Recognizing Perceived Interdependence in Face-to-Face Negotiations through Multimodal Analysis of Nonverbal Behavior

Dudzik, Bernd; Columbus, Simon; Hrkalovic, Tiffany Matej; Balliet, Daniel; Hung, Hayley

DOI
10.1145/3462244.3479935

Publication date
2021

Document Version
Final published version

Published in
ICMI 2021 - Proceedings of the 2021 International Conference on Multimodal Interaction

Recognizing Perceived Interdependence in Face-to-Face Negotiations through Multimodal Analysis of Nonverbal Behavior

Bernd Dudzik
Delft University of Technology
Delft, The Netherlands
B.J.W.Dudzik@tudelft.nl

Simon Columbus
University of Copenhagen
Copenhagen, Denmark
simon@simoncolumbus.com

Tiffany Matej Hrkalovic
Vrije Universiteit Amsterdam
Amsterdam, The Netherlands
t.matejhrkalovic@vu.nl

Daniel Balliet
Vrije Universiteit Amsterdam
Amsterdam, The Netherlands
D.P.Balliet@vu.nl

Hayley Hung
Delft University of Technology
Delft, The Netherlands
H.Hung@tudelft.nl

ABSTRACT

Enabling computer-based applications to display intelligent behavior in complex social settings requires them to relate to important aspects of how humans experience and understand such situations. One crucial driver of peoples' social behavior during an interaction is the interdependence they perceive, i.e., how the outcome of an interaction is determined by their own and others' actions. According to psychological studies, both the nonverbal behavior displayed by (1) persons *themselves* and (2) that of *others* interacting with them may facilitate inferences about their perceptions of interdependence. Motivated by this, we present a series of experiments to automatically recognize interdependence perceptions in dyadic face-to-face negotiations using these sources. Concretely, our approach draws on a combination of features describing individuals' *Facial*, *Upper Body*, and *Vocal Behavior* with state-of-the-art algorithms for multivariate time series classification. Our findings demonstrate that differences in some types of interdependence perceptions can be detected through the automatic analysis of nonverbal behaviors. We discuss implications for developing socially intelligent systems and opportunities for future research.

CCS CONCEPTS

• **Human-centered computing**;

KEYWORDS

Social Signal Processing; Situation Perception; User-Modeling

ACM Reference Format:
Bernd Dudzik, Simon Columbus, Tiffany Matej Hrkalovic, Daniel Balliet, and Hayley Hung. 2021. Recognizing Perceived Interdependence in Face-to-Face Negotiations through Multimodal Analysis of Nonverbal Behavior. In *Proceedings of the 2021 International Conference on Multimodal Interaction (ICMI '21), October 18–22, 2021, Montréal, QC, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3462244.3479935

ICMI '21, October 18–22, 2021, Montréal, QC, Canada

ACM ISBN 978-1-4503-8481-0/21/10.
https://doi.org/10.1145/3462244.3479935

1 INTRODUCTION

Modern computer systems are increasingly envisioned to closely collaborate with humans in social environments, e.g., as partners offering support in education [21], or mental healthcare [6]. Technological systems can benefit from estimating human's conceptualization and experience of social situations as contextual information about an interaction to display intelligent behavior in such collaboration settings [55]. In particular, intelligent systems could use such information about social perceptions to meaningfully adapt their functionality and behavior by facilitating predictions of (1) what a human might do (2) why they might do it, as well as (3) how a user might experience a system's actions.

Motivated by this, one strand of *Social Signal Processing (SSP)* strives to facilitate technology that can interpret the meaning of behavioral cues in terms of social concepts relevant for human interaction and their judgments [55]. Different theoretical accounts from psychology can guide this effort since they describe how individuals' interpretations and judgments of the social situation, together with their judgments of other persons, drive behavioral and emotional responses [34, 43].

One important aspect of how individuals are thought to conceptualize social interactions is the interdependence present in the situation [3]: their perceptions of how potential outcomes of an interaction are determined by their own and others' actions. In particular, when people perceive themselves as more mutually dependent, they are more willing to behave cooperatively; in contrast, perceived conflicts of interest undermine cooperation. Differences in relative power may also affect one's willingness to cooperate [10]. This connection makes interdependence a highly desirable construct for intelligent systems to be aware of in order to reason whether people may be more or less likely to collaborate with one another (or potentially another system). Research in psychology has identified multiple dimensions of perceived interdependence [19]. These dimensions include (1) *Mutual Dependence*, (2) *Conflict of Interest*, (3) *Future Interdependence*, (4) *Information Certainty*, as well as (5) *Power*, and can be assessed using a validated instrument [19] (See *Table 1* for definitions). Together, these prior findings

Table 1: Dimensions of Situational Interdependence (as defined by Gerpott et al. [19])

DIMENSION	DESCRIPTION
Mutual Dependence (MD)	Degree of how much each person's outcomes are determined by how each person behaves in that situation.
Conflict of Interest (C)	Degree to which the behavior that results in the best outcome for one individual results in the worst outcome for the other.
Future Interdependence (FI)	Degree to which own and others' behavior in the present situation can affect own and others behavior and outcomes in future interactions.
Information Certainty (IC)	Degree to which a person knows their partner's preferred outcomes and how each person's actions influence each other's outcomes.
Power (P)	Degree to which an individual determines their own and others' outcomes, while others do not influence their own outcome.

provide a solid framework for developing an automated approach for modeling human perceptions of social situations. Moreover, while existing efforts in SSP have not systematically touched upon modeling such social situation perceptions, there is a substantial body of work on related constructs indicating the technological feasibility of doing so (see *Section 2.2* breakdown).

Motivated by the potential for improving the capacities of socially intelligent systems, we investigate the potential of recognizing an individual's interdependence perceptions through an analysis of audiovisual recordings of dyadic conversational interaction. While the process of interdependence perception is still a subject of ongoing research, interaction partners' nonverbal behavior is likely one source of information utilized in it [3]. For example, preliminary findings point towards the importance of facial expressions and gestures [19]. Moreover, individuals' perceptions of interdependence may guide their actions in social situations [3].

Concretely, we present the following contributions:

- We explore the feasibility of automatically recognizing the interdependence perceived by individuals in face-to-face negotiations by analyzing audiovisual data about the behavior displayed by (1) themselves as well as (2) their interaction partner. Concretely, we focus on information about individuals' *Facial Expressions*, their *Upper Body Behavior*, as well as their *(Non-verbal) Vocal Behavior*.
- We present our approach for predictive modeling as a baseline for future technological research on this task. It relies on a State-of-the-Art approach to classify multivariate time series of behavioral features.
- We discuss the benefits of providing intelligent technologies with the capacity to predict individuals' interdependence perceptions and point out targets for future research.

2 BACKGROUND AND RELATED WORK

2.1 Situation Perception and Interdependence

The situational context in which a person finds themselves shapes their cognition and affect. Conversely, however, a person also forms an impression of the situation, which may influence their behavior [3, 44]. Forming impressions of situations might help individuals to orient themselves and navigate their everyday life [3, 43]. Consequently, it has been suggested that people form such impressions along dimensions that constrain their behavior and determine the outcomes of their actions [3]. One important set of dimensions on which social situations differ is interdependence.

Interdependence describes how people mutually control their own and others' outcomes in a situation. Experimental evidence shows that objective interdependence causes variation in cooperative behavior, with especially conflicting interests and power differences having strong effects on negotiation behavior and outcomes [5, 12, 22, 28].

While social situations can objectively and structurally differ in terms of their interdependence, in real-world interactions, how individuals perceive the interdependence present in their dealings with others determines their affect, cognition, and behavior. Subjective interdependence refers to these perceptions [3]. People readily understand and describe social situations along dimensions such as mutual dependence, conflict of interests, and relative power [19]. Such perceptions track objective features of an interaction, but they are also influenced by stable and situational features such as frames through which people perceive the situation [11] and of the person, such as their personality [19] (for a broader framework for situation perception, see [44]). Empirical research has shown that perceptions of interdependence are associated with cooperative behavior [10]. In negotiations, in particular, perceiving the negotiation as a 'fixed pie'—i.e., high in conflict of interests—has been linked to worse negotiation outcomes [51, 52]; and perceptions of power have been linked to lower offers [42]. More broadly, subjective perceptions account for a significant share of the variation in cooperative behavior across situations and between people [10, 11]. Initial experimental evidence also suggests that this link is partly causal—i.e., differences in perceived interdependence cause differences in cooperative behavior [11].

2.2 Interdependence and (Analysis of) Nonverbal Behavior

A person's perceptions of interdependence in a situation may be reflected in their nonverbal behavior and informed by their interaction partner's nonverbal behavior. As such, past studies have explored both whether differences in interdependence are associated with expressed nonverbal behavior and whether people infer interdependence from others' nonverbal behavior. Moreover, computational work has modeled the relationship between nonverbal behavior and various constructs related to individual dimensions of interdependence. In the following, we provide an overview of empirical findings linking each interdependence dimension with either perceived or expressed nonverbal behavior, combined with examples for related work from Social Signal Processing and Affective Computing. Together, these findings highlight both the existence of (1) links between nonverbal behaviors and the various interdependence dimensions, as well as (2) the relevance of situational interdependence as a construct concerning existing technological research.

2.2.1 Mutual Dependence (MD):. Perceptions of MD have been linked to eye contact and attention to one's interaction partner [3].

A related construct addressed extensively in Social Signal Processing is team cohesion. For example, Hung and Perez [25] investigate the automatic analysis of task-based group meetings based on audiovisual analysis, while Zhang et al. explore assessment based on nonverbal behavior collected through wearable sociometers in a longitudinal study [61].

2.2.2 Conflict of Interest (C):. Prior research has established that CI is positively associated with anger and negatively with happiness [3, 41]. Moreover, people infer corresponding interests from smiles [54], nodding, leaning forward, or affective touching, whereas crossing arms and leaning away is linked to perceived conflict [3]. Several projects have explored the detection of conflict in social interactions from nonverbal cues, e.g., during discussions [29, 38]. Moreover, detecting expressions of aggression and disagreement – indicators of conflict – is a prominent line of research in Affective Computing (e.g., using vocal features [31]).

2.2.3 Future Interdependence (FI):. Little direct empirical evidence exists on the connection between FI perceptions and either expressed or perceived nonverbal behavior. However, there is evidence for individuals differing in nonverbal behavior when interacting with either strangers (individuals on which they may not expect to depend on in the future) and familiars (individuals on which they may depend on further down the line), e.g., in terms of emotional facial expressions [56]. Similarly, exploiting this link to predict the relationship between individuals has been explored in computational work [60]. Moreover, computational work has attempted multimodal predictions of peoples' desire to engage in future interactions with each other, for example, in the context of job interviews [36].

2.2.4 Information Certainty (IC):. Existing findings show that nonverbal behavioral cues can facilitate insights into the feelings of ones' conversation partner about a subject of discussion, leading to a decrease in situational ambiguity [23]. Moreover, the intensity of emotional responses a person displays might indicate the degree of importance they assign to a particular outcome [37]. Similarly, there is evidence linking the richness of behavioral expressions (e.g., in terms of accessible cues) to the reduction of ambiguity in emotion perceptions [58]. Together, this makes it plausible that individuals can rely on conversation partners' nonverbal affective behavior to reason about their preferred outcome (e.g., [41]), and that the presence (or absence) of relevant nonverbal signals relates to the perceived certainty of those inferences. Meanwhile, certainty judgments during interactions have not been directly addressed in automatic behavioral analysis to the best of our knowledge. However, there exists a substantial amount of work on recognizing surprise from nonverbal expressions (in particular from the face, e.g., [33]), which indicates unanticipated or novel stimuli.

2.2.5 Power (P):. Perceptions of higher power have been linked to more frequent expressions of anger [18, 40, 49], but also smiling, the loudness of voice, interruptions, expansive posture, and direct eye gaze [24]. People may also attribute high power to others based on expressions of pride they display [53], free posture, direct eye gaze, more frequent hand gestures [9] and low power from expressions of appreciation, fear, and sadness [53]. These findings are in line with the assumption that individuals who feel more powerful might be more assertive and relaxed [45]. Broadly, the approach/inhibition model of power suggests that low power leads to inhibition of emotional expressiveness, whereas high power is associated with expression of anger and happiness [27]. Finally, there exists a substantial body of research using automatic behavioral analysis to identify nonverbal patterns indicating dominance – the behavioral expression of power [7]. For example, Jayagopi et al. [26] identify the dominant members of a conversation through analysis of their nonverbal behavior as captured by audiovisual recordings. Similarly, there exists a substantial amount of technical work exploring the analysis of nonverbal expressions to estimate feelings of confidence, e.g., in public speaking settings [59].

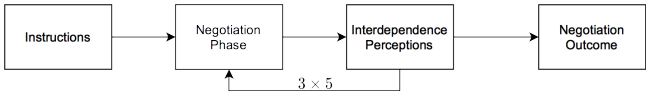

Figure 1: Overview of Data Collection Protocol

3 DATASET

Here we provide an overview of the corpus that we use for modeling and our steps for cleaning and preprocessing. It captures people's perceptions of interdependence during a face-to-face negotiation task with another person. It includes audiovisual recordings of their behavior throughout this conversation.

3.1 Data Collection

3.1.1 Procedure: Participants were assigned the roles of *applicant* or *HR manager* in a simulated job negotiation involving eight different issues (e.g. salary, annual bonus). Each issue had five options for which participants could earn a varying number of points. Dyads were assigned to either a low or a high conflict treatment. In all treatments, two issues were integrative, meaning that the two negotiators could trade one issue off against the other to maximize joint gain. In the low (high) conflict treatment, five (one) issues were compatible, meaning that both negotiators preferred the same option, and one (five) issue was distributive, meaning that one negotiator's best outcome was the other's worst outcome. In addition, dyads differed in the consequences of not agreeing. All started with an outside option worth 40% of an even split. In one-third of dyads, one negotiator received an increase to 60%, in another third, a decrease to 20%. In the remaining third, both negotiator's outside options remained unchanged. Dyads were seated face-to-face, each with a laptop in front of them. Messages and surveys were presented on the laptop. At the outset, participants received information only about their own payoffs. The negotiation proceeded in three five-minute phases (T1, T2, T3); see *Figure 1* for an overview of the protocol. Participants were free to interact during these phases. Thirty seconds after the start of T2, one participant was informed about a change in their outside option. At the end of T3, participants had to come to a negotiated agreement. Points earned were paid out (up to €20 per participant, on top of a show-up fee of €15).

3.1.2 Collected Measures: Before the start of the negotiation and after each phase, participants reported their perceptions of interdependence on the 10-item short form of the *Situational Interdependence Scale* [19]. Additionally, each participant's behavior

throughout the negotiation was recorded on video. The footage was captured using a Logitech C920 camera set on a laptop facing the participant, at a resolution of 1280 ∗ 720 with 30 frames per second.

3.1.3 Curation and Preprocessing: Video recordings were transcoded to a resolution of 640 ∗ 360 pixel for further processing and split into 5 minute segments delineating the specific negotiation phases they cover. We removed records of negotiation phases with incomplete or invalid self-reports from the corpus to create a coherent multimodal dataset. Furthermore, 4 participants in the remaining dataset completed their final negotiation phase (T3) early, and their records for this phase have a duration less than the targeted 5 minutes. Since the algorithms that we deploy for predictive modeling (see *Section 4.1*) require sequences of a fixed length, we discarded data from these specific negotiation phases. Finally, we only kept records from interactions where data was available for both participants after filtering, removing an additional 3 participants. The curated corpus contains combined records (perceived interdependence and behavioral recordings) for 632 negotiation phases (T1: 212, T2: 212, and T3: 208) from a total of 106 sessions. The 212 participants were mostly in their 20s (Age: $M(SD) = 22.306(7.097)$) born in the Netherlands (Netherlands: $N = 182$; Other/Undisclosed: $N = 30$), with a majority identifying as female (Female: $N = 124$; Male: $N = 70$; Other/Undisclosed: $N = 18$).

3.2 Multimodal Contents

3.2.1 Audiovisual Recordings of Negotiation-Phases: The audiovisual recordings in the datasets show participants from a frontal perspective as they are seated at a table and capture their upper body and face throughout the negotiation with their partner. Overall, this is a highly controlled setting, including stable lighting conditions and no environmental background noise, and as such should be well suited for automatic behavioral analysis. However, manual inspection of the data reveals occasional instances of participants' faces being outside of the frame or being occluded by sheets of paper that they hold in their hands. Moreover, since they are close in physical space, there are moments where one person's microphone picks up the voice of their conversation partner, introducing a potential source of noise. These conditions are well within the range of natural behavior in front of a screen (e.g. during video conferencing). As such, any approach for recognizing interdependence from behavioral data in the real world will need to be capable of functioning despite them.

3.2.2 Reports of Perceived Interdependence: An overview of the distribution for the intensity of interdependence perceptions across all negotiation-phases can be seen in *Table 2*. Apart from ratings for P and C the distributions of self-reported interdependence perceptions are biased towards higher levels of intensity (MD: $Median = 4$; FI: Median=3.5; IC:Median=4). This imbalance indicates substantial intersubjective agreement among participants, likely reflecting comparatively stable situation characteristics inherent in the negotiation setting.

For predictive modeling, we split the continuous ratings for each interdependence perception into distinct classes. For this purpose, we first bin the interval of the continuous variables (i.e., $[1 - 5]$) into three equally spaced ranges, indicating low $[0, 1.666]$, medium $[1.667 - 3.334]$ and high intensity $[3.335 - 5]$. Due to the substantial negative skew of the data, this results in very few examples for low-intensity perceptions, and we merge this class with the medium-intensity category. Consequently, we phrase the recognition of interdependence perceptions as a series of binary classification-tasks with the goal of differentiating between negotiations with either a (1) *Low-Medium (LM)* or (2) *High (H) Intensity* along a particular dimension of interdependence. The reason for this splitting approach is to preserve the mapping to the original rating scale. Using an alternative strategy, e.g., a median-split would have broken this connection. An overview of amount of negotiation-phases in the dataset falling into each of these classes is also available in *Table 2*.

Table 2: Distribution of Interdependence-Perceptions

DIMENSION	M(SD)	Min/Max	#LM	#H	#T
Mutual Dependence (MD)	4.212 (0.672)	2./5.	71	561	632
Conflict of Interest (C)	3. (1.075)	1./5.	365	267	632
Future Interdependence (FI)	3.502 (0.672)	1./5.	258	374	632
Information Certainty (IC)	3.646 (1.003)	1./5.	231	401	632
Power (P)	3.100 (0.673)	1./5.	173	459	632

4 PREDICTIVE MODELING

4.1 Overview

In this section, we describe our pipeline for automatic recognition of interdependence perceptions from nonverbal behavioral signals. A central motivation of our approach is to facilitate the interpretation of models in future work regarding the importance of types of human behavioral cues for their predictive performance. As such, we opt for a solution that is grounded in interpretable features of nonverbal behavior as an intermediate representation for predictions, rather than end-to-end learning directly from audiovisual data (i.e., using Deep Learning). Moreover, learning effective representations for behavioral analysis from audiovisual media generally requires large-scale datasets to be effective (see, e.g., the recent review by Rouast et al. in the context of Affect Detection [46]). As such, they are unlikely to be applicable to a dataset of our size. Here, we rely on existing technologies to extract behavioral features from audiovisual recordings (face and pose features from individual video frames and speech features from short audio segments; see below). We then combine and concatenate these features along the time-axis into a single multivariate time series (MTS) for further processing and prediction.

An important component of multimodal social signal processing is combining information from different sources for automatic analysis and prediction. Within the context of Affective Computing, research has extensively explored feature-level fusion – where feature vectors for different modalities are concatenated directly with each other for predictions – and decision-level fusion – where separate models are trained on the features from each of the different information sources and then combined using a meta-estimator or a voting scheme. Traditionally, neither approach has consistently demonstrated performance benefits over the other [15]. Here, we extract a shared representation from the MTS to capture patterns in temporal dynamics within and across the individual channels of behavioral signals (using a multivariate version of the MINIROCKET

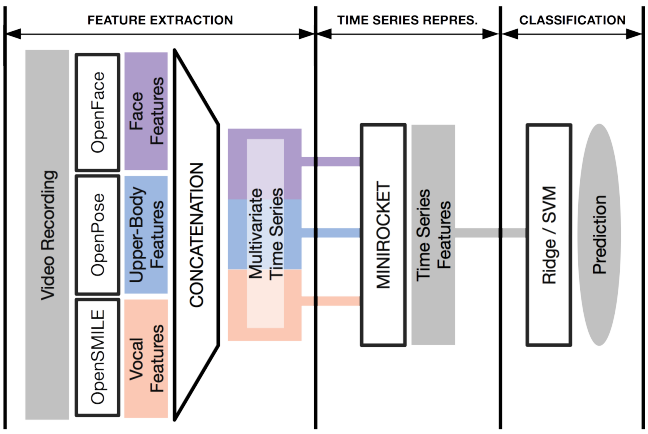

Figure 2: Overview of our Predictive Modeling Pipeline for Automatic Recognition of Interdependence Perceptions.

algorithm [14]; details below). Classification of the MTS is then based on this shared representation. Consequently, our approach relies on a form of feature-level fusion, and as such, is potentially capable of capturing relationships across individual features and their dynamics over time.

We present a graphical overview of the entire machine learning pipeline that we deploy for recognition of interdependence perceptions in *Figure 2*. Processing in it is undertaken in three stages: (1) *Feature Extraction*, (2) *Time Series Representation*, and (3) *Classification*. We describe each stage in greater detail below. The entire pipeline is deployed separately for predictions of each dimension of the situational interdependence construct, taking either a video of the person's own expressed behavior or that of their conversation partner as input. Crucially, we deploy this approach in two different variations in our empirical investigations, which we denote as (1) *MINIROCKET-Ridge*, and (2) *MINIROCKET-SVM*, differing the algorithm that they rely on for classification.

4.2 Feature Extraction

4.2.1 Facial Behavior: We deploy the software *OpenFace 2.0* [4] for extracting feature-sets relating to different aspects of individuals' facial behavior. To characterize *Facial Expressions*, we rely on the activation-intensity of the 17 Facial Action Units provided by OpenFace (*AU Intensities*). Intensities range from $0-5$, whereby a value of 0 denotes no activation of the action unit in question, and a value of 5 an activation at maximum intensity. To capture pattern in *Eye Gaze* during negotiation, we use the gaze-direction vectors provided by OpenFace, which comprise of angles in radians along the (x, y)-axis in world coordinates averaged across both eyes. Finally, to obtain an indication of participants *Head Pose*, we use estimates provided by OpenFace for the head's position ((x, y, z)-coordinates), and orientation ($(pitch, yaw, rotation)$-angles) relative to the camera. Frame-level vectors for each of these feature sets for a particular recording are then concatenated along the time axis to result in a 25-dimensional multivariate time series of length $L = 9000$. In a further processing step, feature values for frames in which OpenFace provided a low-confidence score (i.e., $< .1$) are replaced with the mean value for that feature across the previous and following steps in the time series (3% of all frames were affected by this procedure).

4.2.2 Upper-body Behavior: To characterize individuals' upper-body behavior, we use *OpenPose* [8] to extract the positions of 7 anatomical key-points ((x, y)-coordinates) at the frame-level: (1) Neck, (2) Left Shoulder, (3) Right Shoulder, (4) Left Elbow, (5) Right Elbow, (6) Left Wrist and (7) Right Wrist. All frame-level vectors for a recording are concatenated along the time axis to form a 14-dimensional multivariate time series with length $L = 9000$. Frames for which OpenPose does not output keypoints are replaced with feature vectors containing all zeros. Moreover, coordinates of key points in a frame for which OpenPose outputs a confidence rating below a cut-off value ($< .1$) are replaced with the mean (x, y)-values calculated from the previous and following steps in the time series. The proportion of frames affected by this procedure varies substantially across detected joints: For Neck detections, 8.5% of all frames were affected, while for Shoulder joints, this rose to 18.6%, Wrists 75.7%, and finally about 86.2% of Elbow data. These rates indicate that especially the hands are often not visible in the recordings (see *Section 6.2* for a discussion of this potential limitation).

4.2.3 Vocal Behavior: For extracting vocal-features, we first split each video's audio track into a separate file, before using *openSMILE* [17] to extract a set of low-level descriptors (*LLDs*) of the audio signal. In the *ComParE* configuration, the software provides a set of 64 descriptors forming the traditional baseline for the annual INTERSPEECH paralinguistics challenge (see, e.g., the 2020 iteration [48]). Descriptors in this set relate to different acoustic characteristics of an audio signal (see Weniger et al. for a detailed description [57]) and originate from a variety of fields (e.g., Speech Processing and Music Information Retrieval). They have been applied in a broad array of social signal processing tasks over the years to infer the states and traits of speakers. Extracting LLDs results in 65-dimensional multivariate time series of length $L = 30000$. We downsample this series to a length of $L = 9000$ for multimodal alignment.

4.3 Time Series Representation

ROCKET (RandOm Convolutional KErnel Transformation) is a recent approach for time series representation, and classification achieving state-of-the-art performance on benchmarks for this task [14]. The algorithm applies a large amount (i.e., 10000 as a default) of convolutional kernels to an input time series and then calculates aggregate features over each kernel's feature map (i.e., pooling). For ROCKET, pooling operations include computing the maximum (MAX) and proportion of positive values (PPV), i.e., the proportion of values for which the output of the convolution is positive. In contrast to convolutional kernels in deep neural networks, the parameters for these kernels (e.g., their bias, length, and dilation) are not learned from the data but are sampled at random from a range of sensible choices. Pooling features are then fed to a linear classifier for prediction (either a Ridge Classifier or Logistic Regression). MINIROCKET (MINImally-deterministic ROCKET) is an improved version of the ROCKET displaying greater computational efficiency without losing performance [14]. In contrast to a random sampling of kernel parameters, it relies on a fixed set of 84-kernels combined with a variable amount of dilation for each. Additionally, it only

uses the PPV operation for pooling. While originally developed for univariate time series, ROCKET-variants have been extended for multivariate data as well and were found to be outperforming existing alternatives in this task (including state-of-the-art deep neural network architectures) in a recent large-scale comparison of algorithms [47]. Instead of applying a kernel/dilation-combination to all channels of a multivariate input time series, the current multivariate implementation of MINIROCKET [1] assigns a random subset of channels (max. 9) to each. In our experiments, we optimize the following hyperparameters of MINIROCKET: (1) number of features of the representation produced from {10000, 20000}, and (2) the maximum amount of dilation per kernel from {32, 64}.

4.4 Classification

In line with the original MINIROCKET algorithm, we feed the time series representation resulting from the transformation to a Ridge Classifier (i.e., a linear model with L2-regularization). We denote this variant of our pipeline as MINIROCKET-RIDGE. However, this approach assumes a strictly linear decision boundary between classes in terms of the time series representation. As such, it does not account for potential interactions among the individual PPV features resulting from the transformation, each of which indicates the prevalence of a particular pattern in the input MTS. To explore the potential benefits of accounting for a non-linear relationship in modeling interdependence perceptions, we rely on a Support Vector Machine (SVM) with a Radial Basis Function (RBF)-kernel as a classifier. SVMs have been widely used in Affective Computing research [15] and are often deployed as a generic baseline for more specialized technological approaches to compare against (e.g., in dataset papers [30]). Apart from being comparatively data-efficient, SVMs are well suited for dealing with classification problems in high-dimensional spaces [20]. In the following, we refer to this variant of our pipeline *MINIROCKET-SVM*. For both models, we rely on the implementation from the python library *Scikit-Learn* [39]. Furthermore, we train both classifiers with loss-terms weighted inversely proportional to the class frequencies to ameliorate the adverse effects of the imbalanced distributions on learning.

5 EMPIRICAL INVESTIGATION

We conduct two series of machine learning experiments to explore the capacity for automatic recognition of interdependence perceptions based on nonverbal behavior. In the first one, we investigate our approach's performance based on a perceivers' own behavior. In contrast, we explore predictive performance in the second series using information about their conversation partner's behavior. In each series we collect one set of samples for the test-performance on recognizing *Mutual Dependence (MD)*, *Conflict (C)*, *Future Interdependence (FI)*, *Information Certainty (IC)*, and *Power (P)* respectively. We use these samples for statistical analysis, comparing the performance of our approach to that of a majority classifier as a naive baseline. Additionally, we compare the relative performance of the two variations of our pipeline to each other. This second comparison offers additional insights into the potential importance of accounting for non-linear relationships between the temporal pattern in the behavioral signals and interdependence perceptions.

[1] https://github.com/angus924/minirocket

5.1 Experimental Setup

For training and evaluation of our models, we rely on a *5-Fold Leave-Session-Out Cross-Validation procedure (5-Fold LOSOCV)*. This procedure creates folds so that no data points from the same dyad are simultaneously available for both training and evaluation of a model. We opt for this procedure since individual records for negotiation-phases are not independent of each other but instead were collected using a repeated measures design (negotiation-phases nested in participants, which are again nested in dyads). We repeat the LOSOC-procedure 6 times, splitting the data into new folds for each iteration. Together, this results in a total of $N = 30$ data points for the test performance on unseen data per variant of our pipeline for each targeted dimension we investigate in our experiments. Averaging across different train-test splits in such a way provides a more robust estimate of models' performance on unseen data compared to a single train-test split. Please note that splits are re-used across pipeline variations for the same target and input video to produce comparable results (i.e., the input data is identical for both a MINIROCKET-RIDGE and MINIROCKET-SVM model). Finally, for MINIROCKET-RIDGE/MINIROCKET-SVM, the training folds generated by each LOSOCV-procedure are used as a development set to optimize the hyperparameters of the machine learning algorithms. Parameters are identified through a grid search in combination with an additional inner LOSOCV. To further account for the imbalanced distribution of class labels in our data, we evaluate the performance of all our models using the *Balanced Accuracy* (Acc_B) metric. It is computed as the arithmetic mean of sensitivity (true positive rate) and specificity (true negative rate): $Acc_B = \frac{1}{2}(\frac{TP}{TP+FN} + \frac{TN}{TN+FP})$. Scores lie in the interval of $[0, 1]$, with a classifier exploiting only the prevalence of the majority class scoring $Acc_B = .5$.

5.2 Results and Analysis

We provide a graphical overview of test performance for our pipelines for each targeted interdependence dimension in *Figure 3*. Moreover, *Table 3* shows a statistical comparison of these samples against the performance score for a majority classifier (i.e., $Acc_B = .5$) [2].

5.2.1 Predictions based on Own Behavior: Comparisons against a majority classifier indicate the possibility of detecting differences in perceptions of C, FI, P by analyzing components of individuals' own nonverbal behavior. In particular, our models showed the highest average performance across all our experiments when predicting C and P perceptions. While both variants of our pipelines performed significantly above baseline for these dimensions, only the one using SVM also facilitated predictions for FI. Finally, our comparisons reveal no significant performance when targeting MD or IC.

5.2.2 Predictions based on Partner Behavior: Our experiments demonstrate that analysis of the nonverbal behavior displayed by a person's interaction partner can detect differences in perceptions of C and IC. However, neither variant of our approach offers significant insights into the MD, FI, or P dimensions.

5.2.3 Effect of Classifier on Test Performance: To assess the effect of a non-linear classifier on our pipeline's performance across the

[2] Test samples were taken from different repetitions of the 5-Fold LOSOCV procedure and as such are not independent. To control for this nesting in our statistical analysis, we rely on significance tests using clustered bootstrapping [13] ($B = 10000$ repetitions).

Table 3: Test-performance (Balanced Accuracy-Score Acc_B) of our pipelines for recognizing the intensity of interdependence perceptions, including comparison versus Majority Classifier (fixed value of $Acc_B = .5$). Significant improvements are bold.

		OWN BEHAVIOR				PARTNER BEHAVIOR			
		Acc_B		vs. Majority		Acc_B		vs. Majority	
Target	Pipeline Variant	$M(SD)$	Min/Max	$\Delta M(Acc_B)$	p	$M(SD)$	Min/Max	$\Delta M(Acc_B)$	p
MD	MINIROCKET-Ridge	.468 (0.039)	.383/.534	-.032	<.001***	.470 (0.035)	.387/.526	-.030	<.001***
	MINIROCKET-SVM	.518 (0.048)	.444/.612	+.018	.099	.461 (0.056)	.291/.500	-.039	<.001***
C	MINIROCKET-Ridge	.528 (0.054)	.390/.660	**+.028**	<.001***	.552 (0.052)	.450/.688	**+.052**	<.001***
	MINIROCKET-SVM	.568 (0.048)	.481/.681	**+.068**	<.001***	.542 (0.048)	.438/.620	**+.042**	<.001***
FI	MINIROCKET-Ridge	.505 (0.041)	.422/.570	+.005	.448	.504 (0.043)	.426/.579	+.004	.626
	MINIROCKET-SVM	.521 (0.044)	.491/.649	**+.021**	<.001***	.498 (0.036)	.427/.553	-.002	.61
IC	MINIROCKET-Ridge	.493 (0.038)	.433/.570	-.007	.108	.536 (0.058)	.403/.690	**+.036**	<.001**
	MINIROCKET-SVM	.485 (0.034)	.401/.560	-.015	<.001*	.538 (0.042)	.430/.632	**+.038**	<.001***
P	MINIROCKET-RIDGE	.533 (0.047)	.455/.660	**+.033**	<.001***	.498 (0.052)	.399/.653	-.002	.842
	MINIROCKET-SVM	.568 (0.045)	.492/.679	**+.068**	<.001***	.511 (0.049)	.422/.605	+.011	.293

MD: Mutual Depdendence; C: Conflict of Interest; FI: Future Interdependence; IC: Information Certainty; P: Power

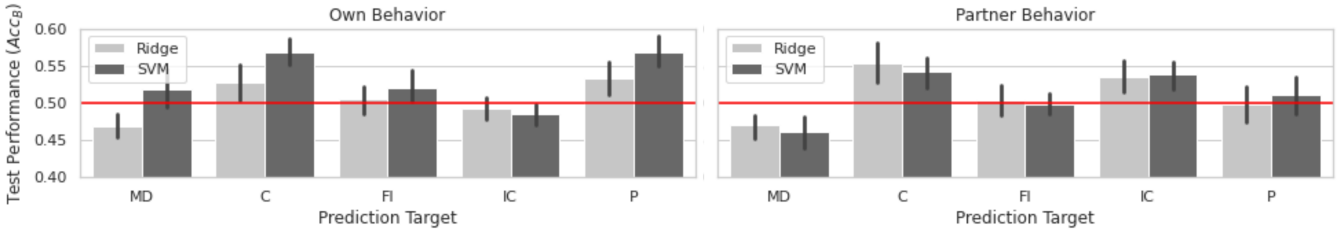

Figure 3: Test Performance (Acc_B) of our Predictive Modeling Pipeline. Error bars denote the 95% confidence interval.

different interdependence dimensions, we conduct separate multi-way ANOVAs for the data from the two experiment series. To do so, we construct linear mixed-effects models with the Acc_B-score as the dependent variable and include fixed-effects for (1) the type of interdependence dimension targeted by the pipeline (DIM), (2) the type of classifier it uses for predictions (CLF), and (3) their two-way interaction. To account for the nesting of performance samples in our analysis, we include random effects for the identity of repetitions within dimensions. For each task, we compare performance only on those dimensions with at least one classifier performing above baseline (i.e. for predictions based on own behavior C, FI and P; for other behavior C and IC). Results for test performance using own behavior reveal significant main effects for both targeted dimension (DIM_{own}: $F(2, 12) = 17.204, p < .001$) and classifier ($CLF_{own}$: $F(1, 342) = 8.26, p < .001$), but not for their two-way interaction. In contrast to this, the analysis of test performance for predictions based on partners' behavior reveals only a significant main effect for the targeted dimension ($DIM_{partner}$: $F(1, 240) = 28.326, p < .001$). Together, this indicates that the choice of classifier can have a substantial influence on the performance of predictions (i.e., for predictions based on own behavior, using SVM improves average performance).

6 DISCUSSION

6.1 Empirical Findings and Implications

The findings from our experiments demonstrate the possibility of predicting some interdependence perceptions from nonverbal behavior. In particular, they show that our approach for automatic analysis could exploit patterns in peoples' own behavior to differentiate between the intensity of some perceptions for Conflict of Interest (C) and Power (P). This result is broadly plausible, given the existing evidence on the expression of power in nonverbal behavior [7], and the strong link between negative emotional responses to situations where ones' goals are perceived as being impeded by others (e.g., anger [41]). Similarly, our results show that information about the behavior of a person's interaction partner can facilitate predictions of both Conflict of Interest (C), Future Interdependence (FI), and Information Certainty (IC). This finding also seems to agree with existing evidence for the importance of nonverbal information when inferring others' feelings, preferences, and ultimately, interdependence within a situation. Additionally, the finding that both a person's own nonverbal behavior and that of their conversation partner facilitates predictions of C indicates that these might provide complementary information that could be exploited in a joint model. Importantly, the overall pattern of predictive performance across interdependence-dimensions indicates a complementary relationship between the two sources of behavior: recognition of some interdependence perceptions requires only analysis of one's own behavior, while others need information about one's interaction partner. This pattern has crucial implications for intelligent systems since it might enable developers to make meaningful trade-offs when sensing interdependence in applications. For example, in the negotiation scenario covered by our dataset, an intelligent support system might require insights into user's perceptions of P and C to detect moments requiring guidance. Our findings suggest that it might suffice for an application to monitor only the user's own

nonverbal behavior in this setting without also requiring access to information about the partner's behavior (e.g., to preserve privacy).

However, there were several dimensions of perceived interdependence for which our approach showed very low or no performance at all in our experiments. One potential reason might be the class imbalance in the dataset for examples related to some interdependence dimensions (see also limitations below). However, this fact alone seems an unlikely cause for the drop in performance, given that examples for FI are almost balanced while examples for P are clearly not. Overall, it seems plausible that either nonverbal behavior provides no insights into these perceptions (at least in the setting covered by our dataset) or that the signal is substantially weaker and more nuanced.

Finally, the significant effect of the choice of classifier (Ridge vs. SVM) on performance demonstrates that the type of behavioral pattern picked up by the MINIROCKET in the multivariate time series can have a complex, non-linear relationship to the intensity of interdependence perceptions. For this reason, it is likely that more sophisticated classifiers built on top of ROCKET-variants (in addition to a wider variety of pooling strategies [50]) might result in better predictive performance on this task and could be used as a first step to extend our modeling approach.

6.2 Limitations and Future Work

One substantial limitation for our results is the imbalance in our dataset for some dimensions of perceived situational interdependence. While present across all dimensions, a skew in the distribution is particularly pronounced for low-intensity IC and MD. As such, research on the automatic perception of situational interdependence will benefit from datasets with more systematic variation along these dimensions. Similarly, because the interactions captured in the dataset describe only one type of situation (i.e., negotiations), it is unclear how well the behavioral patterns recognized by our models generalize to different types of interactions (e.g., team meetings or dinner dates). To better understand the latter, a collection of cross-situational datasets measuring situational interdependence will be needed. Recent events, such as the COVID-19 pandemic, have lead to a shift in the way people interact with each other - promoting remote interactions over the in-person interactions individuals are more used to engage in. Given the overall similarity in setting, i.e., being seated in front of a computer, with separate video streams for participants, our findings seem feasible to generalize to online interactions like video conferencing. Thus, a promising first target for collecting relevant cross-situational data might be interactions other than negotiations in that setting.

Finally, the overall low confidence with which wrist- and hand-joints were detected in our study shows arms may be a sparse signal in this recording format. As such, to study the role of hand-based gestures for interdependence perceptions and their automatic detection may require a recording setup that specifically strives for their inclusion.

We plan to develop context-sensitive approaches to situation perception using information about participants' background (e.g., Demographics, Personality) as part of automatic recognition in future work. Context-sensitivity is a primary challenge for Social Signal Processing [55], and Affective Computing [16], and plays an important role in humans ability to infer qualities of others' situational understanding accurately (e.g., in terms of emotional appraisals [32]). While we have focused here on exploring the general potential of nonverbal behavior for predictions, we strive for a more fine-grained understanding of the contribution provided by particular modalities and cues in the future. In particular, we aim to explore techniques providing explanations for predictions at a temporal level (e.g., using the *timeXplain*-framework [35]) to isolate particular behavioral patterns that discriminate between types of interdependence perceptions. Finally, we plan to account for the interactive nature of conversations in modeling explicitly. A first step could be to extend our approach by incorporating features from both individuals for predictions, i.e., to address phenomena related to behavioral coordination, such as mimicry [2].

7 SUMMARY AND CONCLUSION

Recognizing how individuals think and feel in social interactions is envisioned to provide intelligent systems with contextual information that is important for them to adapt their behavior in a social environment [55], especially in circumstances that require close collaboration with humans [1]. Perceptions of the interdependence present in a social interaction – how a person's outcome depends on their own actions and the actions of other parties involved – are considered an important driver for how people behave in any kind of social situation. As such, the construct of situational interdependence can provide intelligent systems with a representation for describing individuals' social experiences that is generically relevant for adaptation and personalization. Moreover, it relates to a wide variety of existing research topics in Social Signal Processing.

Prior research from psychology has implied that non-verbal behaviors can covary with perceived interdependence. However, no research has relied on multimodal analysis to investigate how non-verbal behaviors are associated with the five dimensions of perceived interdependence. In this article, we have presented a series of machine learning experiments that demonstrate the principal possibility for automatic recognition of some interdependence perceptions from nonverbal behavior during face-to-face negotiations in a dyadic setting. In particular, this includes perceptions of Conflict of Interest, Future Interdependence, Information Certainty, and Power. Moreover, our experiments indicate that behavioral signals of either the person or their conversation partner can provide complementary information for the automatic recognition of interdependence perceptions. This insight may facilitate meaningful trade-offs in developing intelligent applications, e.g., by analyzing only the behavioral source necessary for predicting particular perceptions. Finally, while recognizing interdependence perceptions is challenging, our empirical investigations' approach for predictive analysis can serve as a baseline for future technological research and improvements.

ACKNOWLEDGMENTS

This research was (partially) funded by the Hybrid Intelligence Center, a 10-year programme funded by the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research, https://hybrid-intelligence-centre.nl, grant number 024.004.022 and the MINGLE project number 639.022.606.

Data collection was funded by an ERC Starting Grant (#635356) awarded to Daniel Balliet.

REFERENCES

[1] Zeynep Akata, Dan Balliet, Maarten de Rijke, Frank Dignum, Virginia Dignum, Guszti Eiben, Antske Fokkens, Davide Grossi, Koen Hindriks, Holger Hoos, Hayley Hung, Catholijn Jonker, Christof Monz, Mark Neerincx, Frans Oliehoek, Henry Prakken, Stefan Schlobach, Linda van der Gaag, Frank van Harmelen, Herke van Hoof, Birna van Riemsdijk, Aimee van Wynsberghe, Rineke Verbrugge, Bart Verheij, Piek Vossen, and Max Welling. 2020. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer* 53, 8 (aug 2020), 18–28. https://doi.org/10.1109/MC.2020.2996587

[2] Shahin Amiriparian, Jing Han, Maximilian Schmitt, Alice Baird, Adria Mallol-Ragolta, Manuel Milling, Maurice Gerczuk, and Björn Schuller. 2019. Synchronization in Interpersonal Speech. *Frontiers in Robotics and AI* 6, November (nov 2019), 1–10. https://doi.org/10.3389/frobt.2019.00116

[3] Daniel Balliet, Joshua M. Tybur, and Paul A. M. Van Lange. 2017. Functional Interdependence Theory: An Evolutionary Account of Social Situations. *Personality and Social Psychology Review* 21, 4 (nov 2017), 361–388. https://doi.org/10.1177/1088868316657965

[4] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 59–66. https://doi.org/10.1109/FG.2018.00019

[5] Robert Böhm, Luca Carduck-Eick, J. Frederik Graff, Christine Harbring, Philipp Sprengholz, and Katja Rebecca Tilkes. 2021. Reviewing and predicting cooperation in prisoner's dilemma games: A meta-study. (2021).

[6] Franziska Burger, Mark A. Neerincx, and Willem-Paul Brinkman. 2020. Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review. *Journal of Medical Internet Research* 22, 1 (jan 2020), e12599. https://doi.org/10.2196/12599

[7] Judee K Burgoon and Norah E Dunbar. 2006. Nonverbal Expressions of Dominance and Power in Human Relationships. In *The SAGE Handbook of Nonverbal Communication*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States, 279–298. https://doi.org/10.4135/9781412976152.n15

[8] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).

[9] Dana R. Carney, Judith A. Hall, and Lavonia Smith LeBeau. 2005. Beliefs about the nonverbal expression of social power. *Journal of Nonverbal Behavior* 29, 2 (2005), 105–123. https://doi.org/10.1007/s10919-005-2743-z

[10] Simon Columbus, Catherine Molho, Francesca Righetti, and Daniel Balliet. 2021. Interdependence and cooperation in daily life. *Journal of Personality and Social Psychology* 120, 3 (mar 2021), 626–650. https://doi.org/10.1037/pspi0000253

[11] Simon Columbus, Jiří Münich, and Fabiola H. Gerpott. 2020. Playing a different game: Situation perception mediates framing effects on cooperative behaviour. *Journal of Experimental Social Psychology* 90 (2020). https://doi.org/10.1016/j.jesp.2020.104006

[12] Carsten K. W. de Dreu. 2010. Social Conflict: The Emergence and Consequences of Struggle and Negotiation. In *Handbook of Social Psychology* (5th ed. ed.), S. T. Fiske, D. T. Gilbert, and G. Lindzey (Eds.). Wiley, New York, NY.

[13] Mathijs Deen and Mark de Rooij. 2020. ClusterBootstrap: An R package for the analysis of hierarchical data using generalized linear models with the cluster bootstrap. *Behavior Research Methods* 52, 2 (apr 2020), 572–590. https://doi.org/10.3758/s13428-019-01252-y

[14] Angus Dempster, Daniel F. Schmidt, and Geoffrey I. Webb. 2020. MINIROCKET: A Very Fast (Almost) Deterministic Transform for Time Series Classification. December (2020). arXiv:2012.08791 http://arxiv.org/abs/2012.08791

[15] Sidney K. D'mello and Jacqueline Kory. 2015. A Review and Meta-Analysis of Multimodal Affect Detection Systems. *Comput. Surveys* 47, 3 (apr 2015), 1–36. https://doi.org/10.1145/2682899

[16] Bernd Dudzik, Michel-Pierre Jansen, Franziska Burger, Frank Kaptein, Joost Broekens, Dirk K.J. Heylen, Hayley Hung, Mark A. Neerincx, and Khiet P. Truong. 2019. Context in Human Emotion Perception for Automatic Affect Detection: A Survey of Audiovisual Databases. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 206–212. https://doi.org/10.1109/ACII.2019.8925446

[17] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile. In *Proceedings of the international conference on Multimedia - MM '10*. ACM Press, New York, New York, USA, 1459. https://doi.org/10.1145/1873951.1874246

[18] Robert H Frank. 1988. *Passions within reason: The strategic role of the emotions*. WW Norton & Co.

[19] Fabiola H. Gerpott, Daniel Balliet, Simon Columbus, Catherine Molho, and Reinout E. de Vries. 2018. How do people think about interdependence? A multidimensional model of subjective outcome interdependence. *Journal of Personality and Social Psychology* 115, 4 (oct 2018), 716–742. https://doi.org/10.1037/pspp0000166

[20] B. Ghaddar and J. Naoum-Sawaya. 2018. High dimensional data classification and feature selection using support vector machines. *Psychological Bulletin* 265, 3 (2018), 993–1004. https://doi.org/10.1016/j.ejor.2017.08.040

[21] Arthur C. Graesser, Mark W. Conley, and Andrew Olney. 2012. Intelligent tutoring systems. In *APA educational psychology handbook, Vol 3: Application to learning and teaching*. American Psychological Association, Washington, 451–473. https://doi.org/10.1037/13275-018

[22] Lindred L Greer and Charles Chu. 2020. Power struggles: when and why the benefits of power for individuals paradoxically harm groups. *Current Opinion in Psychology* 33 (2020), 162–166. https://doi.org/10.1016/j.copsyc.2019.07.040 Power, Status and Hierarchy.

[23] Judith Hall and Marianne Schmid Mast. 2007. Sources of accuracy in the empathic accuracy paradigm. *Emotion* 7, 2 (may 2007), 438–446. https://doi.org/10.1037/1528-3542.7.2.438

[24] Judith A. Hall, Erik J. Coats, and Lavonia Smith LeBeau. 2005. Nonverbal Behavior and the Vertical Dimension of Social Relations: A Meta-Analysis. *Psychological Bulletin* 131, 6 (2005), 898–924. https://doi.org/10.1037/0033-2909.131.6.898

[25] H. Hung and D. Gatica-Perez. 2010. Estimating Cohesion in Small Groups UsingAudio-Visual Nonverbal Behavior. (november 2010), 563–575. https://doi.org/10.1109/TMM.2010.2055233

[26] Dinesh Babu Jayagopi, Hayley Hung, Chuohao Yeo, and Daniel Gatica-Perez. 2009. Modeling Dominance in Group Conversations Using Nonverbal Activity Cues. *IEEE Transactions on Audio, Speech, and Language Processing* 17, 3 (mar 2009), 501–513. https://doi.org/10.1109/TASL.2008.2008238

[27] Dacher Keltner, Deborah H. Gruenfeld, and Cameron Anderson. 2003. Power, Approach, and Inhibition. *Psychological Review* 110, 2 (2003), 265–284. https://doi.org/10.1037/0033-295X.110.2.265 arXiv:arXiv:1011.1669v3

[28] Peter Kim, Robin Pinkley, and Alison Fragale. 2005. POWER DYNAMICS IN NEGOTIATION. *The Academy of Management Review* 30, 4 (2005), 799–822. https://doi.org/10.2307/20159169

[29] Samuel Kim, Maurizio Filippone, Fabio Valente, and Alessandro Vinciarelli. 2012. Predicting the conflict level in television political debates. In *Proceedings of the 20th ACM international conference on Multimedia - MM '12*. ACM Press, New York, New York, USA, 793. https://doi.org/10.1145/2393347.2396314

[30] Jean Kossaifi, Robert Walecki, Yannis Panagakis, Jie Shen, Maximilian Schmitt, Fabien Ringeval, Jing Han, Vedhas Pandit, Antoine Toisoul, Bjorn Schuller, Kam Star, Elnar Hajiyev, and Maja Pantic. 2021. SEWA DB: A Rich Database for Audio-Visual Emotion and Sentiment Research in the Wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 3 (mar 2021), 1022–1040. https://doi.org/10.1109/TPAMI.2019.2944808 arXiv:1901.02839

[31] Iulia Lefter and Catholijn M. Jonker. 2017. Aggression recognition using overlapping speech. *2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017* 2018-January (2017), 299–304. https://doi.org/10.1109/ACII.2017.8273616

[32] Andreas Marpaung and Avelino Gonzalez. 2017. Can an affect-sensitive system afford to be context independent?. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10257 LNAI. Springer, Cham, 454–467. https://doi.org/10.1007/978-3-319-57837-8_38

[33] Shervin Minaee, Mehdi Minaei, and Amirali Abdolrashidi. 2021. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. *Sensors* 21, 9 (apr 2021), 3046. https://doi.org/10.3390/s21093046

[34] Agnes Moors, Phoebe C. Ellsworth, Klaus R. Scherer, and Nico H Frijda. 2013. Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review* 5, 2 (apr 2013), 119–124. https://doi.org/10.1177/1754073912468165

[35] Felix Mujkanovic, Vanja Doskoč, Martin Schirneck, Patrick Schäfer, and Tobias Friedrich. 2020. timeXplain – A Framework for Explaining the Predictions of Time Series Classifiers. *arXiv* (jul 2020), 1–17. arXiv:2007.07606 http://arxiv.org/abs/2007.07606

[36] Laurent Son Nguyen and Daniel Gatica-Perez. 2015. I Would Hire You in a Minute. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, New York, NY, USA, 51–58. https://doi.org/10.1145/2818346.2820760

[37] Andrew Ortony, Gerald L. Clore, and Allan Collins. 1990. *The Cognitive Structure of Emotions*. Cambridge University Press.

[38] Yannis Panagakis, Stefanos Zafeiriou, and Maja Pantic. 2015. Audiovisual Conflict Detection in Political Debates. In *Computer Vision - ECCV 2014 Workshops*, Lourdes Agapito, Michael M Bronstein, and Carsten Rother (Eds.). Springer International Publishing, Cham, 306–314.

[39] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and Others. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, Oct (2011), 2825–2830.

[40] K. Petkanopoulou, R. Rodríguez-Bailón, G. B. Willis, and G. A. van Kleef. 2018. Powerless People Don't Yell But Tell: The Effects of Social Power on Direct and Indirect Expression of Anger. *European Journal of Social Psychology* 44 (2018), 533–547. https://doi.org/10.1002/ejsp.2521

[41] Davide Pietroni, Gerben A. van Kleef, Carsten K. W. de Dreu, and Stefano Pagliaro. 2008. Emotions as strategic information: Effects of other's emotional expressions on fixed-pie perception, demands, and integrative behavior in negotiation. *Journal of Experimental Social Psychology* 44, 6 (2008), 1444–1454. https://doi.org/10.1016/j.jesp.2008.06.007

[42] Davide Pietroni, Gerben A. van Kleef, Enrico Rubaltelli, and Rino Rumiati. 2009. When happiness pays in negotiation. The interpersonal effects of 'exit option': directed emotions. *Mind and Society* 8, 1 (2009), 77–92. https://doi.org/10.1007/s11299-008-0047-9

[43] John F. Rauthmann, David Gallardo-Pujol, Esther M. Guillaume, Elysia Todd, Christopher S. Nav, Ryne A. Sherman, Matthias Ziegler, Ashley Bell Jones, and David C. Funder. 2014. The situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology* 107, 4 (2014), 677–718. https://doi.org/10.1037/a0037250

[44] John F. Rauthmann, Ryne A. Sherman, and David C. Funder. 2015. Principles of Situation Research: Towards a Better Understanding of Psychological Situations. *European Journal of Personality* 29, 3 (2015), 363–381. https://doi.org/10.1002/per.1994

[45] S.Martin Remland. 2016. *Nonverbal communication in everyday life*. SAGE Publications.

[46] Philipp V. Rouast, Marc Adam, and Raymond Chiong. 2018. Deep Learning for Human Affect Recognition: Insights and New Developments. *IEEE Transactions on Affective Computing* 14, 8 (2018), 1–20. https://doi.org/10.1109/TAFFC.2018.2890471 arXiv:arXiv:1901.02884v1

[47] Alejandro Pasos Ruiz, Michael Flynn, James Large, Matthew Middlehurst, and Anthony Bagnall. 2020. *The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances*. Vol. 35. Springer US. 401–449 pages. https://doi.org/10.1007/s10618-020-00727-3

[48] Björn W. Schuller, Anton Batliner, Christian Bergler, Eva Maria Messner, Antonia Hamilton, Shahin Amiriparian, Alice Baird, Georgios Rizos, Maximilian Schmitt, Lukas Stappen, Harald Baumeister, Alexis Deighton MacIntyre, and Simone Hantke. 2020. The INTERSPEECH 2020 computational paralinguistics challenge: Elderly emotion, breathing & masks. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* 2020-October (2020), 2042–2046. https://doi.org/10.21437/Interspeech.2020-32

[49] Aaron Sell, John Tooby, and Leda Cosmides. 2009. Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences of the United States of America* 106, 35 (2009), 15073–15078. https://doi.org/10.1073/pnas.0904312106

[50] Chang Wei Tan, Angus Dempster, Christoph Bergmeir, and Geoffrey I. Webb. 2021. MultiRocket: Effective summary statistics for convolutional outputs in time series classification. (2021). arXiv:2102.00457 http://arxiv.org/abs/2102.00457

[51] Leigh L. Thompson and Reid Hastie. 1990. Social perception in negotiation. *Organizational Behavior and Human Decision Processes* 47, 1 (1990), 98–123. https://doi.org/10.1016/0749-5978(90)90048-E

[52] Leigh L. Thompson and Dennis Hrebec. 1996. Lose-lose agreements in interdependent decision making. *Psychological Bulletin* 120, 3 (1996), 396–409. https://doi.org/10.1037/0033-2909.120.3.396

[53] Larissa Z. Tiedens, Phoebe C. Ellsworth, and Batja Mesquita. 2000. Sentimental Stereotypes: Emotional Expectations for High-and Low-Status Group Members. *Personality and Social Psychology Bulletin* 26, 5 (2000), 560–575. https://doi.org/10.1177/0146167200267004

[54] Evert A. van Doorn, Marc W. Heerdink, and Gerben A. van Kleef. 2012. Emotion and the construal of social situations: Inferences of cooperation versus competition from expressions of anger, happiness, and disappointment. *Cognition and Emotion* 26, 3 (2012), 442–461. https://doi.org/10.1080/02699931.2011.648174

[55] Alessandro Vinciarelli, Maja Pantic, Dirk Heylen, Catherine Pelachaud, Isabella Poggi, Francesca D'Errico, and Marc Schröder. 2012. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing* 3, 1 (2012), 69–87. https://doi.org/10.1109/T-AFFC.2011.27

[56] H. L. Wagner and Jayne Smith. 1991. Facial expression in the presence of friends and strangers. *Journal of Nonverbal Behavior* 15, 4 (1991), 201–214. https://doi.org/10.1007/BF00986922

[57] Felix Weninger, Florian Eyben, Björn W. Schuller, Marcello Mortillaro, and Klaus R. Scherer. 2013. On the Acoustics of Emotion in Audio: What Speech, Music, and Sound have in Common. *Frontiers in Psychology* 4, MAY (2013), 1–12. https://doi.org/10.3389/fpsyg.2013.00292

[58] Matthias J. Wieser and Tobias Brosch. 2012. Faces in context: A review and systematization of contextual influences on affective face processing. *Frontiers in Psychology* 3, NOV (nov 2012), 471. https://doi.org/10.3389/fpsyg.2012.00471

[59] Torsten Wortwein, Louis-Philippe Morency, and Stefan Scherer. 2015. Automatic assessment and analysis of public speaking anxiety: A virtual audience case study. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 187–193. https://doi.org/10.1109/ACII.2015.7344570

[60] Zhou Yu, David Gerritsen, Amy Ogan, Alan W. Black, and Justine Cassell. 2013. Automatic prediction of friendship via multi-model dyadic features. *SIGDIAL 2013 - 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference* August (2013), 51–60.

[61] Yanxia Zhang, Jeffrey Olenick, Chu-Hsiang Chang, Steve Kozlowski, and Hayley Hung. 2018. TeamSense: Assessing Personal Affect and Group Cohesion in Small Teams through Dyadic Interaction and Behavior Analysis with Wearable Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–22. https://doi.org/10.1145/3264960