We Click, We Align, We Learn: Impact of Influence and Convergence Processes on Student Learning and Rapport Building

Tanmay Sinha
Articulab
School of Computer Science
Carnegie Mellon University
tanmays@cs.cmu.edu

Justine Cassell
Articulab
School of Computer Science
Carnegie Mellon University
justine@cs.cmu.edu

ABSTRACT

Behavioral convergence has been identified as one (largely subconscious) contributor to successful conversations [27], while rapport is one of the central constructs that explains development of personal relationships [6] between these speakers over time. Social factors such as these have been shown to play a potent role in learning. Therefore, in this work, we investigate the relationship in dyadic peer tutoring conversations of convergence, building up of interpersonal rapport over time, and student learning, while positing a novel mechanism that links these constructs. We develop an approach for hierarchical computational modeling of convergence by accounting for time-based dependencies that arise in longitudinal interaction streams, and can thus a)quantify the effect of one partner's behavior on the other and differentiate between driver and recipient (*Influence*), b)extrapolate the outcome of directional influence to determine adaptation in partners' behaviors (*Convergence*). Our results illustrate that influence, convergence and rapport in the peer tutoring dialog are correlated with learning gains and provide concrete evidence for rapport being a causal mechanism that leads to convergence of speech rate in the interaction. We discuss the implications of our work for the development of peer tutoring agents that can improve learning gains through convergence to and from the human learner's behavior.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: Models and Principles—*User/Machine Systems,Human Factors*

Keywords

Convergence; Influence; Rapport; Learning

1. INTRODUCTION

Conversation is like an intricate partner dance, and better alignment between partners can lead to shared understanding, changed beliefs, and increased rapport. Accommodation is a particularly important aspect of such interactions as it may lead not only to the communicative success of the interaction, by decreasing misunderstandings and attaining goals faster [27], but also to its social success by building rapport and affiliation [17]. Communication Accommodation Theory (CAT) [11] illuminates the dynamics underlying an interaction by describing the tendency of individuals to accommodate towards their partners along an adaptation (tendency to become similar over time) - maintenance (no effect) - differentiation (tendency to exaggerate their differences) continuum. These three kinds of accommodation have been demonstrated to arise from a variety of sources, from a tendency to focus on a comparison with others in new social environments, to a consistent kind of non-conscious mimicry [21].

In this work, we examine the nature of one of the forms of accommodation, i.e adaptation (also referred to as alignment, entrainment or convergence), in dyadic peer tutoring conversations over time as a part of our research program on the social infrastructure of learning, with an eye towards implementing more effective educational technologies. We follow prior work [14] in believing that analyses of the cognitive and social aspects of learning are best conducted together for a more complete picture of how to increase learning gains through the use of intelligent systems. In addition, we follow foundational work that studies language as a form of joint action [10], challenging the traditionally held assumption in cognitive psychology of perception, action, and higher-level cognitive processes being best understood by investigating individual minds in isolation. Accordingly, in the current work the dyad is the unit of analysis.

To fully understand what leads conversational partners to converge in their behaviors over time and its impact on learning, we therefore study the dynamics of interaction at a fine (30 second interaction segment) level of granularity. While prior work [12] models such dynamic manifestations in terms of synchrony (partners exhibit temporally or simultaneously similar behaviors) and asynchrony (partners do not modulate their behaviors in tandem), they don't distinguish between the driver and recipient. In their comprehensive synthesis of joint action studies, [30] have emphasized that assessing the mutual influences of two or more actors on each other is an important step toward investigating the mechanisms whereby individuals coordinate their actions. Therefore, to fill this gap in theorizing about the social nature of

INTERPERSONAL '15, November 09-13 2015, Seattle, WA, USA

ACM 978-1-4503-3986-5/15/11 ...$15.00
DOI: http://dx.doi.org/10.1145/2823513.2823516.

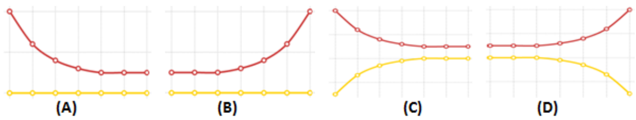


Figure 1: Cases A and B represent unidirectional influence resulting in convergence & divergence. Cases C and D represent bidirectional influence resulting in convergence & divergence

cognition, as a next step, in addition to investigating the presence of convergence, we also operationalize and investigate the direction of influence or which individual affects his interlocutor's behavior. We further show that the direction of influence is important in determining who learns more in a dyad, and perhaps why.

The CAT theory describes how much of interaction occurs as a direct response to the behavior of another person. Because our interest is in the effects of convergence on peer collaborative learning, it is important to examine how partners work together to produce an interaction. This involves understanding how people affect their conversational partners in the interaction. Better modeling of mutual influence during an interaction could bring insights into the design of a virtual peer [7] capable of accommodating and eliciting accommodation, in an attempt to engage in more effective peer tutoring.

Prior work [3] sketch two important hypotheses that clarify the directionality of behavior matching. The "Ideal Input" hypothesis reflects a "uni-directional" dominant view of communication and assumes one of the conversational partners to be a source of ideal behavior towards which the other partner (target) accommodates, while "Least Collaborative Effort" posits that partners "bi-directionally" try to achieve a level of mutual intelligibility while engaging in joint construction of meaning. Other work such as [12] has assumed the latter as a way of describing the mechanism behind convergence in the interaction. In order to computationally quantify and empirically evaluate these mechanisms in our work, we develop representations derived from automatically harvestable low-level linguistic features that highlight convergence and influence, taking into account time-contingency effects, which have largely been ignored in prior computational approaches. Figure 1 presents a brief pictorial description of these interpersonal processes in peer tutoring.

Finally, we investigate the relationship among convergence and influence in longitudinal dyadic interactions, pre-/post-test learning gains, and interpersonal rapport. The power of rapport or being "in-sync" with one's partner during social interactions, has been well documented in the context of learning [25, 32, 33]. In our work, we measure rapport in two ways: a)self report via questionnaire where rapport is broken into its component parts of attentiveness, positivity and coordination, and b)perceived rapport via annotation by naive raters of every 30 second slice of the video of the peer tutoring (and social time in between the peer tutoring), presented to the annotators in a randomized order.

2. RELATED WORK

In this section, we flesh out the rationale for a more careful investigation of interpersonal processes. Recent studies have investigated forms of prosodic entrainment such as coordination in intensity, pitch, voice quality and speaking rate [22], to find cues for engagement in human-human and human-agent interactions. The work of [23] closely follows [22] and is most relevant to our research. Here, the authors found that certain forms of prosodic entrainment such as proximity, convergence and synchrony were correlated with segments of the interaction having similar levels of perceptual and self-reported rapport. While these results are interesting, there are a couple of limitations of the study that we hope to overcome: First, a small number of two minute dialog segments that referred only to solving math problems were manually selected. This ignores the potential effect of "off-task" social talk that didn't pertain to the task but may grease the wheels of the interaction [8]. Here we examine the entire hour-long peer tutoring interaction. Second, their Pearson correlation-based entrainment measures ignored time-based dependency, which we do incorporate in our modeling of influence and convergence. Third, we acknowledge the dynamics of rapport by differentiating between self-reported and perceived rapport when relating them to our measures of accommodation and learning, since these are fundamentally different forms of rapport assessment collected at differing time granularities (at the end of session versus every 30 second). Fourth, in addition to presenting correlational evidence about the impact of entrainment on rapport, we also establish a causal pathway between the two constructs.

The learning science literature provides evidence that convergence is significantly associated with learning. For instance, in [15], linear regression modeling of time (number of conversational exchanges) and syntactic/semantic cues demonstrated that entrainment was positively correlated with learning gain and task success. Other work such as [26] has found that the tendency to adapt to patterns in a partner's utterances predicts collaboration quality and subsequently learning gains. However, the approach to accommodation in prior collaborative learning literature [34] has had a strong cognitive focus - with emphasis on the joint construction of activities that move the group toward problem solving goals and hence knowledge convergence. This work has suggested a cognitive explanation for the impact of convergence on learning, such that it indexes greater shared understanding, and hence leads to improved learning. However, other literature [5] suggests that greater similarity is an index of increased connectedness and interpersonal rapport which, in turn, leads to greater willingness to examine misconceptions, and hence to improved learning. We make an attempt to disambiguate these mechanisms in our work by testing causal relations between perceived rapport and convergence in peer tutoring interactions. We thus add an interactional perspective to the methodological toolkit for examining the effect of accommodation on learning that differentiates between the social and cognitive impacts. Finally, unlike most prior work that has relied on coarse-grained division of the interaction into two or three sub-sessions for investigating longitudinal changes, we zoom in on the conversational dynamics every 30 seconds to study fine-grained attuning of behavior.

3. STUDY CONTEXT

Reciprocal peer tutoring data was collected from 12 American English-speaking dyads (6 dyads were friends and 6 strangers; 6 were 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 on procedural and conceptual aspects of linear equations. Prior work demonstrates that peer tutoring is an effective paradigm that results in student learning [31], making this an effective context to study dyadic interaction and learning. Each session began with social chit-chat, after which the first tutoring period started, followed by another small social interlude, a second tutoring session with role reversal between the tutor and tutee, and then the final social time. For the purposes of the current study, we selected a fairly balanced convenience sample of 15 dyadic conversational sessions (1/4th of our data), in terms of #friend v/s #stranger dyads and session.

4. METHODOLOGY

4.1 Operationalizing Feature Dimensions

We first performed feature extraction from the peer tutoring sessions that had been transcribed and segmented into syntactic clauses. Then, we examined the sequential organization of talk by looking at response pairs. Specifically, for each consecutive 30 second segment in the tutoring sessions, we computed the following features for each speaker: a)# words spoken, b)message density, which is the #independent clauses uttered, divided by the time difference between the first and last utterance within the 30 second segment, c)content density, which is the #characters spoken divided by the #independent clauses uttered, d)#overlaps (a joint event where the two interlocutors speak at once), and e)#laughter. The first three features are representative of speaking rate and two dimensions of burstiness that characterize it. Frequent turn taking will increase the message density, more elaborate or detailed talk between interlocutors will increase the content density, while more #overlaps might potentially indicate a well-coordinated conversation. Laughter may signal enjoyment and affiliation, but also serves to release tension and our data demonstrated examples of what our coders called "nervous laughter" - this is, however, more often the case in early sessions. Prior work on identifying strategies of accommodation in dyadic conversation has included speech rate, overlap (simultaneous speech frequency), laughing and smiling behaviors to facilitate understanding of discourse management [11, 17].

4.2 Operationalizing Joint Constructs

4.2.1 *Quantifying Influence*

To model the causal effects of one partner's behavior on behavior of the other partner over the course of an interaction session, we applied Granger causality [16], widely used in econometrics and neuroscience. This approach is based on asymmetric prediction accuracies of one time series on the future of another. Specifically, a time series X_1 is said to Granger-cause X_2 if the inclusion of past observations (lagged values) of X_1 reduces the prediction error of X_2 in a linear regression model of X_2 and X_1 (unrestricted regression UR), as compared to a model including only the previous observations of X_2 (restricted regression R). In our case, X_1 and X_2 refer to the 30 second sliced feature dimensions for both the partners in a dyad, for instance, #words spoken. Such longitudinal time series interaction data is highly likely to be non-stationary (i.e, is expected to possess a joint probability distribution that changes when shifted in time), and therefore using it in the raw form will lead to spurious and unreliable results.

Fundamentally, time series analysis is based on the assumption that processes generating the data are stationary in time, i.e, they possess no trend (long-term increase or decrease) and seasonality (periodicity). Thus, we first a) detrended the data by removing the best straight line fit, b) performed single exponential smoothing to filter out noise, placing 50% weight on the current data point (smoothed value $s_i{=}\alpha x_i + (1-\alpha)s_{i-1}, \alpha = 0.5, x_i =$raw feature value), c) augmented the Granger causality autoregressive model (where output variable depends linearly on its own previous values) using a lag length of 3, i.e, a time window of 90 seconds (30*3). Intuitively, this means that if we want to test whether time series X_1 Granger-causes X_2, we additionally use feature values from time slices 'i-1', 'i-2' and 'i-3' in X_1 for the unrestricted regression, d)computed statistical significance using an F-test under the null hypothesis that one time series does not granger cause the other ($F(M, n-k)$ $= ((RSS_R - RSS_{UR}) * (n-k))/(M * RSS_{UR})$, where M is the number of lagged X_1 terms and K is the number of estimated in the restricted regression), and finally e)leveraged the F-statistics as granger causality magnitudes (Influence Strength) at 1% LOS.

Thus, unidirectional influence represents the case where only one partner significantly granger-causes the conversational partner (one time series tends to follow the other), while bidirectional influence represents the case when both partners significantly granger-cause each other (affect one another's behavior), along any specific feature dimension.

4.2.2 *Quantifying Convergence*

Following prior literature, we operationalized convergence as the degree to which speakers become similar over the course of the entire conversation. However, we were skeptical about computing it by "traditionally employed" Pearson correlation approaches between time and the absolute difference between a speaker and partner's behavioral feature value at an adjacent turn. The reason is that the correlation measure is not designed to accommodate the interdependent nature of a time point on previous data points in longitudinal dyadic interaction.

Therefore, in our work, following the tests of convergence hypothesis applied in [32] to quantify behavioral convergence, we a)computed the difference in raw behavioral feature values for partner i and partner j engaged in the dyadic conversation for every 30 second slice (call this differenced series y), b)formulated the autoregressive model as $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + ... + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$, where α (constant term) is the drift or change of the average value of the stochastic process, βt is the deterministic time trend and p is the lag length (which is quantified as 3, similar to the prior influence computation), c)tested the presence of unit-root in this time series framework using the Augmented Dickey Fuller (ADF) test at 1% LOS, following proposition 5 [2].

Intuitively, if the ADF test statistic is significant, we reject the null hypothesis that the differenced behavioral time series has a unit root and accept the alternative hypothesis that the variable was generated by a stationary process, which is an evidence for convergence. On the contrary, if ADF test statistic is not significant, we accept the null hypothesis of the presence of a unit root, in turn indicating that the process (change) is not stationary and the definition of convergence is violated.

4.2.3 Index Development

To construct a composite score, the F test statistic for Influence and ADF test statistic for Convergence (call this x) along each feature dimension, were firstly scaled between 0 and 1 using the formula $(x - minimum(x))/(maximum(x) - minimum(x))$, with an intuition to provide transparency and comparability. Secondly, in weighting across features, different feature dimensions were equally weighted (averaged).

4.3 Outcome Measures

4.3.1 Pre-Post Tests

We administered a pre-test to students $before$ the first tutoring session and a post-test $after$ the fifth session (≈ delayed post-test) comprising upto 10 algebra questions. Normalized learning gain for each individual in the dyad was computed by comparing the percentage points gained or lost compared to the maximum possible gainable points from pre- to post-test: (Post-assessment - Pre-assessment)/(100% - Pre-assessment). Then, following a standard approach, the composite learning gain for a dyad was calculated using the average of the individual learning gains. For the 22 subjects used in the current analysis, there was a ceiling effect on pre-test scores. Performing a paired t-test revealed no significant difference in the pre-test and post-test scores (t = 1.6964, df = 21, p < 0.1) [1]

4.3.2 Questionnaires

After each session, both participants in the dyad completed 7 point likert scale (1 = Disagree Strongly; 7 = Agree Strongly) questionnaires, reflecting the dimensions of Attentiveness (3-item scale indexing interest, attention and respectfulness of the partner towards the speaker, Cronbach α =0.42), Positivity (2-item scale indexing friendliness and warmth towards the partner, α =0.72), Coordination (3-item scale indexing whether partners felt in sync, could say everything that they wanted to say and that the interaction was not frustrating, α =0.64), and Long Term Rapport (3-item scale indexing whether the partners felt that they knew each other, were more comfortable and had greater liking compared to the previous interaction session, α =0.78). In addition, the questionnaire asked about Self Efficacy (7-item scale indexing whether the partners thought they were good tutors, learned a lot from tutoring and were concerned about tutoring quality, motivation and impact on the tutee, α =0.5).

4.3.3 Thin Slice Rapport

Person perception research has demonstrated that rapidly-made (< 5 minutes), judgments of others based on brief exposure to their verbal and nonverbal behavior is an extremely accurate assessment of interpersonal dynamics [1]. We employed such "thin-slice" judgments for our work, where raters were provided with a simple definition of rapport and three raters annotated rated every 30 second video segment [1] of the peer tutoring sessions for rapport using a 7 point likert scale (1 to 7). The segments were presented to the annotators in random order so as to ensure that raters were not actually annotating the delta of rapport over the course of the session.

[1]Tests for all dyads were graded twice, except for dyad 7

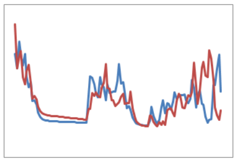
Note that as a part of the thin-slice experimental design (where raters are supposed to be naive), we did not train (and re-train) the raters on how to interpret the criteria for judging rapport and consistently apply the rating scale, and therefore the level of "objectivity" in the annotation process was not very high. Interestingly, for our data, we found that on an average across all dyadic sessions, the "consensus" estimate of inter-rater reliability was 0.37 (as measured by single measures for Intra-class correlation), while the "consistency" estimate of inter-rater reliability was 0.68 (as measured by Cronbach's alpha).

This lack of inter-rater reliability, which is not reported in the thin slice literature, may be caused by the fact that thin slice annotation is usually conducted on one single (or a few) slices from each video, whereas our raters annotated the whole peer tutoring session to serve as ground-truth data for ongoing rapport dynamics, and to inform reinforcement learning algorithms for rapport state updates, for use in our virtual peer tutor. In order to mitigate rater bias and account for label over-use and under-use, we used weighted majority rule with inverse based bias correction [20] to pick a single rapport rating for each 30 second video segment. The inverse based bias correcting rule is given by $1/Freq_i(k)$, where $Freq_i(k)$ is the number of times rater i chooses category k. Finally, to summarize the ratings for each 30 second video segment j, we picked category $k^* \in K$ (total #categories) that maximizes the weighted sum we get when we add up the weights for every individual annotation of j with a given category, using the weights prescribed by the bias correcting rule.

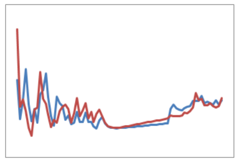
5. ANALYSIS AND RESULTS

First we conducted an explanatory data analysis to illuminate different ways in which influence and convergence of automatically gathered linguistic variables play out. The notions of influence and convergence are illustrated via a visual time series description culled from our conversational data, as illustrated in figure 2. We could successfully distinguish 6 combinations. For instance, Figure 2.a and 2.f depict the two extremes (time-series), clearly illustrating the spatial and temporal difference in the dyadic interaction that can be explained by the underlying processes of influence and convergence.

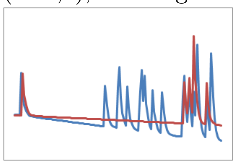
Preliminary results revealed that virtually all dyads and all sessions seemed to converge on message density (roughly, speech rate). This bi-directional accommodation is the kind of classic entrainment that indexes engagement in the interaction - engagement that we also saw indexed with the content of the conversations, particularly during social periods. For content density, on the other hand, influence was more likely to be unidirectional, indicating that when an individual spoke less elaborately, the partner's behavior became less elaborate, and vice versa. For overlaps and laughter, we found that, for both phenomena, more than half of the dyadic sessions showed uni-directional influence while roughly half of the sessions showed presence of convergence, indicating that students exhibited high accommodation across both these phenomena. This result is particularly interesting as laughter and overlap can fairly easily be generated for the human partner to entrain to and can be entrained to in human-agent dialog. In the following sections, we describe correlations (Pearson r, Spearman Rank ρ, Point Biserial correlation r_{pb}) computed in order to find



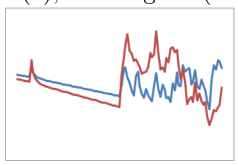
a. ✓✓(0.48,0); Convergence(0.57)



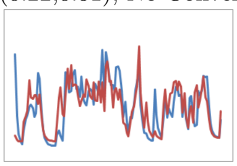
b. ✓(0); Convergence(0.95)



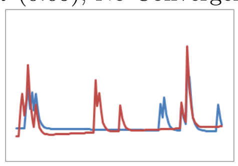
c. ✓✓(0.22,0.51); No Convergence



d. ✓(0.05); No Convergence



e. ✗; Convergence(0.57)



f. ✗; No Convergence

Figure 2: Time Series depicting examples of 6 distinct cases that we can successfully identify from the dyadic conversational sessions. X axis represents the 30 sec sliced timeline. Y axis represents detrended and smoothed feature values. Blue and Red lines represent the 2 conversational partners in the dyad. Notations: ✓✓: bi-directional influence, ✓: uni-directional influence, ✗: no influence. Numeric values in brackets represent Influence/Convergence strength (scaled between 0 and 1)

relationships between our joint constructs and the outcome measures described in section 4.3, while testing significance of the correlation via two tailed t-test. The summarized results are presented in Table 1.

5.1 Effect of Influence and Convergence on Socio-Cognitive Processes

First, at the dyadic level, we correlated our measures of convergence strength with learning. As shown in Table 1, results showed that the maximum convergence strength of a dyad across different sessions was positively associated with their composite learning gains. These results suggest that a virtual peer tutor that both mimics its human partner and produces salient and highly mimicable behaviors may be a more effective learning partner.

We then looked at the association between convergence strength and self-reported rapport. Because the self-reported questionnaire difference metric was dichotomous, we employed Point Biserial correlation measure (r_{pb}) to discern its association with composite learning gains. For self-reported questionnaire ratings x and y provided by two individuals in a dyad after the end of each session, we empirically quantified the difference $d(x, y)$ along any questionnaire dimension as follows: $d(x, y) = 0$, if $|x - y| <= \#questions - 1$ and $d(x, y) = 1$, if $|x - y| > \#questions - 1$, with an intuition to capture at least some reasonable difference in ratings. We found that when the dyad rated the attentiveness questionnaire dimension identically, their convergence strength was in fact lesser.

Next, at an individual level, we observed significant positive correlation between influence strength and learning gains for the influencing partner (correlation here was assessed using Spearman's ρ since a visual plot of the variables revealed that influence strength and learning gains tended to move in the same relative direction, but not necessarily at a constant rate (monotonic relationship)). The CAT literature shows that the lower status person in a conversation is more likely to be the one who accommodates. Perhaps here "higher status" means the tutor who, as the learning science literature has shown, is more likely to learn during a tutoring interaction than is the tutee. In addition, no significant effects for learning gains were found for the influenced partner in the interaction. In line with the previous result, this may be due to the fact that the influenced partner is more likely to be the tutee, who generally (perhaps paradoxically) learns less in peer tutoring than the tutor.

However, interestingly, we found that influence strength had a significant negative correlation with the questionnaire construct of positivity for the influencer (meaning that the influencer felt less friendly and warm towards the influenced) which goes against the CAT literature. Influence strength was also negatively correlated with the questionnaire construct of coordination for both the influencer and partner influenced. We note that looking at these constructs at the individual level may not be as valuable as looking at them in the dyad. For this reason, in Section 5.4 below, we look at these same constructs from a dyadic perspective.

5.2 Effect of Perceived Rapport on Socio-Cognitive Processes

If students are to critique the ideas of their peers, offer tentative suggestions and interpret others' critiques as valuable, they need to trust each other and feel a sense of warmth and belonging. Positive effects of short-term (instant) rapport on enhancing students' math performance has already been demonstrated [18]. However, if we were to build a virtual learning partner that works with a student over a long period of time, that agent would need to adapt its use of socially expressive behaviors with a student over time, increasing and then maintaining rapport, as is done by human peers, as described in [9] and [36].

It was thus legitimate to hypothesize that the deepening rapport in later sub-sessions might be more connected to greater learning. Therefore, to assess the relationship between perceived rapport (as measured by thin slice annotations) and composite learning gains, we divided each hourly dyadic session into 5 equal sub-sessions and considered correlations for the averaged perceived ratings for each sub-session. A linear regression highlighted significant coefficient estimates (coeff=0.22, $p<0.05^*$) for perceived rapport in the fifth sub-session of the interaction when regressed with composite learning gain as the dependent variable. These findings supported our hypothesis that changes in rapport during the final segments of an interaction were significantly predictive of learning gains.

Given these and results in the prior subsection, at this point, it becomes essential to think about the interplay between convergence, rapport and learning. Does convergence correlate with better learning in the dyad via shared cognitive representations, because shared representations bring about better collaborative learning ("cognitive hypothesis"), or, is convergence associated with better learning due to the impact of rapport, because students learn better when paired with students they like and feel rapport with ("social hypothesis")? While we know from literature (discussed in

Table 1: Results of Correlational Analysis at the session level (significance assessed via 2-tailed t-test: +:p=0.06, *: p<0.05, **: p<0.01, ***: p<0.001)

Construct (both continuous)	**Pearson's r or Spearman's ρ**
1) Convergence Strength in a dyadic session (maximum across different interaction sessions) and Composite Learning Gain	$r = 0.57^{+}$
2) Influence Strength in a dyadic session (averaged across different interaction sessions) and Individual Learning Gain for Influencer	$\rho = 0.44^{*}$
3) Avg Influence Strength in a dyadic session and Avg Positivity for Influencer (across different interaction sessions)	$r = -0.6^{**}$
4) Avg Influence Strength in a dyadic session and Avg Coordination for Influencer (across different interaction sessions)	$r = -0.44^{*}$
5) Avg Influence Strength in a dyadic session and Avg Coordination for Influenced (across different interaction sessions)	$r = -0.75^{***}$
Construct (1 dichotomous, 1 continuous)	**Point Biserial r_{pb}**
1) Difference in Attentiveness in a dyadic session and Convergence Strength in a dyadic session	$r_{pb} = 0.57^{*}$
2) Difference in Attentiveness in a dyadic session and Composite Learning Gain	$r_{pb} = -0.51^{*}$
3) Difference in Coordination in a dyadic session and Composite Learning Gain	$r_{pb} = -0.55^{*}$
4) Difference in Self-Efficacy in a dyadic session and Composite Learning Gain	$r_{pb} = -0.59^{*}$
5) Difference in Attentiveness in a dyadic session and Thin Slice Rapport (5th segment of interaction session)	$r_{pb} = -0.43^{+}$
6) Difference in Coordination in a dyadic session and Thin Slice Rapport (4th segment of interaction session)	$r_{pb} = -0.47^{+}$
7) Difference in Coordination in a dyadic session and Thin Slice Rapport (5th segment of interaction session)	$r_{pb} = -0.63^{**}$
8) Difference in Self-Efficacy in a dyadic session and Thin Slice Rapport (5th segment of interaction session)	$r_{pb} = -0.54^{*}$

related work) that the former is a plausible explanation for convergence, we were interested in finding a strong causal support for the "social hypothesis".

5.3 Interplay: Convergence & Rapport

We sought to investigate whether rapport leads to convergence in the longitudinal interaction. Mediation analysis [24] was one of the potential choices for assessing whether convergence statistically mediated the predictive (correlational) relationship between rapport and learning. However, since a)we were interested in testing for causation rather than correlation and, b)perceived rapport, convergence (∼time series) and learning (∼aggregated measure) are of fundamentally different nature and granularity, this technique was inappropriate for our scenario. After careful consideration, we decided to test for causality between perceived rapport and convergence via granger causality leveraging a similar procedure applied to quantify influence in section 4.2.1.

Specifically, to perform the causal analysis, we: a)formalized convergence time series (say T_1) as the difference in detrended and smoothed low level feature value for the two students in every dyad at lag 0 (e.g, #words spoken), for every 30 second segment in the hourly session, b)formalized rapport time series (say T_2) as the detrended and smoothed value of thin slice rapport, c)inferred whether T_2 significantly granger-causes T_1 (at 5% LOS) leveraging a similar procedure applied to quantify influence in section 4.2.1 (using a time window of 90 seconds in the autoregressive model formulation, meaning that we additionally looked at the rapport rating for time slices 'i-1', 'i-2' and 'i-3' in inferring about convergence at time slice 'i' apart from just using convergence information from time 0 to 'i-1').

We found that for 13/15 sessions used in our study, rapport significantly led to convergence in message density (that indexes speech rate in the interaction). To evaluate the robustness of our findings, we also altered the definition of convergence time series T_1 as the difference in detrended and smoothed low level feature value for the two students in every dyad at lag 1 and 2 respectively, the intuition being to include a time contingency effect for similarity in behaviors over time. However, the causal effects for rapport on convergence in message density were still significantly present for 13/15 dyadic sessions. The second strongest causal effects were found for #words (9/15 sessions) and #overlaps (8/15 sessions), meaning that increased rapport led to similarity in #words spoken and overlapping expressions used. Interestingly, by reversing the direction of causality and testing whether convergence and rapport work together in a feedback loop (rapport ⇔ convergence) to help in interaction regulation, we found very low support (roughly $1/3^{rd}$ sessions) for significant causal effects of convergence on rapport in the dyadic sessions, along any of the feature dimensions.

These results, derived from our fine-grained causal analysis, substantiate the crucial role of rapport as an influencing social mechanism that leads to behavioral convergence in the interaction - convergence, which we showed above to be associated with learning. This has direct implications for the development of virtual companions that can improve students' performance by establishing positive relationship with them and in turn also potentially triggering verbal and non-verbal behavioral convergence. Such a form of social facilitation and socially adaptable behavior generation by a virtual agent has been shown to have a positive effect on student effort and performance on math tasks [18].

5.4 Effect of Self-Reported Rapport on Socio-Cognitive Processes

Finally, we looked at how the self-reported rapport questionnaires correlated to learning gains. At a dyadic level, we found that smaller values on the dichotomous difference ($d(x, y)$) measure for attentiveness, coordination and self efficacy (i.e, 0) were associated with higher average learning gains (significant negative correlation), meaning that when the dyad rated their level of attentiveness, coordination and self efficacy identically, they were likely to learn more. This means that convergence in assessment of rapport of one another is correlated with increased learning gains.

Partners with similar levels of reported self-efficacy are likely to feel more conscious ("I was afraid that I might not understand how to tutor", "I was worried that I might not understand the math as much as I'd like", "It was important that I show that I know the math in front of my partner"), engrossed ("I think I learned a lot from tutoring", "I wanted my partner to learn as much as possible") and motivated ("my fear of my partner performing poorly is often what motivated me") during the tutoring process and hence learn more ("I think I was a good tutor"). Since there wasn't enough variation in self reported positivity ratings for individuals in the dyad, we could not find any significant relation between the difference in positivity and composite learning gains. Overall, these results provide further evidence for in-sync behavior characterized by mutual responsiveness and receptivity, which in turn enhances coordination and leads to improved learning gains.

Lastly to close the loop, we tested whether difference between the scores given by both individuals on their questionnaire ratings for the whole session was associated with average perceived rapport for each sub-session. Results showed that difference in attentiveness and coordination scores for a session were negatively correlated with average perceived rapport for the fourth and fifth interaction segment ($\approx$ last 25 minutes), while difference in self-efficacy scores for a session were negatively correlated with average perceived rapport for the fifth interaction segment ($\approx$ last 12 minutes). This means that when both partners rated attentiveness, coordination and self-efficacy identically, their average perceived rapport (annotated by independent raters) was higher for the final two interaction segments of that session. Overall these results a)substantiate the importance of mutual attentiveness as an important way in which interactants learn enough about the other so as to adapt behavioral expectations and build rapport, b)highlight the goal of coordination as a path to rapport by signaling that as common ground increases between interlocutors, mutual responsiveness and behavioral synchrony become more prominent [36].

6. DISCUSSION AND CONCLUSION

In this paper, we developed principled approaches to capture the synchrony of communicative behaviors by modeling joint constructs of influence and convergence, which in turn exhibited strong correlations with important outcome measures such as learning gains and interpersonal rapport. By allowing for a fine-grained representation of dyadic indices that accounted for time-based dependencies in the longitudinal interaction, we avoided false assumptions about behavioral independence of conversational partners. In essence, we learned that: a)Influence and Convergence have a positive effect on learning - a virtual peer tutor that both converges to its human partner and invites convergence may be a more effective learning partner, b)Deepening rapport in final segments of the interaction is connected to greater learning - a virtual peer that elicits increasing relational closeness from the learner may be a more effective learning partner, c)There is a significant causal effect of rapport on convergence - a virtual peer that builds rapport might lead to students speaking and behaving like the virtual peer. Therefore, in sum, a virtual peer that can establish rapport with students and in turn trigger behavioral convergence can improve learning gains.

In order to explain these results, one might turn to the following mechanism. Learning can be explained as the side effect of cognitive [13] and social [19] processes triggered by the conversational interactions (explanation, argumentation, mutual regulation etc) students engage in so as to develop shared understanding. The collaborative learning literature [29] suggests that the effort necessary to build shared understanding [34] is what actually leads to learning. In parallel, the entrainment literature provides theoretical evidence [27] of convergence (alignment processes) being one of the important indices of shared understanding between interlocutors, allowing them to sufficiently reconstruct the meaning of the interaction. Moreover, the Interactive Alignment Model [27] also posits that such shared mental representation is caused by greater similarity in observable low-level behaviors (for e.g - lexical, acoustic, prosodic levels) and subsequently in internal (higher level semantic) representations. Interactive priming [4], which links these neighboring levels of representation, has been described as one of the underlying mechanisms for observable convergence. Empirically [28] too, priming has been utilized for operationalizing convergence in the dialog as a visible measure of shared mental models, which was in turn shown to be positively associated with learning [15, 35]. Thus, to be clear, in the results described above, we are positing a novel mechanism for learning in situations of interpersonal closeness by positing a relationship among rapport, convergence and learning. When there is rapport, it leads to convergence in the interaction - this social phenomenon of convergence causes shared mental representation, which in turn leads to learning.

The goal of this work is to have a roadmap for integrating convergence into our dialog-based reciprocal peer tutoring virtual agent, in such a way as to evoke alignment, and to detect and remedy decreasing alignment, between the tutor and student in real time. Foreseeable next steps include working on a)matching the tutee's problem solving pace and regulating flow of the interaction through adjusting the balance between message density and content density, b)improving tutor-tutee alignment by entraining on overlapping behavior to signal acknowledgment or understanding of what the tutee says, c)predicting learning outcomes based on current level of influence and convergence, so as to provide early scaffolding as opposed to delayed scaffolding at the end of the entire interaction, d)differentiating friendship status (friends/strangers), tutor/tutee and gender based entrainment effects. Once these features have been integrated, the effects on learners will be evaluated.

7. ACKNOWLEDGMENTS

This work was supported by a graduate fellowship from the R.K. Mellon Foundation and a Focused Award from Google on improving adaptive learning technologies.

8. REFERENCES

[1] N. Ambady and R. Rosenthal. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, 111(2):256, 1992.

[2] A. B. Bernard and S. N. Durlauf. Interpreting tests of the convergence hypothesis. *Journal of econometrics*, 71(1):161–173, 1996.

[3] H. Bortfeld and S. E. Brennan. Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes*, 23(2):119–147, 1997.

[4] H. Branigan, M. Pickering, and A. Cleland. Syntactic co-ordination in dialogue. *Cognition*, 75(2):B13–B25, 2000.

[5] W. Cade, N. Dowell, A. Graesser, Y. Tausczik, and J. Pennebaker. Modeling student socioaffective responses to group interactions in a collaborative online chat environment. In *EDM*, 2014.

[6] J. N. Cappella. On defining conversational coordination and rapport. *Psychological Inquiry*, 1(4):303–305, 1990.

[7] J. Cassell, M. Ananny, A. Basu, T. Bickmore, P. Chong, D. Mellis, K. Ryokai, J. Smith, H. Vilhjálmsson, and H. Yan. Shared reality: physical collaboration with a virtual peer. In *CHI'00 extended*

abstracts on Human factors in computing systems, pages 259–260. ACM, 2000.

[8] J. Cassell and T. Bickmore. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction*, 13(1-2):89–132, 2003.

[9] J. Cassell, A. J. Gill, and P. A. Tepper. Coordination in conversation and rapport. In *Proceedings of the workshop on Embodied Language Processing*, pages 41–50. ACL, 2007.

[10] H. H. Clark. *Using language*. Cambridge university press, 1996.

[11] J. Coupland, N. Coupland, and H. Giles. Accommodation theory. communication, context and consequences. *Contexts of Accommodation. Cambridge & Paris: Cambridge University Press & Éditions de la maison des sciences de lâĂŹhomme*, pages 1–68, 1991.

[12] C. De Looze, S. Scherer, B. Vaughan, and N. Campbell. Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, 58:11–34, 2014.

[13] P. Dillenbourg and P. Jermann. Designing integrative scripts. In *Scripting computer-supported collaborative learning*, pages 275–301. Springer, 2007.

[14] S. K. D'Mello, A. C. Strain, A. Olney, and A. Graesser. Affect, meta-affect, and affect regulation during complex learning. In *International handbook of metacognition and learning technologies*, pages 669–681. Springer, 2013.

[15] H. Friedberg, D. Litman, and S. B. Paletz. Lexical entrainment and success in student engineering groups. In *Spoken Language Technology Workshop (SLT), 2012 IEEE*, pages 404–409. IEEE, 2012.

[16] C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.

[17] E. Jones, C. Gallois, V. Callan, and M. Barker. Strategies of accommodation: Development of a coding system for conversational interaction. *Journal of Language and Social Psychology*, 18(2):123–151, 1999.

[18] B. Karacora, M. Dehghani, N. Krämer-Mertens, and J. Gratch. The influence of virtual agents' gender and rapport on enhancing math performance. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, pages 563–568, 2012.

[19] K. Kreijns, P. A. Kirschner, and W. Jochems. Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in human behavior*, 19(3):335–353, 2003.

[20] J. Kruger, U. Endriss, R. Fernández, and C. Qing. Axiomatic analysis of aggregation methods for collective annotation. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1185–1192. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

[21] J. L. Lakin and T. L. Chartrand. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological science*, 14(4):334–339, 2003.

[22] R. Levitan and J. B. Hirschberg. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. 2011.

[23] N. Lubold and H. Pon-Barry. Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, pages 5–12. ACM, 2014.

[24] D. P. MacKinnon. Mediation analysis. *The Encyclopedia of Clinical Psychology*, 2008.

[25] A. Ogan, S. Finkelstein, E. Walker, R. Carlson, and J. Cassell. Rudeness and rapport: Insults and learning gains in peer tutoring. In *Intelligent Tutoring Systems*, pages 11–21. Springer, 2012.

[26] R. Pea. The effect of mutual gaze perception on students' verbal coordination. In *Educational Data Mining 2014*, 2014.

[27] M. J. Pickering and S. Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(02):169–190, 2004.

[28] D. Reitter, F. Keller, and J. D. Moore. Computational modelling of structural priming in dialogue. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 121–124. Association for Computational Linguistics, 2006.

[29] D. L. Schwartz. The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences*, 4(3):321–354, 1995.

[30] N. Sebanz, H. Bekkering, and G. Knoblich. Joint action: bodies and minds moving together. *Trends in cognitive sciences*, 10(2):70–76, 2006.

[31] A. M. Sharpley, J. W. Irvine, and C. F. Sharpley. 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, 1983.

[32] T. Sinha and J. Cassell. Fine-grained analyses of interpersonal processes and their effect on learning. In *Artificial Intelligence in Education*, pages 781–785. Springer, 2015.

[33] T. Sinha, R. Zhao, and J. Cassell. 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, 2015.

[34] S. D. Teasley, F. Fischer, A. Weinberger, K. Stegmann, P. Dillenbourg, M. Kapur, and M. Chi. Cognitive convergence in collaborative learning. In *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 3*, pages 360–367. International Society of the Learning Sciences, 2008.

[35] A. Ward and D. Litman. Dialog convergence and learning. *Frontiers in Artificial Intelligence and Applications*, 158:262, 2007.

[36] R. Zhao, A. Papangelis, and J. Cassell. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *Intelligent Virtual Agents*, pages 514–527. Springer, 2014.