), we also calculated effect size via Cohen's  $d$  to test for generalizability of results.

Third, for visual behaviors like smile, eye gaze, head nod, we binarized these features by denoting their presence (1) or absence (0) in one clause. If an individual shifts gaze during a particular spoke conversational strategy, we might have multiple types of eye gaze represented. We performed  $\chi^2$  test to see whether the appearance of visual annotations were independent of whether the utterance belonged to a particular conversational strategy or not. For all significant  $\chi^2$  test statistics, odds ratio ( $o$ ) was computed to explore co-occurrence likelihood. Majority of the features discussed in the subsequent sub-sections were drawn from qualitative observations and note-taking, during and after the formulation of our coding manuals.

### 5.1 Verbal

We used Linguistic Inquiry and Word Count (LIWC 2015) (Pennebaker et al., 2015) to quantify verbal cues of interest that were semantically associated with a broad range of psychological constructs and could be useful in distinguishing conversational strategies. The input to LIWC were conversational transcripts that had been tran-

scribed and segmented into syntactic clauses.

**Self-disclosure:** We observed personal concerns of students (sum of words identified as belonging to categories of work, leisure, home, money, religion and death etc) to be significantly higher, than in non self-disclosure utterances with a moderate effect size ( $d=0.44$ ), signaling that students referred significantly more to their personal concerns during self-disclosure. Next, due to the fact that self-disclosures are often likely to comprise of emotional expressions when revealing one's likes and dislikes (Sparrevoorn and Rapee, 2009), we used the LIWC dictionary to capture words representative of negative emotions ( $d=0.32$ ) and positive emotion words ( $d=0.18$ ). Also, to formalize the intuition that when people reveal themselves in an authentic or honest way, they are more personal, humble, and vulnerable, the standardized LIWC summary variable of Authenticity ( $d=1.16$ ) was taken into account. Finally, as expected, we found self-disclosure utterances had significantly higher usage of first person singular pronouns ( $d=1.62$ ).

**Reference to shared experience:** We looked at three LIWC categories: (1) Affiliation drive, which comprises words signaling a need to affiliate such as ally, friend, social etc ( $d=0.92$ ), (2) Time Orientation words, which capture past (mostly in ROE) , present (mostly in RIE) and future focus and comprises words such as ago, did, talked, today, is, now, may, will, soon etc ( $d=0.95$ ). Such words are not only used by interlocutors to index commonality within a time frame (Enfield, 2013), but also to signal an increased need for affiliation with the conversational partner, perhaps to indicate common ground (Clark, 1996), (3) First person plural such as we, us, our etc. In line with expectations, this feature had high effect size ( $d=0.93$ ), since interlocutors focused on both themselves and the conversational partner.

**Praise:** We looked at positive emotions ( $d=2.55$ ), since praise is one form of verbal persuasion that increases the interlocutor's confidence and boosts self efficacy (Bandura, 1994). Most of the praise utterances in our dataset were not very specific or directed at the tutee's performance or effort. Also, the LIWC standardized summary variable of Emotional Tone from LIWC was considered for the sake of completeness, which puts positive emotion and negative emotion dimensions into a single summary variable, such

that the higher the number, the more positive the tone ( $d=3.56$ ).

**Social norm violation:** We looked at different categories of off-task talk from LIWC, such as social processes comprising words related to friends, family, male and female references ( $d=0.78$ ), biological processes comprising words belonging to the categories of body, health etc ( $d=0.30$ ) and personal concerns ( $d=0.24$ ). The effect sizes across these categories ranged from moderate to low. Next, we looked at usage of swearing words like fuck, damn, shit etc and found low effect size ( $d=0.13$ ) for this category in utterances of social norm violation. For the LIWC category of anger (words such as hate, annoyed etc), the effect size was moderate ( $d=0.27$ ).

In our qualitative analysis of social norm violation utterances, we had discovered interactions of students to be reflective of need for power, meaning attention to or awareness of relative status in a social setting (perhaps this could be a result of putting one student in the tutor role). We formalized this intuition from the LIWC category of power drive that comprises words such as superior etc ( $d=0.18$ ). Finally, based on prior work (Kacewicz et al., 2009) that found increased use of first-person plural to be a good predictor of higher status, and increased use of first-person singular to be a good predictor of lower status, we posited that when students violated social norms, they were more likely to freely make statements that involved others. However, the effect size for first-person plural usage in utterances of social norm violation was negligible ( $d=0.07$ ). Table 2 in the appendix provides complete set of results.

## 5.2 Vocal

In our qualitative observations, we noticed the variations of both pitch and loudness when interlocutors used different conversational strategies. We were thus motivated to explore the mean difference of those low-level vocal descriptors as differentiators among the different conversational strategies. By using Open Smile (Eyben et al., 2010), we extracted two sets of basic features - for loudness features, pcm-loudness and its delta coefficient were tested; for pitch-based features, jitterLocal, jitterDDP, shimmerLocal, F0final and also their delta coefficients were tested. pcm-loudness represents the loudness as the normalised intensity raised to a power of 0.3. F0final is the

smoothed fundamental frequency contour. Jitter-Local is the frame-to-frame pitch period length deviations. JitterDDP is the differential frame-to-frame jitter. ShimmerLocal is the frame-to-frame amplitude deviations between pitch periods.

**Self-disclosure:** We found a moderate effect size for pcm-loudness-sma-amean ( $d=0.26$ ). Despite often becoming excited when disclosing things that they loved or liked, sometimes students also seemed to hesitate and spoke at a lower pitch when they revealed a transgressive act. However, the effect size for pitch was negligible. One potential reason for our results not aligning with hypothesis could be consideration of utterances with annotations of enduring states as well as transgressive acts together.

**Reference to shared experience:** We found a moderate negative effect size for the shimmerLocal-sma-amean ( $d=-0.32$ ).

**Praise:** We found negative effect size for loudness ( $d=-0.51$ ), meaning the speakers spoke in a lower voice when praising the interlocutor (mostly the tutee). We also found positive and moderate effect sizes for jitterLocal-sma-amean ( $d=0.45$ ) and shimmerLocal-sma-amean ( $d=0.39$ ).

**Social norm violation:** We found high effect sizes for pcm-loudness-sma-amean ( $d=0.72$ ) and F0final-sma-amean ( $d=0.61$ ) and interestingly, negative effect sizes for jitter ( $d=-0.09$ ) and shimmer ( $d=-0.16$ ). One potential reason could be that when student violate social norms, their behaviors are likely to become outliers compared to their normative behaviors. In fact, we noticed usage of “joking” tone of voice (Norrick, 2003) and pitch different than usual, to signal a social norm violation. When the content of the utterance was unaccepted by the social norms, students also tried to lower down their voice, which could be a way of hedging these violations. Table 2 in the appendix provides complete set of results.

### 5.3 Visual

Computing the odds ratio  $o$  involved comparing the odds of occurrence of a non-verbal behavior for a pair of categories of a second variable (whether an utterance was a specific conversational strategy or not). Overall, we found that that smile and gaze were significantly more likely to occur in utterances of self-disclosure ( $o(\text{Smile})=1.67$ ,  $o(\text{gP})=2.39$ ,  $o(\text{gN})=0.498$ ,  $o(\text{gO})=0.29$ ,  $o(\text{gE})=2.8$ ) compared to a non self-disclosure utterance. A similar

trend was observed for reference to shared experience ( $o(\text{Smile})=1.75$ ,  $o(\text{gP})=3.02$ ,  $o(\text{gN})=0.58$ ,  $o(\text{gO})=0.31$ ,  $o(\text{gE})=4.19$ ) and social norm violation ( $o(\text{Smile})=3.35$ ,  $o(\text{gP})=2.75$ ,  $o(\text{gN})=0.8$ ,  $o(\text{gO})=0.47$ ,  $o(\text{gE})=1.67$ ) utterances, compared to utterances that did not belong to these categories.

The high odds ratio for gP in these results suggests that an interlocutor was likely to gaze at their partner when using specific conversational strategies, signaling attention towards the interlocutor. The extremely high odds ratio for smiling behaviors during a social norm violation is also interesting. However, for praise utterances, we did not find all kinds of gaze and smile to be more likely to occur than non-praise utterances. Only gazing at partner ( $o(\text{gP})=0.44$ ) or their worksheet ( $o(\text{gN})=4.29$ ) or gazing elsewhere ( $o(\text{gE})=0.30$ ) were among the non-verbals that were significantly greatly present in praise utterances. Table 3 in the appendix provides complete set of results for the speaker (as discussed above) and also for the listener.

## 6 Machine Learning Modeling

In this section, our objective was to build a computational model for conversational strategy recognition. Towards this end, we first took each clause, or the smallest units that can express a complete proposition, as the prediction unit. Next, three sets of features were used as input. The first set  $f_1$  comprised verbal (LIWC), vocal and visual features of the speaker, informed from the qualitative and quantitative analysis as discussed above. While LIWC features helped in categorization of words used during usage of a particular conversational strategy, they did not capture contextual usage of words within the utterance. Thus, we also added bigrams, part of speech bigrams and word-part of speech pairs from the speaker’s utterance.

In addition to the speaker’s behavior, we also added two sets of interlocutor behavior to capture the context around usage of a conversational strategy. The feature set  $f_2$  comprised visual behaviors of the interlocutor (listener) in the current turn. The feature set  $f_3$  comprised verbal (bigrams, part of speech bigrams and word-part of speech pairs), vocal and visual features of the interlocutor in the previous turn.

Finally, early fusion was applied on these multimodal features (by concatenation) and L2 regularized logistic regression with 10-fold cross val-

iation was used as the machine learning algorithm, with rare threshold for feature extraction being set to 10 and performance evaluated using accuracy and kappa<sup>1</sup> measures. The following table shows our comparison with other standard machine learning algorithms such as Support Vector Machine (SVM) and Naive Bayes (NB), where we found Logistic Regression (LR) to perform better in recognition of the four conversational strategies. In next sub-section, we therefore denote the feature weights derived from logistic regression in brackets to offer interpretability of results.

Conversational Strategy	LR	SVM	NB
Self-disclosure	Acc=0.85 $\kappa = 0.7$	Acc=0.84 $\kappa = 0.68$	Acc=0.83 $\kappa = 0.65$
Shared Experience	Acc=0.84 $\kappa = 0.67$	Acc=0.82 $\kappa = 0.64$	Acc=0.79 $\kappa = 0.59$
Praise	Acc=0.91 $\kappa = 0.81$	Acc=0.90 $\kappa = 0.80$	Acc=0.88 $\kappa = 0.76$
Social Norm Violation	Acc=0.80 $\kappa = 0.61$	Acc=0.78 $\kappa = 0.55$	Acc=0.73 $\kappa = 0.47$

Table 1: Comparative Performance Evaluation using Accuracy (Acc) and Kappa ( $\kappa$ ) for Logistic Regression (LR), Support Vector Machine (SVM) and Naive Bayes (NB)

## 6.1 Results and Discussion

**Self-Disclosure:** We could successfully identify self-disclosure from non self-disclosure utterances with an accuracy of **85%** and a kappa of **70%**. The top features from feature set  $f_1$  predictive of speakers disclosing themselves included gazing at partner (0.44), head nodding (0.24) and not gazing at their own worksheet (-0.60) or the interlocutor’s worksheet (-0.21). Head nod is a way to emphasize what one is saying (Poggi et al., 2010), while gazing at the partner signals one’s attention. Higher usage of first person singular by the speaker (0.04) was also positively predictive of self-disclosure in the utterance. The top features from feature set  $f_2$  predictive of speakers disclosing included listener behaviors such as head nodding (0.3) to communicate their attention (Schegloff, 1982), gazing elsewhere (0.12) or at the speaker (0.09) instead of gazing at their own worksheet (-0.89) or the speaker’s worksheet (-0.27). The top features from feature set  $f_3$  predictive of speakers disclosing included no smiling

<sup>1</sup>The discriminative ability over chance of a predictive model, for the target annotation, or the accuracy adjusted for chance

(-0.30), no head nodding (-0.15) and lower loudness in voice (-0.11) from the interlocutor in the last turn.

**Reference to shared experience:** We achieved an accuracy of **84%** and kappa of **67%** for prediction. The top features from feature set  $f_1$  predictive of speakers referring to shared experience included not gazing at own worksheet (-0.66), partner’s worksheet (-0.40) or at the partner (-0.22), no smiling (-0.18) and having lower shimmer in voice (-0.26). Instead, words signaling affiliation drive (0.07) and time orientation (0.06) from the speaker were deployed to index shared experience. The top features from feature set  $f_2$  predictive of speakers using shared experience included listener behaviors such as smiling (0.53) perhaps to indicate appreciation towards the content of the talk, or encourage the speaker to go on (Niewiadomski et al., 2010). Besides, the listener gazing elsewhere (0.50) or at the speaker (0.47), and neither gazing at own worksheet (-0.45) nor head nodding (-0.28) had strong predictive power. The top features from feature set  $f_3$  predictive of speakers using shared experience included lower loudness in voice (-0.58), smiling (0.47), gazing elsewhere (0.59), at own worksheet (0.27) or at the partner (0.22) but not at partner’s worksheet (-0.40) from the interlocutor in the last turn.

**Praise:** For praise, our computational model achieved an accuracy of **91%** and kappa of **81%**. The top features from feature set  $f_1$  predictive of speakers using praise included gazing at partner’s worksheet (0.68) indicative of directing attention to the partner’s (perhaps the tutee’s) work, smiling (0.51), perhaps to mitigate the potential embarrassment of praise (Niewiadomski et al., 2010) and head nodding (0.35) with a positive tone of voice (0.04), perhaps to emphasize the praise. The top features from feature set  $f_2$  predictive of speakers using praise included listener behaviors such as head nodding (0.45) for backchanneling and acknowledgement and not gazing at partner’s worksheet (-1.06), elsewhere (-0.5) or at the partner (-0.49). The top features from feature set  $f_3$  predictive of speakers using praise included smiling (0.51), lower loudness in voice (-0.91) and overlap (-0.66) from the interlocutor in the last turn.

**Violation of Social Norm:** We achieved an accuracy of **80%** and kappa of **61%** for prediction. The top features from feature set  $f_1$  predictive of speakers violating social norms included smiling

(0.40), gazing at partner (0.45) but not head nodding (-0.389). (Keltner and Buswell, 1997) introduced a remedial account of embarrassment, emphasizing that smiles signal awareness of a social norm being violated and serve to provoke forgiveness from the interlocutor, in addition to being a hedging indicator. (Kraut and Johnston, 1979) posited that smiling evolved from primate appeasement displays and is likely to occur when a person has violated a social norm. The top features from feature set  $f_2$  predictive of speakers violating social norms included listener behaviors such as smiling (0.54), gazing at own worksheet (0.32) or at the partner's (0.14). The top features from feature set  $f_3$  predictive of speakers violating social norms included high loudness (0.86) and jitter in voice (0.50), lower shimmer in voice (-0.53), gazing at own worksheet (0.49) and no head nodding (-0.31) from the interlocutor in the last turn.

## 6.2 Implications

We began this paper indicating our interest in better understanding conversational strategies in and of themselves, and in employing automatic recognition of conversational strategies to improve interactive systems. With respect to this first goal, because the current approach takes into account verbal, vocal and visual behaviors, it can identify regularities in social interaction processes that have not been identified by earlier work. This becomes especially important as automatic behavioral analysis increasingly develops new real-time metrics to predict other kinds of conversational strategies related to interpersonal dynamics like politeness, sarcasm etc, that are not easily captured by observer-based labeling. Similar benefits may accrue in other areas of automated human behavior understanding.

With respect to interactive systems, these findings are applicable to building virtual peer tutors in whom rapport improves learning gains as it does for human-human tutors, training military personnel and police to build rapport with the communities in which they work, and trustworthy dialog systems for clinical decision support (DeVault et al., 2013). Improved understanding of conversational strategy response pairs can help us better estimate the level of rapport at a given point in a dialog (Sinha et al., 2015; Zhao et al., 2016), which means that for the design of interactive systems, our work could help improve the capability of a natural language understanding module to capture

user's interpersonal goals, such as those of building, maintaining or destroying rapport.

More broadly, understanding of these particular ways of talking may also help us in building artificially intelligent systems that exhibit and evoke behaviors not just as conversationalists, but also as confidants to whom we can relay personal and emotional information with the expectation of acknowledgement, empathy and sympathy in response (Boden, 2010). These social strategies improve the bond between interlocutors which, in turn, can improve the efficacy of their collaboration. Efforts to experimentally generate interpersonal closeness (Aron et al., 1997) to achieve positive task and social outcomes depend on advances in moving beyond behavioral channels in isolation and leveraging the synergy and complementarity provided by multimodal human behaviors.

## 7 Conclusion

In this work, by performing quantitative analysis of our peer tutoring corpus followed by machine learning modeling, we learnt the discriminative power and generalizability of verbal, vocal and visual behaviors from both the speaker and listener, in distinguishing conversational strategy usage.

We found that interlocutors usually accompany the disclosure of personal information with head nods and mutual gaze. When faced with such self-disclosure listeners, on the other hand, often nod and avert their gaze. When the conversational strategy of reference to shared experience is used, speakers are less likely to smile, and more likely to avert their gaze (Cassell et al., 2007). Meanwhile, listeners smile to signal their coordination. When speakers praise their partner, they direct their gaze to the interlocutor's worksheet, smile and nod with a positive tone of voice. Meanwhile, listeners simply smile, perhaps to mitigate the embarrassment of having been praised.

Finally, speakers tend to gaze at their partner and smile when they violate a social norm, without nodding. The listener, faced with a social norm violation, is likely to smile extensively (once again, most likely to mitigate the face threat of social norm violations such as teasing or insults). Overall, these results present an interesting interplay of multimodal behaviors at work when speakers use conversational strategies to fulfil interpersonal goals in a dialog.

These results have been integrated into a real-time end-to-end socially aware dialog system



(SARA)<sup>2</sup> described in (Matsuyama et al., 2016) in this same volume. SARA is capable of detecting conversational strategies, relying on the conversational strategies detected in order to accurately estimate rapport between the interlocutors, reasoning about how to respond to the intentions behind those particular behaviors, and generating appropriate social responses as a way of more effectively carrying out her task duties. To our knowledge, SARA is the first socially-aware dialog system that relies on visual, verbal, and vocal cues to detect user social and task intent, and generates behaviors in those same channels to achieve her social and task goals.

## 8 Limitations and Future Work

We acknowledge some methodological limitations in the current work. In the current work we under-sampled the negative examples in order to make a balanced dataset. For future work, we will work with corpora that have a more natural distribution and deal with the sparsity of the phenomena through machine learning methods. This will improve applicability to a real-time system where conversation strategies are likely to be less frequent than in our training dataset. Moreover, in current work, we looked at individual modalities in isolation initially, and fused them later via a simple concatenation of feature vectors. Including sequentially occurring features may better exploit correlation and dependencies between features from different modalities. As a next step, we have thus started to investigate the impact of temporal ordering of verbal and visual behaviors that lead to increased rapport (Zhao et al., 2016).

In terms of future work, one concrete example of an application area where we are beginning to apply these findings is the domain of learning technologies. While we know from research on dialog-based intelligent tutoring systems that conversations with such computer systems help students learn (Graesser, 2016), we also know that those students who are academically challenged, perhaps because under-represented in the fields they are trying to learn (Robinson et al., 2005), are most likely to need a social component to their learning interactions. Hence a major critique of existent intelligent tutoring systems is that they serve to fulfil only the task-goal of the interaction. Traditionally (DMello and Graesser, 2013), this is instantiated via an expectation and misconception

tailored dialog directed towards the portions of learning content where student under-performance is noted, and simply blended with some motivational scaffolding.

Despite significant advances in such conversational tutoring systems (Rus et al., 2013), we believe that future systems that provide intelligent support for tutoring via dialog should support the social as well as task nature of natural peer tutoring. Because learning does not happen in a cultural or social void, it is important to think about how we can leverage dialog, the natural modality of pedagogy, to foster supportive relationships that make learning challenging, engaging and meaningful<sup>3</sup>.

We have also begun to use the social conversational strategies described here to complement the curriculum script in a traditional tutoring dialogue comprising knowledge-telling or knowledge-building utterances, shallow or deep question asking, hints and other forms of feedback. We believe this is a step towards building SCEM-sensitive (social, cognitive, emotional and motivational) tutors (Graesser et al., 2010), and towards more accurate computational models of human interaction that will need to underlie those new kinds of intelligent tutors.

Dialog systems that can recognize and use conversational strategies such as self-disclosure, reference to shared experience, and violation of social norms, are also part of a new genre of dialog system that departs from the rigid repetitive natural language generation templates of the olden days, and that can learn to speak with style. It is conceivable that contemporary corpus-based approaches to NLG that introduce stylistic variation into a dialog (Wen et al., 2015) may one day learn on the user's own conversational style, and entrain to it. In a system like that, real-time recognition of conversational strategies like that demonstrated here could play an essential role.

## References

- Arthur Aron, Edward Melinat, Elaine N Aron, Robert Darrin Vallone, and Renee J Bator. 1997. The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin*, 23(4):363–377.
- Albert Bandura. 1994. *Self-efficacy*. Wiley Online Library.

<sup>2</sup>[sociallyawarerobotassistant.net](http://sociallyawarerobotassistant.net)

<sup>3</sup><http://articulab.hcii.cs.cmu.edu/projects/rapt/>

- Margaret A Boden. 2010. Conversationalists and confidants. *Artificial Companions in Society: Perspectives on the Present and Future*, page 5.
- Justine Cassell and Timothy Bickmore. 2003. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction*, 13(1-2):89–132.
- Justine Cassell, Alastair J Gill, and Paul A Tepper. 2007. Coordination in conversation and rapport. In *Proceedings of the workshop on Embodied Language Processing*, pages 41–50. ACL.
- Herbert H Clark. 1996. *Using language*. Cambridge university press.
- Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013a. A computational approach to politeness with application to social factors. *arXiv preprint arXiv:1306.6078*.
- Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013b. No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22nd international conference on World Wide Web*, pages 307–318. International World Wide Web Conferences Steering Committee.
- David DeVault, Kallirroi Georgila, Ron Artstein, Fabrizio Morbini, David Traum, Stefan Scherer, Albert Rizzo, and Louis-Philippe Morency. 2013. Verbal indicators of psychological distress in interactive dialogue with a virtual human. In *SIGDIAL*, Metz, France, August.
- Aimee L Drolet and Michael W Morris. 2000. Rapport in conflict resolution: Accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *Journal of Experimental Social Psychology*, 36(1):26–50.
- Sidney DMello and Art Graesser. 2013. Design of dialog-based intelligent tutoring systems to simulate human-to-human tutoring. In *Where Humans Meet Machines*, pages 233–269. Springer.
- Nick J Enfield. 2013. Reference in conversation. *The handbook of conversation analysis*, pages 433–454.
- Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the international conference on Multimedia*, pages 1459–1462. ACM.
- Anita Fetzer. 2013. ‘no thanks’: a socio-semiotic approach. *Linguistik Online*, 14(2).
- Arthur C Graesser, Fazel Keshtkar, and Haiying Li. 2010. The role of natural language and discourse processing in advanced tutoring systems. In *Oxford Handbook of Language and social psychology*, pages 1–12. New York: Oxford University Press.
- Arthur C Graesser. 2016. Conversations with autotutor help students learn. *International Journal of Artificial Intelligence in Education*, pages 1–9.
- Mohammed Hoque, Louis-Philippe Morency, and Rosalind W Picard. 2011. Are you friendly or just polite?—analysis of smiles in spontaneous face-to-face interactions. In *Affective Computing and Intelligent Interaction*, pages 135–144. Springer.
- Lixing Huang, Louis-Philippe Morency, and Jonathan Gratch. 2011. Virtual rapport 2.0. In *International Workshop on Intelligent Virtual Agents*, pages 68–79. Springer.
- E Kacewicz, J. W Pennebaker, M Davis, M Jeon, and A. C Graesser. 2009. The language of social hierarchies.
- Sin-Hwa Kang, Jonathan Gratch, Candy Sidner, Ron Artstein, Lixing Huang, and Louis-Philippe Morency. 2012. Towards building a virtual counselor: modeling nonverbal behavior during intimate self-disclosure. In *Proceedings of the 11th International Conference on Autonomous Agents and Multi-Agent Systems-Volume 1*, pages 63–70.
- Dacher Keltner and Brenda N Buswell. 1997. Embarrassment: its distinct form and appeasement functions. *Psychological bulletin*, 122(3):250.
- Robert E Kraut and Robert E Johnston. 1979. Social and emotional messages of smiling: An ethological approach. *Journal of personality and social psychology*, 37(9):1539.
- Yoichi Matsuyama, Arjun Bhardwaj, Ran Zhao, Oscar J. Romero, Sushma Akoju, and Justine Cassell. 2016. Socially-aware animated intelligent personal assistant agent. In *17th Annual SIGdial Meeting on Discourse and Dialogue*.
- Radoslaw Niewiadomski, Ken Prepin, Elisabetta Bevacqua, Magalie Ochs, and Catherine Pelachaud. 2010. Towards a smiling eca: studies on mimicry, timing and types of smiles. In *Proceedings of the 2nd international workshop on Social signal processing*, pages 65–70.
- Neal R Norrick. 2003. Issues in conversational joking. *Journal of Pragmatics*, 35(9):1333–1359.
- Amy Ogan, Samantha Finkelstein, Erin Walker, Ryan Carlson, and Justine Cassell. 2012. Rudeness and rapport: Insults and learning gains in peer tutoring. In *Intelligent Tutoring Systems*, pages 11–21. Springer.
- Alexandros Papangelis, Ran Zhao, and Justine Cassell. 2014. Towards a computational architecture of dyadic rapport management for virtual agents. In *Intelligent Virtual Agents*, pages 320–324. Springer.
- Florian Pecune, Magalie Ochs, Catherine Pelachaud, et al. 2013. A formal model of social relations for artificial companions. In *Proceedings of The European Workshop on Multi-Agent Systems (EUMAS)*.

- James W Pennebaker, Ryan Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015.
- Isabella Poggi, Francesca D’Errico, and Laura Vincze. 2010. Types of nods. the polysemy of a social signal. In *LREC*.
- Anna Prokofieva and Julia Hirschberg. 2014. Hedging and speaker commitment. In *5th Intl. Workshop on Emotion, Social Signals, Sentiment & Linked Open Data, Reykjavik, Iceland*.
- Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. Sarcasm as contrast between a positive sentiment and negative situation. In *EMNLP*, pages 704–714.
- Debbie R Robinson, Janet Ward Schofield, and Katrina L Steers-Wentzell. 2005. Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review*, 17(4):327–362.
- Vasile Rus, Sidney DMello, Xiangen Hu, and Arthur Graesser. 2013. Recent advances in conversational intelligent tutoring systems. *AI magazine*, 34(3):42–54.
- Emanuel A Schegloff. 1982. Discourse as an interactional achievement: Some uses of uh huh and other things that come between sentences. *Analyzing discourse: Text and talk*, 71:93.
- Anna M Sharpley, James W Irvine, and Christopher F Sharpley. 1983. An examination of the effectiveness of a cross-age tutoring program in mathematics for elementary school children. *American Educational Research Journal*, 20(1):103–111.
- Tanmay Sinha and Justine Cassell. 2015a. Fine-grained analyses of interpersonal processes and their effect on learning. In *Artificial Intelligence in Education*, pages 781–785. Springer.
- Tanmay Sinha and Justine Cassell. 2015b. We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. In *Proceedings of the 2015 Workshop on Modeling Interpersonal Synchrony, 17th ACM International Conference on Multimodal Interaction*. ACM.
- Tanmay Sinha, Ran Zhao, and Justine Cassell. 2015. Exploring socio-cognitive effects of conversational strategy congruence in peer tutoring. In *Proceedings of the 2015 Workshop on Modeling Interpersonal Synchrony, 17th ACM International Conference on Multimodal Interaction*. ACM.
- Roslyn M Sparrevohn and Ronald M Rapee. 2009. Self-disclosure, emotional expression and intimacy within romantic relationships of people with social phobia. *Behaviour Research and Therapy*, 47(12):1074–1078.
- Helen Spencer-Oatey. 2008. Face,(im) politeness and rapport.
- Henri Tajfel and John C Turner. 1979. An integrative theory of intergroup conflict. *The social psychology of intergroup relations*, 33(47):74.
- Linda Tickle-Degnen and Robert Rosenthal. 1990. The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4):285–293.
- Karen Tracy and Nikolas Coupland. 1990. Multiple goals in discourse: An overview of issues. *Journal of Language and Social Psychology*, 9(1-2):1–13.
- William Yang Wang, Samantha Finkelstein, Amy Ogan, Alan W Black, and Justine Cassell. 2012. Love ya, jerkface: using sparse log-linear models to build positive (and impolite) relationships with teens. In *13th annual SIGdial meeting on discourse and dialogue*, pages 20–29. Association for Computational Linguistics.
- Yi-Chia Wang, Moira Burke, and Robert Kraut. 2016. Modeling self-disclosure in social networking sites. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 74–85. ACM.
- Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, Pei-Hao Su, David Vandyke, and Steve Young. 2015. Semantically conditioned lstm-based natural language generation for spoken dialogue systems. *arXiv preprint arXiv:1508.01745*.
- Zhou Yu, David Gerritsen, Amy Ogan, Alan W Black, and Justine Cassell. 2013. Automatic prediction of friendship via multi-model dyadic features. In *14th Annual SIGdial Meeting on Discourse and Dialogue, Metz, France*.
- Ran Zhao, Alexandros Papangelis, and Justine Cassell. 2014. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *Intelligent Virtual Agents*, pages 514–527. Springer.
- Ran Zhao, Tanmay Sinha, Alan Black, and Justine Cassell. 2016. Socially-aware virtual agents: Automatically assessing dyadic rapport from temporal patterns of behavior. In *16th International Conference on Intelligent Virtual Agents*.

## A Appendix: Complete Statistics for Understanding Conversational Strategies (Section 5)

Conversational Strategy	Verbal/Vocal(Speaker)	t-test value	Mean value	Effect Size
1. Self-Disclosure	LIWC Personal Concerns	t(1013)=7.06***	SD=4.13, NSD=1.58	d=0.44
	LIWC Positive Emotion	t(1013)=2.98**	SD=7.61, NSD=5.50	d=0.18
	LIWC Negative Emotion	t(1013)=5.51***	SD=5.62, NSD=2.22	d=0.32
	LIWC First Person Singular	t(1013)=25.87***	SD=20.12, NSD=7.77	d=1.62
	LIWC Authenticity	t(1013)=18.59***	SD=66.71, NSD=34.07	d=1.16
	pcm-loudness-sma-amean	t(1013)=4.11***	SD=0.64, NSD=0.59	d=0.26
2. Shared Experience	LIWC Affiliation Drive	t(183)=6.22***	SE=4.64, NSE=0.77	d=0.92
	LIWC Time Orientation	t(183)=6.47***	SE=24.89, NSE=15.02	d=0.95
	LIWC First Person Plural	t(183)=6.29***	SE=3.99, NSE=0.48	d=0.93
	shimmerLocal-sma-amean	t(183)=-2.21*	SE=0.18, NSE=0.194	d=0.32
3. Praise	LIWC Positive Emotion	t(166)=16.48***	PR=55.63, NPR=4.56	d=3.56
	LIWC Emotional Tone	t(166)=22.96***	PR=91.1, NPR=33.5	d=2.55
	pcm-loudness-sma-amean	t(166)=-3.33***	PR=0.5, NPR=0.6	d=-0.51
	jitterLocal-sma-amean	t(166)=2.93*	PR=0.1, NPR=0.07	d=0.45
	shimmerLocal-sma-amean	t(166)=2.56*	PR=0.2, NPR=0.18	d=0.39
4. Social Norm Violation	LIWC Social Processes	t(7469)=33.98***	VSN=17.35, NVSN=6.45	d=0.78
	LIWC Biological Processes	t(7469)=12.95***	VSN=4.21, NVSN=1.38	d=0.30
	LIWC Personal Concerns	t(7469)=10.61***	VSN=2.61, NVSN=1.33	d=0.24
	LIWC Swearing	t(7469)=5.85***	VSN=0.49, NVSN=0.11	d=0.13
	LIWC Anger	t(7469)=11.64***	VSN=1.19, NVSN=0.20	d=0.27
	LIWC Power Drive	t(7469)=7.83***	VSN=1.99, NVSN=1.14	d=0.18
	LIWC First Person Plural	t(7469)=3.23**	VSN=0.85, NVSN=0.64	d=0.07
	pcm-loudness-sma-amean	t(7469)=31.24***	VSN=0.69, NVSN=0.56	d=0.72
	F0final-sma-amean	t(7469)=26.6***	VSN=231.09, NVSN=206.99	d=0.61
	jitterLocal-sma-amean	t(7469)=-4.09***	VSN=0.083, NVSN=0.087	d=-0.09
	shimmerLocal-sma-amean	t(7469)=-7.02***	VSN=0.1818, NVSN=0.1897	d=-0.16

Table 2: Complete Statistics for presence of numeric verbal and vocal features in Self-Disclosure (SD)/Non-Self Disclosure (NSD), Shared Experience (SE)/Non-Reference to Shared Experience (NSE), Praise (PR)/Non-Praise (NPR) and Violation of Social Norms (VSN)/Non-Violation of Social Norms (NVSN). Effect size assessed via Cohen's *d*. Significance: \*\*\*: $p < 0.001$ , \*\*:  $p < 0.01$ , \*:  $p < 0.05$

Conversational Strategy	Visual (Speaker) - $\chi^2$ test value - Odds Ratio	Visual (Listener) - $\chi^2$ test value - Odds Ratio
1. Self-Disclosure	Smile - $\chi^2(1,1013)=20.67***$ - $o=1.67$	Smile - $\chi^2(1,1013)=18.63***$ - $o=1.63$
	Gaze (gP) - $\chi^2(1,1013)=93.04***$ - $o=2.39$	Gaze (gP) - $\chi^2(1,1013)=131.34***$ - $o=2.84$
	Gaze (gN) - $\chi^2(1,1013)=35.1***$ - $o=0.49$	Gaze (gN) - $\chi^2(1,1013)=73.23***$ - $o=0.38$
	Gaze (gO) - $\chi^2(1,1013)=173.88***$ - $o=0.29$	Gaze (gO) - $\chi^2(1,1013)=152.12***$ - $o=0.31$
	Gaze (gE) - $\chi^2(1,1013)=120.77***$ - $o=1.8$	Gaze (gE) - $\chi^2(1,1013)=78.92***$ - $o=2.37$
2. Shared Experience	Smile - $\chi^2(1,183)=4.73*$ - $o=1.75$	Smile - $\chi^2(1,183)=7.53**$ - $o=2.07$
	Gaze (gP) - $\chi^2(1,183)=25.37***$ - $o=3.02$	Gaze (gP) - $\chi^2(1,183)=33.36***$ - $o=3.59$
	Gaze (gN) - $\chi^2(1,183)=3.73*$ - $o=0.58$	Gaze (gN) - $\chi^2(1,183)=17.68***$ - $o=0.32$
	Gaze (gO) - $\chi^2(1,183)=27.87***$ - $o=0.31$	Gaze (gO) - $\chi^2(1,183)=16.55***$ - $o=0.41$
	Gaze (gE) - $\chi^2(1,183)=38.13***$ - $o=4.19$	Gaze (gE) - $\chi^2(1,183)=32.45***$ - $o=3.92$
3. Praise	Gaze (gP) - $\chi^2(1,166)=9.94***$ - $o=0.44$	Gaze (gP) - $\chi^2(1,166)=14.22***$ - $o=0.39$
	Gaze (gN) - $\chi^2(1,166)=37.52***$ - $o=4.29$	Gaze (gN) - $\chi^2(1,166)=15.19***$ - $o=0.33$
	Gaze (gO) - N.S	Gaze (gO) - $\chi^2(1,166)=24.23***$ - $o=3.30$
	Gaze (gE) - $\chi^2(1,166)=14.44***$ - $o=0.30$	Gaze (gE) - $\chi^2(1,166)=9.77**$ - $o=0.39$
4. Social Norm Violation	Smile - $\chi^2(1,7469)=871.73***$ - $o=3.35$	Smile - $\chi^2(1,7469)=869.29***$ - $o=3.37$
	Gaze (gP) - $\chi^2(1,7469)=911.89***$ - $o=2.75$	Gaze (gP) - $\chi^2(1,7469)=609.06***$ - $o=2.27$
	Gaze (gN) - $\chi^2(1,7469)=34.82***$ - $o=0.8$	Gaze (gN) - $\chi^2(1,7469)=239.22***$ - $o=0.55$
	Gaze (gO) - $\chi^2(1,7469)=515.26***$ - $o=0.47$	Gaze (gO) - $\chi^2(1,7469)=110.48***$ - $o=0.70$
	Gaze (gE) - $\chi^2(1,7469)=195.17***$ - $o=1.67$	Gaze (gE) - $\chi^2(1,7469)=12.38**$ - $o=1.14$
	Head Nod - $\chi^2(1,7469)=8.06**$ - $o=0.77$	Head Nod - $\chi^2(1,7469)=44.51***$ - $o=0.56$

Table 3: Complete Statistics for presence of binary non-verbal features in Self-Disclosure (SD), Shared Experience (SE), Praise (PR) and Violation of Social Norms (VSN). Odds ratio signals how much more likely is a non-verbal behavior likely to occur in conversational strategy utterances compared to non-conversational strategy utterances. Significance: \*\*\*: $p < 0.001$ , \*\*:  $p < 0.01$ , \*:  $p < 0.05$