

Towards Emotion-Aware Agents For Negotiation Dialogues

Kushal Chawla¹ Rene Clever² Jaysa Ramirez³ Gale Lucas¹ Jonathan Gratch¹

¹University of Southern California, Los Angeles, USA

²CUNY Lehman College, Bronx, USA ³Rollins College, Winter Park, USA

¹{chawla, lucas, gratch}@ict.usc.edu ²rene.clever@lc.cuny.edu ³jramirez@rollins.edu

Abstract—Negotiation is a complex social interaction that encapsulates emotional encounters in human decision-making. Virtual agents that can negotiate with humans are useful in pedagogy and conversational AI. To advance the development of such agents, we explore the prediction of two important subjective goals in a negotiation – outcome satisfaction and partner perception. Specifically, we analyze the extent to which emotion attributes extracted from the negotiation help in the prediction, above and beyond the individual difference variables. We focus on a recent dataset in chat-based negotiations, grounded in a realistic camping scenario. We study three degrees of emotion dimensions – *emoticons*, *lexical*, and *contextual* by leveraging affective lexicons and a state-of-the-art deep learning architecture. Our insights will be helpful in designing adaptive negotiation agents that interact through realistic communication interfaces.

Index Terms—negotiation dialogues, human-agent interaction, emotion models, outcome prediction

I. INTRODUCTION

Negotiation is a core task for studying emotional feelings and expressions in human decision-making [1]. Being a mixed-motive task, it creates both interpersonal and intrapersonal conflicts for the negotiators. Motivational tensions often arise when negotiations pit aspirations for individual accomplishments against the demands of sustaining social connections. This leaves difficult decisions for the negotiators about working towards their own self-oriented outcomes or making sacrifices for others. Such situations can be fraught with emotional encounters. For instance, as a negotiator strives to get as much as possible for themselves, in doing so, they need their partner to go along as well. However, the partner's willingness to go along is essentially an emotional decision. A negotiator that tries to take too much can annoy their partner, and in turn, hurt their likeability in the eyes of their partners and also the partner's affective evaluation of the outcome (that is, their satisfaction). Instead, it is desirable for the negotiator to strive for maximum performance while ensuring that the partner is satisfied [2] and leaves with a positive perception of the partner [3, 4]. Therefore, predicting the partner's satisfaction and perception in advance can be

crucial for an AI assistant that aims to negotiate with its users. These virtual negotiation agents find broad applications in advancing conversational AI such as the Google Duplex agent that books a haircut appointment over the phone [5]. Further, such agents can also be useful in increasing access to social skills training [6].

A number of prior studies rely on user individual difference attributes to explain negotiation behaviors and outcomes. For instance, this includes studies based on gender [7] and personality traits such as Social Value Orientation and Machiavellianism [8]. This research is crucial from the perspective of practical negotiation agents that interact with users on social media platforms since research suggests that demographic and personality attributes can be inferred from past user interactions [9]–[11]. Thus, they can be used to adapt the behavior of the deployed automated systems. However, relying only on these variables misses out on all the other available information such as affective attributes in the negotiation itself, which may further help in predicting the outcomes. While one might expect that such affect variables would help merely because they are manifestations of the individual differences, it is instead possible that affective factors reflect the recent interaction with the partner and might predict in their own right. This leads us to our key research question in this work: *For predicting a negotiator's satisfaction and liking for their partners, is there value in leveraging affective attributes from the negotiation, above and beyond the individual difference variables?*

Most prior efforts in negotiations focus on a menu-driven communication between the human and the agent. Such interfaces allow button clicks for sharing offers or preferences, which helps to keep the design tractable. This concreteness, however, comes at a cost – it hinders the analysis of several realistic aspects in a negotiation such as persuasion and free-form emotion expression. Due to the inherent design restrictions, emotion expression merely reduces to the use of emoticons and predefined sentence templates [12]. Instead, recent efforts have explored a fundamentally different, chat-based interaction that allows human-agent communication in free-form natural language such as English [13, 14], having the potential to fill the gap between existing emotion research and the challenges in real-world negotiations.

To this end, we analyze a large-scale, linguistically rich dataset [14] that is grounded in a realistic campsite negotiation

Research was sponsored by the National Science Foundation and Army Research Office under Cooperative Agreement W911NF-20-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

task. The associated metadata about the participants enables us to study individual differences and linguistic dialogue behaviors in the same setup. Going beyond the capabilities of restrictive menu-driven interfaces, we develop three degrees of emotion recognition techniques: *emoticons*, *lexical*, and *contextual* emotion by leveraging recent advancements in deep learning (Section III). Using correlational and step-wise regression analysis, we quantify the extent to which affective attributes help to account for the variance in the outcomes, beyond just the individual user attributes (Section V). We further discuss how our analysis can guide the development of automated negotiation agents and briefly summarise the ethical considerations around such systems (Section VI).

II. RELATED WORK

How humans negotiate has been extensively studied in multiple disciplines: Game Theory for identifying optimal behaviors [15], Psychology for understanding human decision-making [16], and Computer Science for building automated negotiation agents [17, 18]. Demographics, personality and affective behaviors have played a central role in improving our understanding around negotiations [7, 8] and similar decision-making social interactions [19, 20]. For instance, de Melo and colleagues studied human-agent prisoner’s dilemma interaction and found that the display of guilt after exploitation can elicit more cooperative responses as compared to an agent that smiled [20].

Both satisfaction and the relationship with the negotiation partner are seen as crucial metrics of negotiation performance. Oliver and colleagues studied how profit expectations relate to satisfaction after the bargaining in buyer-seller interactions, finding that higher expectations had the effect of decreased satisfaction [2]. Maintaining a positive relationship with the partner is especially crucial in repeated interactions, where poor relations in earlier negotiations can adversely impact the results of future ones [4]. Relationship also manifests in the context of rapport building [21], favor exchange [22], and reputation effects [23].

However, most efforts in human-agent negotiations are based on highly structured menu-driven designs such as the IAGO negotiation platform [12]. In many cases, the negotiation is seen as simply an exchange of offers [24]. Towards more realistic modes of communication, limited efforts have studied chat-based dialogue systems, that are fundamentally different than menu-driven systems and hence, must be studied separately. Initial datasets in this direction looked at game settings [25], which allow free-form language use but the lack of any semantic context in such negotiations inhibits any rich personal conversations. He and colleagues focused on more realistic buyer-seller price negotiations for the product listings on Craigslist [13]. Unfortunately, none of these datasets incorporated the demographics or the personality of the users. Further, all of them looked only at the objective outcomes of the negotiation performance such as the points scored or the final agreed price. Our focus is on the recently released *CaSiNo* dataset that enables linguistically rich conversations in

a constrained, tractable environment [14]. The dataset further contains the metadata of the participants, that we use to develop our measures in this work (Section III).

Finally, we note that researchers have also looked at face-to-face negotiations [26], including speech and embodied agents that can be natural extensions to our current analysis in chat-based negotiations.

III. DATASET AND METHODS

Our analysis is based on the recently released *CaSiNo* dataset [14]. The design is an instance of Multi-Issue Bargaining Task (MIBT), which is a popular model for studying negotiations computationally [12, 27]. MIBT provides a concrete structure to the negotiation and thus, it helps to keep the conversations focused. Further, the task is grounded in a real-world campsite scenario that promotes rich personal conversations in free-form natural language. Unlike prior datasets in chat-based negotiation systems [13, 25, 28], this design enables the study of a number of realistic aspects in negotiations such as persuasion and emotion expression in a constrained, tractable environment.

A. Participants

A total of 846 unique participants were recruited on Amazon Mechanical Turk for collecting the *CaSiNo* dataset. The participant pool was restricted to the United States. They were paid for their time and also with a performance based bonus. Among the participants, 472 identified themselves as *Female*, 372 were *Male*, and 2 belonged to *Other* category. 73.9% of the participants were *White American*, 1.5% were *Native or Islander*, 8.7% were *Asian American*, 8.6% were *Black or African American*, 4.1% were of *Hispanic or Latino* origin, and remaining 3.1% belonged to *Other* category.

B. Design and Procedure

The *CaSiNo* dataset contains a total of 1030 negotiation dialogues in English. In each dialogue, two participants take the role of campsite neighbors and negotiate for additional packages of camping essential items. Following the MIBT structure, there are three issues: *Food*, *Water*, and *Firewood*. Each issue has an available quantity of three, thereby resulting in four *levels* from 0 to 3. Before the negotiation begins, each participant is randomly assigned a priority order for the three issues, which is a permutation of $\{High, Medium, Low\}$. Once the priorities are assigned, the participants also come up with their own justifications for needing or not needing a particular issue such as needing more water for a planned hike or food for the kids. This exercise *prepares* the participants for their upcoming negotiation, and subsequently results in the collection of linguistically rich conversations.

The dialogues in *CaSiNo* contain 11.6 utterances with 22 tokens per utterance, on average. The participants were also allowed to use emoticons for four basic emotions: *Joy*, *Sadness*, *Anger*, and *Surprise*. We provide a screenshot from the chat interface in Figure 1. Once the negotiation is over, the researchers record the performance of the participants based on

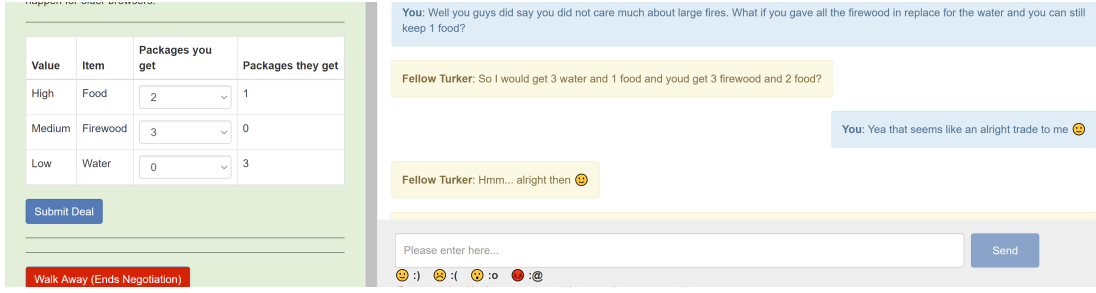


Fig. 1: Screenshot from the CaSiNo data collection interface. Taken from [14].

the number of items that they are able to negotiate for. Alongside, the participants are also asked to report their satisfaction with the outcome, and their perception of the negotiation partner. Analyzing these subjective measures of negotiation performance is the primary focus of this work. The dataset further contains the demographics and personality traits for the participants, which are collected prior to the negotiation. We refer the readers to the original CaSiNo paper [14] for more information on the data collection procedure and a sample dialogue from the dataset.

C. Measures

We cluster the variables into three categories: 1) *Individual difference*, which captures the demographics and personality of the participants, 2) *Affect Variables*, which comprises all the affective attributes extracted from the negotiation utterances, and 3) *Negotiation Outcomes*, where we describe the two primary dependent variables in our analysis.

1) *Individual Difference*: These variables are based on the responses reported by the participants in a survey before their negotiation. The participants self-identified their demographic attributes while the personality traits are based on the standard tests from the psychology literature.

Demographics: There are two continuous variables: *Age*¹ and *Education*. We encoded *Education* as a continuous variable, leveraging the inherent order in the highest level of education. It takes a value from 0 to 8, with an increasing level of education.

Further, there are two discrete demographic variables: *Gender* and *Ethnicity*. For our regression analysis, we dummified these variables based on the categories discussed earlier.

Personality: There are two available measures of individual personality differences: Social Value Orientation [29] and the Big-5 personality traits [30]. These personality attributes have been extensively studied in the context of negotiation research [31, 32]. The Social Value Orientation (SVO) is defined as the stable preferences for certain patterns of outcomes for oneself or others [33]. A participant can either be categorized as *Prosocial*, that tend to be cooperative in their interactions, or *Proself*, that tend to serve their individual interests. In the CaSiNo data collection study, SVO was computed using

the Triple Dominance Measure [29]. 463 participants were classified as *Prosocial*, 364 as *Proself*, while 19 were *Unclassified*. The Big-5 test consists of five personality dimensions: *Extraversion*, *Agreeableness*, *Conscientiousness*, *Emotional Stability*, and *Openness to Experiences*. These dimensions were computed using the Ten-Item Personality Inventory [34]. Each dimension takes a value between 1 and 7.

2) *Capturing Affect*: Natural language based negotiations provide exciting avenues for Affective Computing research. Unlike menu-driven systems, where communication is highly restricted, language allows free expression of emotion and other affective attributes. In order to capture these attributes in CaSiNo, we consider three different degrees of affect recognition techniques: emoticon counts, use of emotive vocabulary based on affect lexicons (specifically LIWC [35]), and utterance-level emotion expressions (based on a pretrained deep learning model called T5 [36]). We discuss these approaches below. We will later compare these methods through our regression analysis for predicting satisfaction and likeness.

Emoticons: Emoticons provide a structured way for emotion expression. Due to this reason, emoticons have been heavily used in menu-driven negotiation research. For instance, the IAGO platform for human-agent negotiations allows human participants to use the emoticons via button clicks [12]. For CaSiNo, the participants were allowed to use shorthands for typing emoticons directly in the chat interface for four basic emotions: *Joy*, *Sadness*, *Anger*, and *Surprise* (see Figure 1). For a given participant and the negotiation dialogue, we count the number of emoticons used by the participant for each of these types and use that as a continuous measure of emotion expression. Approximately 15% of the utterances in CaSiNo make use of one or more emoticons. Among those that use it, nearly 80% use *Joy* emoticon.

Based on a manual inspection of the utterances, we find that the participants tend to use *Joy* emoticon in a number of scenarios, including small talk at the beginning or end of the conversation when showing agreement, and in some cases, interestingly, also when rolling out a strict offer. *Sadness* tends to be used when showing disagreement, showing empathy, or emphasizing personal needs. Some cases where the participants expressed *Surprise* are when pointing out strange behavior from their partner or showing empathy towards a specific need of their partner. *Anger* is used in the cases of

¹One participant reported the age of 3, which we believed to be in error and was removed from all our analysis that uses *Age*.

strong disagreement where for instance, the partner does not empathize with the personal requirements of the participants or when they receive a very unfair offer that largely benefits the negotiation partner.

LIWC: We now go beyond the capabilities of menu-driven systems by extracting emotion attributes from the text in the utterances of the participants. Specifically, we look at word-level expressions by leveraging affect lexicons from the literature. Specifically, we make use of the Linguistic Inquiry and Word Count lexicon, popularly referred to as LIWC in the literature [35]. LIWC consists of a number of word vocabularies corresponding to everyday use that reflect our feelings and personality. Given our focus on affective attributes, we use four LIWC dimensions in this work: *Positive Emotions*, *Sadness*, *Anger*, and *Anxiety*. To compute the corresponding scores for a participant in a given dialogue, we simply count the number of words in the utterances of the participant that are present in the corresponding LIWC lexicons.

As one might expect, we find a number of instances where the participants express emotions without explicitly making use of emoticons. For instance, an excerpt where the participant clearly expresses anger without using the *anger* emoticon: ‘Are you mad. with out water what we do...’. Hence, with LIWC lexicons, we are able to capture more emotive utterances than by simply looking at the emoticon usage. Nearly 60% of the utterances make use of at least one emotive word from our four LIWC categories.

Since LIWC variables are based on word usage, they provide a highly interpretable way of extracting emotions directly from the text. However, merely relying on word-level traits can still miss out on phrase-level or utterance-level emotion expressions. Hence, we next discuss how we make use of a state-of-the-art deep learning model from Natural Language Processing (NLP) to identify emotion in CaSiNo dialogues.

Utterance-level emotion: Large deep learning models based on the highly parallelizable Transformer architecture have gained enormous popularity in the field of NLP in recent years [37]. These models have achieved huge success on a variety of tasks ranging from sentiment analysis to machine translation, fundamentally changing the way researchers approach these tasks. The primary way of training a model for a specific task consists of two steps: 1) *Pre-train* a large language model with freely available text data on the web in an unsupervised or a self-supervised manner, and then 2) *Fine-tune* the model on the task-specific dataset based on supervised learning.

Leveraging these recent advancements, we make use of the Text-to-Text Transfer Transformer (T5) model [36]. T5 is an encoder-decoder architecture that is pretrained on a variety of NLP tasks by converting all the tasks in the same text-to-text format. No matter the task, the input to the model is a text sequence along with a task identifier prefix, and the target output is also a text sequence. Classification tasks can also be converted in this way by simply using a target output sequence that consists of a single word corresponding to the ground-truth label. This approach swiftly unifies a number of tasks,

paving the way for parameter sharing throughout the model architecture. T5 achieves the state-of-the-art performance on a number of NLP tasks such as classification, text summarization, and question answering.

For our purpose, we first pre-process the utterances in CaSiNo by removing all the emoticons. We then use the publicly available T5 model² for making predictions. This model is *fine-tuned* on an emotion recognition dataset based on social interactions on Twitter [38]. Given an input text sequence, the model classifies it into one of the six classes: *Joy*, *Love*, *Sadness*, *Fear*, *Anger*, and *Surprise*. The model performs well on the test set of the emotion data, achieving an accuracy of 93% and a macro F1 score of 90%.

The output prediction of the model is based on the confidence scores that it generates for all the emotion labels, such that the label with the highest confidence is chosen as the prediction. In our case, since we are primarily interested in extracting the emotion scores for the utterances, we do not rely on the final predictions of the model. Instead, we directly use these confidence scores (probability values between 0 and 1) as the corresponding scores for each label. Given a negotiation dialogue and a participant, we simply sum up these confidence scores for all the participant’s utterances and use these as the continuous measures for the six emotion dimensions. We refer to these dimensions (and interchangeably, this method) as *T5-Emotion* in this paper.

All the three methods described above capture the emotion expression in CaSiNo dataset, in fundamentally very different ways. To gain further insight into these methods, in Section IV, we will analyze the relationship between them through a combination of quantitative and qualitative techniques. First, we complete our description for measures by elaborating on the subjective negotiation outcomes.

3) *Negotiation Outcomes:* We focus on two outcome variables: 1) Satisfaction (How satisfied are you with the negotiation outcome?), and 2) Likeness (How much do you like your opponent?). Both variables were self-reported by the participants after the negotiation, using a 5-point Likert scale. We coded the responses from 1 to 5 to create the continuous measures. We present the means and standard deviations for all the continuous variables in Table IV.

IV. ANALYSIS OF EMOTION VARIABLES

Before presenting our results for predicting participant satisfaction and likeness, we first validate *T5-Emotion* measures by understanding how they relate to *Emoticon* and *LIWC*.

A. Correlation among emotion dimensions

Table I presents the correlations of *T5-Emotion* measures among each other and with *Emoticons* and *LIWC* dimensions. For this analysis, we only focus on the dimensions that appear in all three models: *Joy*, *Sadness*, and *Anger*. For *LIWC*, we incorporate the *Positive* dimension. In general, we find significant positive correlations of the emotion dimensions with

²<https://huggingface.co/mrm8488/t5-base-finetuned-emotion>

TABLE I: Correlations of *T5-Emotion* variables (a) among each other, and (b) with corresponding *Emoticon* and *LIWC* variables. We only consider the dimensions that are common across all methods. * denotes $p < 0.05$. ** denotes $p < 0.01$.

(a)				(b)						
	Joy	Sadness	Anger	T5-Emotion	Emoticon Usage			LIWC Variables		
Joy	1	-.288**	-.295**	Joy	.062**	-.043	-.040	Positive Emotions	.387**	-.125**
Sadness	-.288**	1	.277**	Sadness	-.010	.213**	.086**	Sadness	-.125**	-.106**
Anger	-.295**	.277**	1	Anger	-.052*	.113**	.111**	Anger	.550**	.033
									-.185**	.114**
										.169**

TABLE II: Top 5 words based on log-odds ratio for (a) Emoticon (b) LIWC and (c) T5-Emotion methods. * denotes $p < 0.05$. ** denotes $p < 0.01$. *** denotes $p < 0.001$.

(a)							
Joy		Sadness		Anger		Surprise	
Word	Z-score	Word	Z-score	Word	Z-score	Word	Z-score
hope	2.89**	sorry	2.10*	animal	1.36	drive	1.43
buddy	2.72**	spilled	1.84	unacceptable	0.98	bargain	1.43
awesome	2.71**	n't	1.82	covid-19	0.84	awful	1.32
great	2.60**	reconsider	1.38	word	0.84	function	1.32
thank	2.53*	stopped	1.30	joking	0.84	snack	1.23

(b)							
Positive Emotions		Sadness		Anger		Anxiety	
Word	Z-score	Word	Z-score	Word	Z-score	Word	Z-score
good	11.89***	sorry	9.93***	hate	2.08*	worried	7.18***
well	11.87***	low	6.16***	greedy	1.58	afraid	4.07***
okay	9.90***	alone	4.43***	critical	1.35	worry	3.5***
great	9.86***	sugar	2.44*	tricky	1.19	risk	3.22**
ok	9.68***	hear	2.43*	bother	1.06	confused	2.54*

(c)											
Joy		Love		Sadness		Fear		Anger		Surprise	
Word	Z-score	Word	Z-score	Word	Z-score	Word	Z-score	Word	Z-score	Word	Z-score
you	8.62***	hot	6.4***	sorry	10.63***	worried	8.72***	cold	11.5***	funny	1.93
good	7.68***	generous	3.04**	unfortunately	4.68***	afraid	7.05***	no	5.76***	interesting	1.89
great	7.33***	hotter	1.89	dehydrated	4.03***	concerned	4.48***	need	4.77***	surprise	1.08
hello	6.84***	lovely	1.56	low	3.12***	dark	3.88***	thirsty	4.63***	ha	1.03
sounds	6.81***	liking	1.34	suffer	2.98**	scared	3.58***	hungry	4.36***	status	0.85

TABLE III: High confidence sample predictions for *T5-Emotion* model that went *undetected* for both *Emoticon* and *LIWC* methods.

Prediction	Sample Utterances
Joy	1) I think that sounds reasonable to me. 2) I can make that deal work. 3) You sound very resourceful! Sounds like we have a deal then, yes?
Sadness	1) I feel like this deal just keeps getting worse for me. I won't take less than 2 waters. 2) That is unfortunate. How about I give you all my firewood for all your food? 3) Oh, that would make things difficult!
Anger	1) You are not getting everything, thats just selfish 2) I am about to walk, you are being so unfair 3) You gave me one on each, thats unfair man!

their own counterparts across the three models, which aligns well with our expectations and validates the model predictions on the CaSiNo dataset. We observe that although significant, most correlation values are weak to moderate, indicating the fundamental differences in what is captured by these three models. We also note that *Anger* shows significant positive correlations with *Sadness* as well. Since these measures are based on the confidence scores of the model, this suggests that the model might get confused between these two negative emotion dimensions.

TABLE IV: Statistics and correlations with outcome variables for all continuous measures. * denotes $p < 0.05$. ** denotes $p < 0.01$.

Variable	Mean	Std.	Correlations	
			Satisfaction	Likeness
Individual Difference				
Age	36.97	10.81	.110**	.158**
Education	5.23	1.67	-.004	-.009
Extraversion	3.69	1.70	.055*	.060**
Agreeableness	5.27	1.29	.067**	.099**
Conscientiousness	5.60	1.26	.026	.046*
Emotional Stability	4.91	1.56	.062**	.087**
Openness to Experiences	5.04	1.31	.027	.042
Emoticons				
Joy	.77	1.24	.040	.075**
Sadness	.13	.42	-.151**	-.170**
Anger	.02	.18	-.074**	-.108**
Surprise	.07	.27	-.062**	-.059**
LIWC				
Positive Emotions	5.58	3.36	.033	.061**
Sadness	.22	.52	-.057**	-.091**
Anger	.05	.25	-.047*	-.043
Anxiety	.18	.52	-.040	-.014
T5-Emotion				
Joy	3.89	1.09	.010	.026
Love	.22	.21	-.010	-.030
Sadness	.35	.42	-.141**	-.172**
Fear	.43	.43	-.121**	-.117**
Anger	.81	.66	-.217**	-.295**
Surprise	.03	.08	-.011	-.016
Negotiation Outcomes				
Satisfaction	4.17	1.03	1	.702
Likeness	4.11	1.12	.702	1

B. Lexical correlates of emotion

To gain more insights, we perform additional qualitative analysis by generating lexical correlates for the emotion measures. We assigned an emotion label to each utterance for all three methods. For *Emoticons* and *LIWC*, this was based on the label that is reflected in the majority in the utterance. For *T5-Emotion* dimensions, we simply used the final predictions from the model. We next compute the log-odds ratio, informative Dirichlet prior [39] of all the tokens for each emotion dimension relative to all other dimensions. This allows us to study the associations between the tokens and each emotion dimension. We present the top 5 words in each category in Table II along with scaled Z-scores. We observe a number of highly significant associations across all categories that align with our intuition such as words like ‘awesome’, ‘great’, ‘thank’ for *Joy* and *Positive Emotions* categories. We found these words to be common in the beginning and end of the negotiation where participants engage in small talk, resulting in such utterances being classified as *Joy* or *Positive*. Further, the associations are higher for *LIWC* and *T5-Emotion* which, as we observe, are able to recognize many other emotive utterances. Although significant, we observe that some words are less meaningful such as in the case of *Anger*. This suggests that some emotion dimensions may require more utterance-level context for better interpretation.

C. Sample Predictions

For a more contextual analysis, we explicitly observe the sample predictions for *T5-Emotion*. We still focus on the three dimensions that are the most common and are present in all the three emotion methods. Table III presents a few of the most confident predictions by the model. To analyze the utterances that are specifically captured by the deep learning model, we only look at the utterances that are *undetected* by the other two methods. As the table depicts, the model is able to capture contextual emotion, beyond just the lexical or emoticon usage. Based on these observations, we expected the affective attributes extracted using the deep learning model to be better predictors than *Emoticons* and *LIWC* measures.

V. RESULTS

We now present our main results.

A. Correlation with outcomes

Table IV summarizes the correlations of all the continuous measures defined in Section III with the negotiation outcomes. Overall, we observe a number of significant trends across individual differences and emotion variables. We find that satisfaction and partner likeness both improve with age ($r=0.110$, $p < 0.01$). The outcomes are also positively correlated with Agreeableness and Emotional Stability ($ps < 0.01$). We observe that positive emotions tend to relate positively with both the outcomes, showing significant trends for likeness in the case of *Emoticon* usage ($r=0.075$, $p < 0.01$) and *LIWC* ($r=0.061$, $p < 0.01$). We do not observe similar results for *T5-Emotion*. This may be due to the inherent bias that the deep

learning model shows towards *Joy* label, reducing its precision. However, trends show significant negative correlations for both *Sadness* and *Anger* that are usually expressed when the negotiation is not going favorably. This is naturally associated with lower satisfaction of the participants and liking for their negotiation partners.

Among the discrete variables, we find that females are significantly more satisfied ($t=3.6$, $p < 0.001$) and significantly more like their partners ($t=4.1$, $p < 0.001$) than males. We do not observe any significant associations with SVO and Ethnicity, based on t-test and one-way ANOVA respectively.

B. Regression Analysis

We now discuss the results of our regression analysis that answers whether emotion variables extracted from the negotiation dialogue collectively explain more variance in satisfaction and likeness, above and beyond the individual difference variables. To achieve this, we perform regression with three steps where each subsequent step incorporates the following set of variables: 1) Individual difference, 2) Affect variables of the participant, and 3) Affect variables for the negotiation partner. The affect variables come from one of the emotion methods defined in Section III, which also helps to compare their performance. As discussed earlier, these variables only encompass the information about the negotiation that would be entirely *visible* to an automatic negotiation agent, either explicitly from the negotiation dialogue, or implicitly by inferring the individual difference attributes from past online behaviors of the users on social media platforms. Hence, our insights from such an analysis can practically aid in designing sophisticated negotiation agents that incorporate user satisfaction and their perception of the agent itself in their modelling.

Table V and VI summarize the results for satisfaction and likeness respectively. Overall, we observe that the predictions are highly significant with just using the individual difference variables. However, incorporating the affective attributes of the participant and the negotiation partner explains significantly more variance.

For satisfaction, the individual difference variables account for minimal yet significant variance ($F(14, 1997)=3.46$, $p < .001$, $R^2=.024$). Adding the participant’s affect variables based on the *T5-Emotion* model in the second step, helps to account for a much higher variance ($F(20, 1991)=10.41$, $p < .001$, $R^2=.095$), such that this increase in the proportion is itself highly significant ($\Delta F(6, 1991)=26.02$, $p < .001$, $\Delta R^2=.071$). Yet, further variance is explained when the partner’s affect variables are incorporated ($F(26, 1985)=10.88$, $p < .001$, $R^2=.125$; $\Delta F(6, 1985)=11.38$, $p < .001$, $\Delta R^2=.030$).

Similar trends can be observed for likeness as well. Individual difference variables alone account for a significant proportion ($F(14, 1997)=6.05$, $p < .001$, $R^2=.041$) but adding the participant’s *T5-Emotion* attributes further improves the prediction performance ($F(20, 1991)=18.16$, $p < .001$, $R^2=.154$) with a significant increase ($\Delta F(6, 1991)=44.58$, $p < .001$, $\Delta R^2=.114$). Finally, when the partner’s attributes are also incorporated as predictors, this again shows significant improve-

TABLE V: Regression results for predicting Satisfaction. ** denotes $p < 0.01$. *** denotes $p < 0.001$.

Variables	R ²	df	F	R ² Change	F Change
Affect Variables: Emoticons					
Individual Difference	.024	(14, 1997)	3.46***	—	—
+Participant Affect	.051	(18, 1993)	5.92***	.027	14.20***
+Partner Affect	.063	(22, 1989)	6.04***	.012	6.34***
Affect Variables: LIWC					
Individual Difference	.024	(14, 1997)	3.46***	—	—
+Participant Affect	.032	(18, 1993)	3.61***	.008	4.07**
+Partner Affect	.051	(22, 1989)	4.85***	.019	10.14***
Affect Variables: T5-Emotion					
Individual Difference	.024	(14, 1997)	3.46***	—	—
+Participant Affect	.095	(20, 1991)	10.41***	.071	26.02***
+Partner Affect	.125	(26, 1985)	10.88***	.030	11.38***

TABLE VI: Regression results for predicting Likeness. *** denotes $p < 0.001$.

Variables	R ²	df	F	R ² Change	F Change
Affect Variables: Emoticons					
Individual Difference	.041	(14, 1997)	6.05***	—	—
+Participant Affect	.080	(18, 1993)	9.68***	.040	21.50***
+Partner Affect	.097	(22, 1989)	9.70***	.017	9.12***
Affect Variables: LIWC					
Individual Difference	.041	(14, 1997)	6.05***	—	—
+Participant Affect	.055	(18, 1993)	6.39***	.014	7.32***
+Partner Affect	.072	(22, 1989)	7.06***	.018	9.56***
Affect Variables: T5-Emotion					
Individual Difference	.041	(14, 1997)	6.05***	—	—
+Participant Affect	.154	(20, 1991)	18.16***	.114	44.58***
+Partner Affect	.200	(26, 1985)	19.07***	.046	18.83***

ments in the explained variance ($F(26, 1985)=19.07$, $p < .001$, $R^2=.200$; $\Delta F(6, 1985)=18.83$, $p < .001$, $\Delta R^2=.046$).

Lastly, we note that in line with our linguistic analysis, the ability for *T5-Emotion* to capture more contextual emotive utterances helps to account for more variance, as compared to *Emoticons* and *LIWC* independently. In our experiments, we also observed that even when *Emoticons* and *LIWC* are combined, *T5-Emotion* still achieves a much better performance.

VI. DISCUSSION AND CONCLUSIONS

Our aim was to empirically investigate the extent to which the affect variables extracted from the negotiation itself are helpful for prediction of two important metrics for negotiation agents – participant’s outcome satisfaction and liking for the partner. We presented an extensive analysis based on a large-scale dataset of human-human negotiation dialogues, grounded in a realistic camping scenario. We devised three degrees of emotion dimensions, from leveraging emoticon usage similar to prior work in menu-driven systems, to going beyond and extracting emotion expression directly from the textual chat utterances. Our results show that such affect variables explain variance in these subjective outcomes, above and beyond the impact of individual difference variables that are available before the negotiation begins³.

One might expect that the affect variables would help merely because they are manifestations of the individual difference such as the social value orientation, agreeableness,

³We also tried controlling for the objective negotiation performance of the partner in the regression analysis. We still observed similar significant trends for both satisfaction and likeness.

or gender. If this was indeed the case, there would have been no need to incorporate such affective variables, but merely the individual difference variables would have sufficed. However, to the extent to which the affect variables help in predictions, above and beyond these individual differences, this suggests that there is utility in extracting these emotion dimensions explicitly for improved agent’s performance, being especially well-suited for more subjective outcomes in a negotiation.

We further note that the individual difference variables that we use are based on self-identification and standard personality tests from the psychology literature. This makes these variables more reliable than if they were to be implicitly inferred from past social behaviors of the users, as attempted in a number of previous works analysing interactions on social media platforms [9]–[11]. If affective factors show significant improvements above the individual differences measured with minimal error, this further attests to their utility when the demographics and personality information is inferred, and thus, less reliable.

Since negotiations and other mixed-motive situations can be fraught with emotional decisions, our findings suggest that it would be useful for designers of agents that negotiate with humans to be *armed* with algorithms for understanding the unfolding emotions displayed by users in terms of their emoticon use and natural language. Our work suggests that demographic and personality attributes inferred from past social interactions (such as on social media) are not sufficient, and that providing information to negotiating agents from affective channels tracked during the negotiation itself is important for developing agents that can predict or understand if their human counterpart is satisfied and likes them. Ultimately, outcomes like satisfaction and likeness will be essential for such agents to cultivate if they, like human negotiators, hope to successfully negotiate with that same partner in the future [4].

In light of the theme for this year’s ACII on *Ethical Affective Computing*, we briefly discuss the ethical considerations around the development of automated negotiation agents. There has been a significant amount of work on ethics in negotiation literature. The acts of emotion manipulation, deception, bias, and misinterpretation have been central concerns [40]. It is plausible that these behaviors also get reflected in the negotiation agents that are grounded in the insights gained from human-human negotiation datasets such as CaSiNo. To mitigate the impact of these ethical concerns when automated negotiation systems are deployed to the end users, we lay out a few recommendations. We believe that transparency is the key. Being upfront and open about the identity, capabilities, and any known misbehaviors of the agent would prevent any misinterpretation and would set the right expectations for the end users. Further, we note that our work encourages the agent to adapt its behavior by incorporating affective attributes along with user demography and personality. If the deployed agent is indeed adaptive towards these traits, we recommend a regular monitoring cycle for unexpected behaviors of the agent, preventing against any offensive or discriminative actions.

REFERENCES

- [1] J. Gratch, D. DeVault, G. M. Lucas, and S. Marsella, "Negotiation as a challenge problem for virtual humans," in *International Conference on Intelligent Virtual Agents*. Springer, 2015, pp. 201–215.
- [2] R. L. Oliver, P. S. Balakrishnan, and B. Barry, "Outcome satisfaction in negotiation: A test of expectancy disconfirmation," *Organizational Behavior and Human Decision Processes*, vol. 60, no. 2, pp. 252–275, 1994.
- [3] J. Mell, J. Gratch, R. Aydoğan, T. Baarslag, and C. M. Jonker, "The likeability-success tradeoff: Results of the 2nd annual human-agent automated negotiating agents competition," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2019, pp. 1–7.
- [4] R. Aydoğan, T. Baarslag, K. Fujita, J. Mell, J. Gratch, D. de Jonge, Y. Mohammad, S. Nakadai, S. Morinaga, H. Osawa *et al.*, "Challenges and main results of the automated negotiating agents competition (anac) 2019," in *Multi-Agent Systems and Agreement Technologies*. Springer, 2020, pp. 366–381.
- [5] Y. Leviathan and Y. Matias, "Google duplex: An ai system for accomplishing real-world tasks over the phone," *URL https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html*, vol. 3, 2018.
- [6] E. Johnson, G. Lucas, P. Kim, and J. Gratch, "Intelligent tutoring system for negotiation skills training," in *International Conference on Artificial Intelligence in Education*. Springer, 2019, pp. 122–127.
- [7] A. F. Stuhlmacher and A. E. Walters, "Gender differences in negotiation outcome: A meta-analysis," *Personnel Psychology*, vol. 52, no. 3, pp. 653–677, 1999.
- [8] Y. Xu, P. Sequeira, and S. Marsella, "Towards modeling agent negotiators by analyzing human negotiation behavior," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2017, pp. 58–64.
- [9] Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla, "Inferring user demographics and social strategies in mobile social networks," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 15–24.
- [10] A. Ortigosa, J. I. Quiroga, and R. M. Carro, "Inferring user personality in social networks: A case study in facebook," in *2011 11th International Conference on Intelligent Systems Design and Applications*. IEEE, 2011, pp. 563–568.
- [11] S. Adali and J. Golbeck, "Predicting personality with social behavior: a comparative study," *Social Network Analysis and Mining*, vol. 4, no. 1, pp. 1–20, 2014.
- [12] J. Mell and J. Gratch, "Iago: interactive arbitration guide online," in *AAMAS*, 2016, pp. 1510–1512.
- [13] H. He, D. Chen, A. Balakrishnan, and P. Liang, "Decoupling strategy and generation in negotiation dialogues," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 2333–2343.
- [14] K. Chawla, J. Ramirez, R. Clever, G. Lucas, J. May, and J. Gratch, "Casino: A corpus of campsite negotiation dialogues for automatic negotiation systems," *arXiv preprint arXiv:2103.15721*, 2021.
- [15] J. F. Nash Jr, "The bargaining problem," *Econometrica: Journal of the Econometric Society*, pp. 155–162, 1950.
- [16] P. J. Carnevale and D. G. Pruitt, "Negotiation and mediation," *Annual review of psychology*, vol. 43, no. 1, pp. 531–582, 1992.
- [17] C. Beam and A. Segev, "Automated negotiations: A survey of the state of the art," *Wirtschaftsinformatik*, vol. 39, no. 3, pp. 263–268, 1997.
- [18] T. Baarslag, M. J. Hendriks, K. V. Hindriks, and C. M. Jonker, "A survey of opponent modeling techniques in automated negotiation," in *15th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2016*. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), 2016, pp. 575–576.
- [19] G. Stratou, R. Hoegen, G. Lucas, and J. Gratch, "Emotional signaling in a social dilemma: An automatic analysis," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2015, pp. 180–186.
- [20] C. M. de Melo, P. Carnevale, and J. Gratch, "The effect of expression of anger and happiness in computer agents on negotiations with humans," in *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, 2011, pp. 937–944.
- [21] J. Nadler, "Rapport in negotiation and conflict resolution," *Marq. L. Rev.*, vol. 87, p. 875, 2003.
- [22] J. Mell, G. M. Lucas, and J. Gratch, "An effective conversation tactic for creating value over repeated negotiations," in *AAMAS*, vol. 15, 2015, pp. 1567–1576.
- [23] G. Zacharia and P. Maes, "Trust management through reputation mechanisms," *Applied Artificial Intelligence*, vol. 14, no. 9, pp. 881–907, 2000.
- [24] T. Baarslag, K. Fujita, E. H. Gerding, K. Hindriks, T. Ito, N. R. Jennings, C. Jonker, S. Kraus, R. Lin, V. Robu *et al.*, "Evaluating practical negotiating agents: Results and analysis of the 2011 international competition," *Artificial Intelligence*, vol. 198, pp. 73–103, 2013.
- [25] M. Lewis, D. Yarats, Y. Dauphin, D. Parikh, and D. Batra, "Deal or no deal? end-to-end learning of negotiation dialogues," in *EMNLP*, 2017.
- [26] D. DeVault, J. Mell, and J. Gratch, "Toward natural turn-taking in a virtual human negotiation agent," in *AAAI Spring Symposia*. Citeseer, 2015.
- [27] C. Fershtman, "The importance of the agenda in bargaining," *Games and Economic Behavior*, vol. 2, no. 3, pp. 224–238, 1990.
- [28] N. Asher, J. Hunter, M. Morey, B. Farah, and S. Afantenos, "Discourse structure and dialogue acts in multiparty dialogue: the stac corpus," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 2721–2727.
- [29] P. A. Van Lange, E. De Bruin, W. Otten, and J. A. Joireman, "Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence," *Journal of personality and social psychology*, vol. 73, no. 4, p. 733, 1997.
- [30] L. R. Goldberg, "An alternative" description of personality": the big-five factor structure," *Journal of personality and social psychology*, vol. 59, no. 6, p. 1216, 1990.
- [31] S. Bogaert, C. Boone, and C. Declerck, "Social value orientation and cooperation in social dilemmas: A review and conceptual model," *British Journal of Social Psychology*, vol. 47, no. 3, pp. 453–480, 2008.
- [32] R. G. Curtis, T. D. Windsor, and A. Soubelet, "The relationship between big-5 personality traits and cognitive ability in older adults—a review," *Aging, Neuropsychology, and Cognition*, vol. 22, no. 1, pp. 42–71, 2015.
- [33] C. G. McClintock, "Social values: Their definition, measurement and development," *Journal of Research & Development in Education*, 1978.
- [34] M. G. Ehrhart, K. H. Ehrhart, S. C. Roesch, B. G. Chung-Herrera, K. Nadler, and K. Bradshaw, "Testing the latent factor structure and construct validity of the ten-item personality inventory," *Personality and individual Differences*, vol. 47, no. 8, pp. 900–905, 2009.
- [35] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, no. 2001, p. 2001, 2001.
- [36] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, pp. 1–67, 2020.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
- [38] E. Saravia, H.-C. T. Liu, Y.-H. Huang, J. Wu, and Y.-S. Chen, "Carer: Contextualized affect representations for emotion recognition," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 3687–3697.
- [39] B. L. Monroe, M. P. Colaresi, and K. M. Quinn, "Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict," *Political Analysis*, vol. 16, no. 4, pp. 372–403, 2008.
- [40] R. J. Lewicki, B. Barry, and D. M. Saunders, *Essentials of negotiation*. McGraw-Hill, 2016.