



# eNTERFACE'19

The 15<sup>th</sup> International Summer Workshop on Multimodal Interfaces

## *PROCEEDINGS* *eNTERFACE'19*

Summer Workshop  
on Multimodal Interfaces

July 8 – August 2, 2019  
Bilkent University – Turkey

Editors:  
Hamdi Dibeklioglu, Elif Sürer

# MP-BGAAD: Multi-Person Board Game Affect Analysis Dataset

Arjan Schimmel<sup>(1)</sup>, Metehan Doyran<sup>(1)</sup>, Pinar Baki<sup>(2)</sup>, Kübra Ergin<sup>(3)</sup>, Batıkan Türkmen<sup>(2)</sup>, Almıla Akdağ Salah<sup>(1)</sup>, Sander Bakkes<sup>(1)</sup>, Heysem Kaya<sup>(1)</sup>, Ronald Poppe<sup>(1)</sup>, Albert Ali Salah<sup>(1)</sup>  
<sup>(1)</sup> Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands  
<sup>(2)</sup> Department of Computer Engineering, Boğaziçi University, Istanbul, Turkey  
<sup>(3)</sup> Sahibinden.com, Istanbul, Turkey

a.e.schimmell@students.uu.nl, m.doyran@uu.nl, pinar.baki@boun.edu.tr, kubraergin3@gmail.com,  
 batikan.turkmen@boun.edu.tr, a.a.akdag@uu.nl, s.c.j.bakkes@uu.nl, hkaya@nku.edu.tr,  
 r.w.poppe@uu.nl, a.a.salah@uu.nl

**Abstract**—Board games are fertile grounds for the display of social signals, and they provide insights into psychological indicators. In this work, we introduce a new dataset collected from four-player board game sessions, recorded via multiple cameras. Recording four players at once provides a setting richer than dyadic interactions. Emotional moments are annotated for all game sessions. Additional data comes from personality and game experience questionnaires. We present a baseline for affect analysis and discuss some potential research questions for the analysis of social interactions and group dynamics during board games.

**Index Terms**—Board game, Dataset, Affect Analysis, Facial Modality, Social Interaction, Group Dynamics.

## I. INTRODUCTION

Multiplayer board games are excellent tools to stimulate specific interactions both for children and adults. Many board games have been adopted for therapeutic purposes by psychologists that work with children [1], [2]. Players of board games exhibit a wealth of social signals. As such, they enable to study of affective responses to game events and other players and emotion contagion, possibly in interaction with personal and interpersonal factors. Board games have been used by therapists to assess behavioural patterns, a child’s cognitive abilities, and attitudes [3], [4]. The assessments in turn may be employed for constructing playful interventions for children. Even though this approach is not a typical part of the toolkit of psychologists working with adults, a lot can be learned from analysing the game behaviour of adults as well.

Using board games for therapeutic purposes presents several methodological challenges. First, although a game may elicit valuable behavioural and affective responses, the amount of time when such a response can be observed during play is typically relatively brief [5]. Second, exhibited play behaviour (typically) cannot be easily annotated; accurate behavioural classifications not only depend on insight on human affect and decision-making processes, but also factors such as player personality and motivation, the state of the game, and the dynamics of the social context. Finally, manually coding a player’s behaviour is inherently time-intensive. As such, while the potential for employing board games as analysis and

intervention tools is clear, at present it generally is time-consuming for therapists to fully exploit this potential.

With rapidly developing computational approaches to behavioural analysis, it is becoming increasingly more feasible to automatically process large amounts of play observations and, if needed, prepare indices for therapists. As such, an effective computational approach to behavioural analysis would mitigate the above-mentioned challenges. First, depending on the observed behaviour, a limited number of observations may suffice for accurate analysis. Second, multiple modalities such as the face, body and the voice can be analysed simultaneously; information from one modality may be used to assess the accuracy of classifications derived from other modalities. For example, knowing a person’s head orientation may tell us something about the expected accuracy of facial expression analysis (*cf.* [6]). Third, automated analysis can be expected to be significantly more efficient than manual analysis. The drawbacks of fully automatic analysis are the limited generalization capabilities of such algorithms, their dependence on rich annotations (which may imply a small number of affective states as target variables, or in the case of continuous affect space annotations, a non-trivial mapping to practically useful labels), and the lack of a semantic grounding, which makes interpretation of rare events and idiosyncratic displays impossible. However, as the capabilities of the automatic analysis tools grow, they are expected to play larger roles in the toolbox of the analysts.

In this paper, we investigate approaches for automated multimodal behavioural analysis of adults interacting with each other while playing different types of board games. We introduce MP-BGAAD, a dataset with recordings of 62 game sessions, each involving four players. Using MP-BGAAD, we investigate to what extent we can derive information on personality traits, emotional states and social interactions of adults from recordings of their behaviour. Our setup includes the recording of videos of interacting players and the game board, the collection of personality traits for each player, and an assessment of the game experience after each played game.

This paper makes the following contributions:

- 1) We introduce a multi-person dataset with three levels of

annotations (segment-level, session-level, and user level, respectively) of recorded game sessions.

- 2) We present baseline evaluation results on this dataset by using state-of-the-art feature extraction and classification methods.
- 3) We analyse and discuss the effectiveness of the employed classifiers.

We proceed with a discussion of related works on affect analysis and datasets. Section III introduces our dataset, Multi-Person Board Game Affect Analysis Dataset (MP-BGAAD). We explain the board games, characteristics of the participants, recording setup, annotation process and the questionnaires we have used for assessing personality and game experience. In Section IV, the feature extraction process and classification methods are detailed. We present baseline scores for several automated analysis tasks in Section V and conclude with a discussion in Section VI.

## II. RELATED WORK

Affective computing aims at equipping computer programs with the ability to sense affective cues exhibited by humans, in the hope of using these for the design of more responsive interactive applications [7]. However, such analysis can also be directly used, for example by psychologists who observe, describe, and quantify behaviours during long-term therapy. Since the type of features that can be automatically derived from human behaviour analysis is vast [8], [9], a comprehensive review is not included here. Rather, we focus on the automatic analysis of player behaviour during game play.

Since the setting we use involves board games, we focus on a scenario where multiple persons are sitting around a table to play a game with materials on it. The most interesting states during such a scenario involve responses to the game events, or to other players, such as frustration, anger, elation, boredom, excitement, disappointment, concentration, puzzlement, expectation, pride and shame. Of particular importance is the display of these emotions in children, as play scenarios are particularly suitable for them. The behaviours giving away these states mostly happen above the table, so the focus lies on the upper body. While fidgeting may also be indicative, putting a camera under the table is not a desired solution.

The face is regarded as the most expressive part of the body [10], and there are works specialised in processing faces of children during game play or other activities such as problem solving [11], [12]. The eyes are in particular shown to be good indicators of a person's engagement with an activity [13], [14]. The use of the bodily motions alone in affect recognition is less common than using facial expressions [15].

One of the challenges in affect analysis with a broad range of affective states to detect is the fact that each particular affective state, with the possible exception of happiness manifesting in a smile, happens rarely. Thus, these problems are typically severely unbalanced in terms of sample distributions, and it is very important to study them in natural conditions. Facial displays are by themselves difficult to fully catch these states automatically, as the face is also deformed via non-emotional speech. The use of a multimodal system can increase the

performance. Moreover, facial and bodily modalities are the most widely used signals for automatic analysis of interactions [16]. Filntis *et al.* addressed affect recognition during child-robot interaction, and illustrated how the combination of face and bodily cues in a machine learning algorithm could yield better results than the use of a single modality [17]. A similar finding in the application domain of serious games was reported by Psaltis *et al.*, where decision level fusion was employed and the individual modalities were fused with the help of confidence levels [18].

The use of the body as a modality provides some challenges. For facial expressions, Ekman and Friesen introduced the Facial Action Coding System (FACS) [19], [20], which provides an objective way to describe facial movements of the face. However, there is no clear and unambiguous mapping from action units to expressions; there are only indicators for a number of expressions, some strongly correlated, and some not. For example, the upwards movement of lip corners, coded as AU12, is a good indicator of a smile. Yet it does not immediately tell us whether it is due to genuine enjoyment, or used as a social back-channel signal [21].

For the body, such a system does not exist. The body language associated with certain emotions is usually described by how body parts move, but it is much more idiosyncratic [22], [23].

How to represent emotions and affect is still up for debate [15]. In 1981, and Kleinginna created an overview of the definitions of emotion that existed until then [24]. This gave 92 different definitions. There has since been many works on affect and what it precisely is [25], [26]. A working definition is given by Desmet [27]: “*emotions are best treated as a multifaceted phenomenon consisting of the following components: behavioural reactions (e.g. approaching), expressive reactions (e.g. smiling), physiological reactions (e.g. heart pounding), and subjective feelings (e.g. feeling amused)*”. This definition agrees with our aims, as in this project, our ambition is to create a dataset where participants' subjective feelings during gameplay and their expressive reactions can be predicted.

Visual behaviour and affect analysis have been applied to gameplay contexts [28]. Action recognition methods are used by many researchers for analysing sports games such as tennis [29], basketball [30] and football [31]. There are game consoles (such as XBox) designed to have capabilities to analyse users through audiovisual cues, for instance for showing relevant ads to them, depending on their age or behaviour. Some researchers point out the need for emotional appraisal engines for video games in order to achieve human-like interaction between the players and the non-player characters [32], [33]. This can be achieved to a degree through visual analysis of faces through a camera [34]. Elsewhere, face and head gestures are combined with posture to recognise affective states of people playing serious games [35]. Some existing datasets provide researchers with audio, visual or audiovisual data to aid research on affective computing and social interaction analysis. The Tower Game Dataset [36], Static Multimodal Dyadic Behavior (MMDB) dataset [37], Mimicry database [38] and the PInSoRo database [39] are some of the important resources with which it is possible to

Name	Year	Modality	Subj.	Subj. per Session	Sessions	Annotations	Labels
The Tower Game Dataset [36]	2015	V + A	39	2	112	Manual, continuous	Eye gaze, body language, simultaneous movement, tempo similarity coordination and imitation are rated using a six-point Likert scale
Static MMDB Dataset [37]	2016	V + A	98	2	98	Manual, discrete	Actions are classified
Mimicry Database [38]	2011	V + A	40	2	54	Semi-automatic, discrete	- Behavioural expression labels (smile, head nod, head shake, body leaning away, body leaning forward) - Mimicry/ non mimicry labels - Conscious / unconscious
PInSoRo Dataset [39]	2018	V + A	120	1 or 2 with 1 robot	75	Manual, discrete	- Task engagement - Social engagement - Social attitude
MP-BGAAD	2019	V	58	4	62	Manual, discrete	Emotional moments are annotated based on shown expressions

TABLE I  
RECENT GAME BEHAVIOUR DATASETS. V = VIDEO, A = AUDIO

study social interactions between two adults, or child-adult and child-robot interactions.

The Tower Game Dataset [36] consists of audio-visual recordings of two players and focuses on the joint attention and entertainment during a game. Annotation of the dataset has been done with Essential Social Interaction Predicates (ESIPs). The static MMDB dataset [37] focuses on dyadic interactions between adults and 15- to 30-month old children. The dataset is annotated based on the action-reaction dynamics. Multimodal Mimicry database [38] is recorded during two experiments: a discussion on a political topic and a role-playing game, respectively. The annotation consists of a number of social signaling cues and conscious/non-conscious labels illustrating the status of these cues. The PInSoRo dataset [39] has recordings of both child-child and child-robot interactions. This dataset is annotated using three different social interaction codes, which are task engagement, social engagement and social attitude, respectively. These databases are all based on structured, short, two-person video segments. In Multi-Person Board Game Affect Analysis Dataset, four participants of a board game are recorded simultaneously during each session, which affords for more complex interactions between the participants. Table I summarises available game behaviour datasets and their characteristics.

### III. DATASET

In this section, we introduce the Multi-Person Board Game Affect Analysis Dataset (MP-BGAAD). MP-BGAAD is collected during the eNTERFACE 2019 Summer Workshop on Multimodal Interfaces<sup>1</sup>. The dataset features participants playing cooperative (co-op) and competitive board games. Every game session consisted of four participants, recorded by two separate cameras, and an additional recording of the board game itself to allow for the detection of in-game events. Every participant filled in a HEXACO personality test [40] and after every game, they completed the in-game and social modules of

the Game Experience Questionnaire (GEQ) [41]. In total, there are 62 sessions recorded. The study received ethical approval from the Internal Review Board for Ethical Questions by the Scientific Ethical Committee of Boğaziçi University.

In the following subsections, we will describe the games, participants, recordings, annotations and questionnaires. All images are reproduced with explicit permission from the participants.

#### A. Games

According to [42], there are four categories of games for therapeutic use: communication games, problem-solving games, ego-enhancing games, and socialization games, respectively. We use two types of games in the construction of the MP-BGAAD: communication games and ego-enhancing games, respectively. In communication games, competition plays a smaller role, and inter-player communication is the key [43]. Ego-enhancing games on the other hand trigger stress, feelings of competition and challenge. This potentially leads to conflicts between game players, creating emotional states like frustration, disappointment, anger, but also relief, triumph, elation, etc.

Each session consisted of four participants that played one of six multiplayer games, see Table II. The game that was played was chosen by the participants. Before playing, the rules of the game were explained by the experimenters. We briefly describe each of the six games and the benefits of using such a game.

**Magic Maze** is the most played game in MP-BGAAD. It is a cooperative game where players work together to achieve a common goal. The players win by collectively managing four game characters exploring a maze. These characters need to steal certain items from specific locations of the maze, and use specific escape locations to complete the task against a running hourglass. Players do not take turns and are allowed to move whenever they can. Each player has a complementary set of moves. The game is played in real-time and if the hourglass (green circle in Figure 2) runs out, the players lose the game.

<sup>1</sup>For more information about the workshop: <http://web3.bilkent.edu.tr/enterface19/>.





Fig. 1. A screenshot from the recording stream, where four players respond to a players mistake.



Fig. 2. A moment in a Magic Maze game, where the red cone was just placed in front of the player on the left, who is confused about his expected moves.

Players are not allowed to speak with each other during most of the game. The only way they can communicate is using a big red cone (red circle in Figure 2), which can be placed in front of another player to indicate that the other player needs to do something. In Magic Maze, players naturally show emotions due to the tension generated by the game. The time pressure prompts the players to perform well, as every wrong move will set back the group as a whole. The most stress-related emotions can be seen at moments when the hourglass is about to run out and players try to reverse it by visiting special squares in the maze. Another clear moment is during the use of the red cone. If players place it in front of another player, this is generally done with a lot of enthusiasm to prevent face loss. The player who gets it might show a number of emotions, mostly frustration or confusion (e.g. left player in Figure 2). A game of Magic Maze takes around 10-15 minutes.

**Qwixx** is a competitive game, primarily based on luck. The players throw dice every turn and take some of them to cross off numbers, based on certain restrictions, on their own sheet. At the end of the game, the player with the most crosses wins. Each action disables a number of future actions (e.g. crossing a number may disable crossing smaller or larger numbers of the same color for the rest of the game), thus

Type	Games	Sessions	Minutes	Participants
Cooperative	Magic Maze	39	405	156 (57)
	Pandemic	2	78	8 (4)
	The Mind	1	6	4 (4)
Competitive	Qwixx	10	203	40 (17)
	Kingdomino	8	140	32 (17)
	King of Tokyo	2	73	8 (5)

TABLE II  
THE GAMES PLAYED IN MP-BGAAD. NUMBERS BETWEEN BRACKETS ARE UNIQUE PARTICIPANTS.

the game requires the players to calculate and take risks. The emotions that are shown during a Qwixx game are mostly moments of surprise, both happy and sad when players see the results of the dice throw. Another commonly occurring type of emotion is ‘schadenfreude,’ i.e., enjoyment of an other player’s demise. When a player cannot cross something off, other players typically enjoy these moments.

**Kingdomino** is also a competitive game, where players compete to create the most valuable kingdom. Every turn, players take a piece of land to place it in their kingdoms. The pieces work just like domino stones and have similar placement restrictions. New pieces are revealed at the start of every turn. This typically evokes emotions such as positive and negative surprise (see Figure 3 for an example). A player’s choices directly influence the other players, as the piece of land can only be chosen by one player. This creates moments of friction between the players. In Kingdomino, boredom sometimes occurs when a player takes a long time to think. Players also take the opportunity to talk to other players to convince them to take a certain piece. Those moments show negotiation skills and how players react to each other.

**Pandemic** is a cooperative game where players try to save the world from an epidemic. Players need to work together to keep the outbreaks of diseases under control, while at the same time finding the cures for these diseases. The game decides where the next outbreak is, based on a deck of cards which players need to draw from every turn. This creates a lot of tension in these moments, because depending on which card is drawn, the game can swing in favor of the players or it

Name	Year	Modality	Subj.	Subj. per Session	Sessions	Annotations	Labels
The Tower Game Dataset [36]	2015	V + A	39	2	112	Manual, continuous	Eye gaze, body language, simultaneous movement, tempo similarity coordination and imitation are rated using a six-point Likert scale
Static MMDB Dataset [37]	2016	V + A	98	2	98	Manual, discrete	Actions are classified
Mimicry Database [38]	2011	V + A	40	2	54	Semi-automatic, discrete	- Behavioural expression labels (smile, head nod, head shake, body leaning away, body leaning forward) - Mimicry/ non mimicry labels - Conscious / unconscious
PInSoRo Dataset [39]	2018	V + A	120	1 or 2 with 1 robot	75	Manual, discrete	- Task engagement - Social engagement - Social attitude
MP-BGAAD	2019	V	58	4	62	Manual, discrete	Emotional moments are annotated based on shown expressions

TABLE I  
RECENT GAME BEHAVIOUR DATASETS. V = VIDEO, A = AUDIO

study social interactions between two adults, or child-adult and child-robot interactions.

The Tower Game Dataset [36] consists of audio-visual recordings of two players and focuses on the joint attention and entertainment during a game. Annotation of the dataset has been done with Essential Social Interaction Predicates (ESIPs). The static MMDB dataset [37] focuses on dyadic interactions between adults and 15- to 30-month old children. The dataset is annotated based on the action-reaction dynamics. Multimodal Mimicry database [38] is recorded during two experiments: a discussion on a political topic and a role-playing game, respectively. The annotation consists of a number of social signaling cues and conscious/non-conscious labels illustrating the status of these cues. The PInSoRo dataset [39] has recordings of both child-child and child-robot interactions. This dataset is annotated using three different social interaction codes, which are task engagement, social engagement and social attitude, respectively. These databases are all based on structured, short, two-person video segments. In Multi-Person Board Game Affect Analysis Dataset, four participants of a board game are recorded simultaneously during each session, which affords for more complex interactions between the participants. Table I summarises available game behaviour datasets and their characteristics.

### III. DATASET

In this section, we introduce the Multi-Person Board Game Affect Analysis Dataset (MP-BGAAD). MP-BGAAD is collected during the eNTERFACE 2019 Summer Workshop on Multimodal Interfaces<sup>1</sup>. The dataset features participants playing cooperative (co-op) and competitive board games. Every game session consisted of four participants, recorded by two separate cameras, and an additional recording of the board game itself to allow for the detection of in-game events. Every participant filled in a HEXACO personality test [40] and after every game, they completed the in-game and social modules of

the Game Experience Questionnaire (GEQ) [41]. In total, there are 62 sessions recorded. The study received ethical approval from the Internal Review Board for Ethical Questions by the Scientific Ethical Committee of Boğaziçi University.

In the following subsections, we will describe the games, participants, recordings, annotations and questionnaires. All images are reproduced with explicit permission from the participants.

#### A. Games

According to [42], there are four categories of games for therapeutic use: communication games, problem-solving games, ego-enhancing games, and socialization games, respectively. We use two types of games in the construction of the MP-BGAAD: communication games and ego-enhancing games, respectively. In communication games, competition plays a smaller role, and inter-player communication is the key [43]. Ego-enhancing games on the other hand trigger stress, feelings of competition and challenge. This potentially leads to conflicts between game players, creating emotional states like frustration, disappointment, anger, but also relief, triumph, elation, etc.

Each session consisted of four participants that played one of six multiplayer games, see Table II. The game that was played was chosen by the participants. Before playing, the rules of the game were explained by the experimenters. We briefly describe each of the six games and the benefits of using such a game.

**Magic Maze** is the most played game in MP-BGAAD. It is a cooperative game where players work together to achieve a common goal. The players win by collectively managing four game characters exploring a maze. These characters need to steal certain items from specific locations of the maze, and use specific escape locations to complete the task against a running hourglass. Players do not take turns and are allowed to move whenever they can. Each player has a complementary set of moves. The game is played in real-time and if the hourglass (green circle in Figure 2) runs out, the players lose the game.

<sup>1</sup>For more information about the workshop: <http://web3.bilkent.edu.tr/enterface19/>.

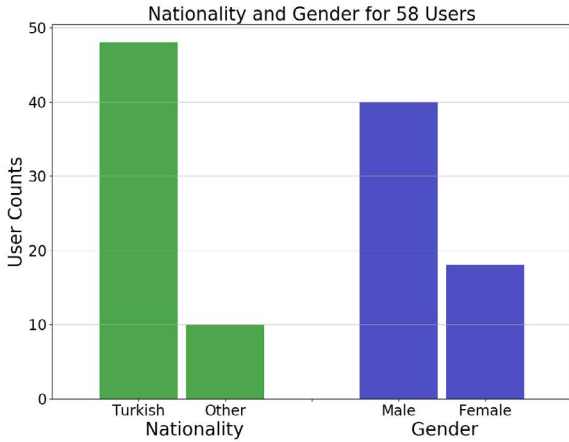


Fig. 6. Nationality and Gender histograms for all participants.

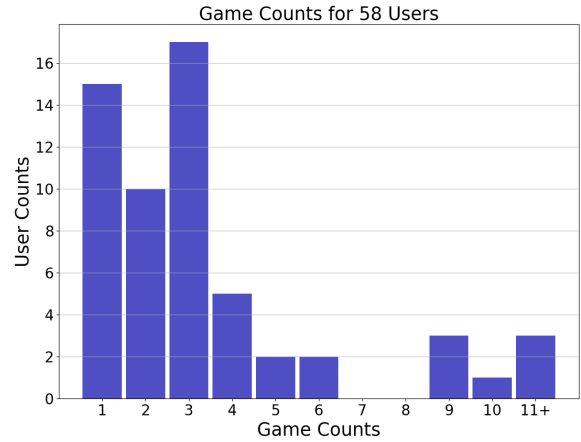


Fig. 8. Game count histogram for all participants.

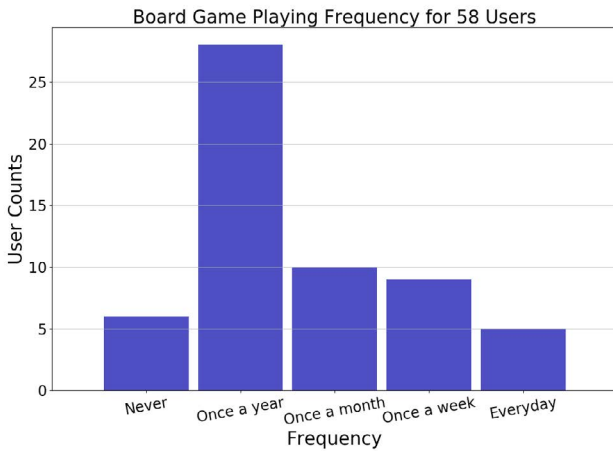


Fig. 7. Board Game playing frequency histogram for all participants.

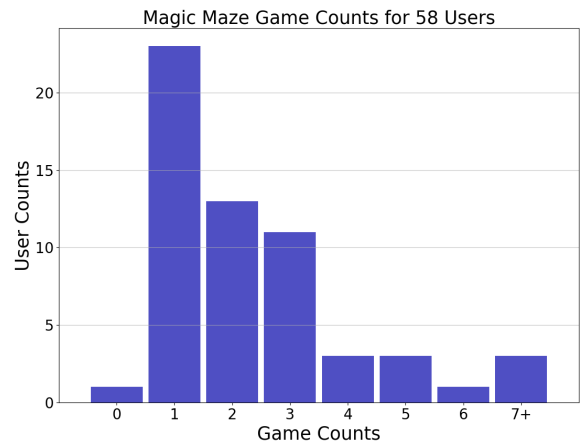


Fig. 9. Only Magic Maze game count histogram for all participants.

cameras are placed opposite to record both pairs (see left and middle frame in Figure 1). A third camera recorded the board and was placed slightly higher to have a better view (right frame in Figure 1). The three videos were merged into a single one (Figure 1) using Open Broadcaster Software<sup>2</sup> (OBS) to synchronize them for annotation purposes, but automatic processing is performed on the individual streams. The videos that were recorded of the participants (left and middle frame in Figure 1) have a resolution of  $1280 \times 720$  and the recording of the board game state (right frame in Figure 1) had a resolution of  $800 \times 448$ . All the recordings were captured in 30 fps. We decided not to focus on the audio in the recordings, because our recordings took place in a noisy environment. This would render the audio modality largely unsuitable. Furthermore, our participants were from different nationalities and they were not using their native language.

<sup>2</sup><https://obsproject.com/>

#### D. Annotation

To mark expressive moments in the videos, we annotated for each player the deviations from a neutral facial expression. We used ELAN<sup>3</sup> to create seven different annotations, Table III describes each in detail. People do not always show their emotions in the same way. For example, negative emotions can be expressed with a smile. If an anomaly shows but it was not clear what the label should be, the board game state was used to determine the label.

The dataset is annotated by two annotators with high inter-rater reliability. At the start of the annotating process, two videos were annotated by both annotators separately, and the final versions were compared. Annotators trained themselves further by discussing discrepancies in their annotations. After the training period, each video was annotated by a single annotator.

To measure the inter-rater reliability between the two annotators, we calculated Cohen’s Kappa [44] on two videos that both annotators coded. Preliminary experiments have shown a

<sup>3</sup><https://tla.mpi.nl/tools/tla-tools/elan/>

Label	Name	Meaning
+	Positive	Highest annotation in positive valence space. For example laughter and open mouth smiles.
+?	Small positive	Placed in positive valence space. Closer to the neutral state. For example gentle smiles.
'No label'	Neutral	Represents the state of the player that is generally shown throughout the game. Each player has a different neutral state, so annotations are done considering the most occurring state of that player.
-?	Small negative	Placed in negative valence space. Closer to the neutral state. For example short frowns and lowering of mouth corners.
-	Negative	Lowest annotation of negative valence space. For example looking angry and another player.
f	Focus	Not ranked in valence space. Player gives full attention to the board game. For example narrowed eyes and lower blink rate.
f?	Small focus	Less intense version of the focus label.
x	Non-Game event	For example taking a call or talking with another person outside of the game.

TABLE III  
LABELS USED IN ANNOTATION.

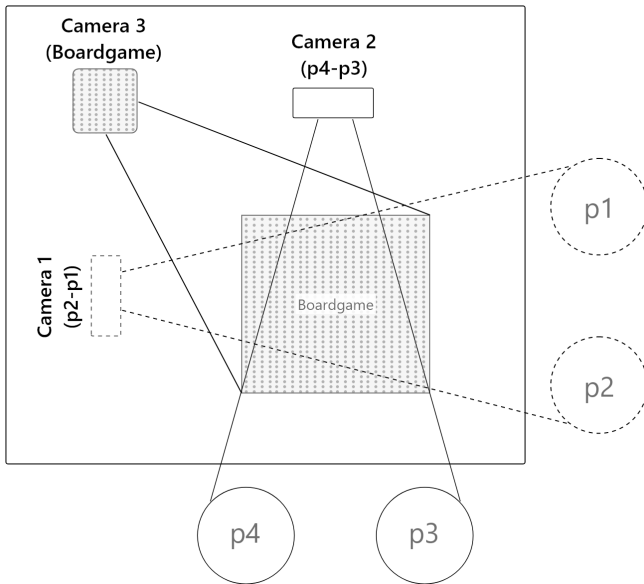


Fig. 10. Recording setup.

frame length of 50 to be adequate for segment-level coding. The Kappa score of our annotators was 0.735.

E. Questionnaires

The annotations of facial affect serve as in-game ground truth for the affective state of the player. To get social ground truths, the participants filled in two different questionnaires, which provided an opportunity to establish if there are correlations between the results of the recorded game data and the experience of the players reported by themselves. This also gave insights about the participants, and a baseline about checking if certain in-game events can be linked to certain personality traits.

Each participant filled in a 60-item HEXACO-PI-R test (HEXACO-60) [40] to assess personality in six dimensions: Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience, respectively. Participants rated 60 statement sentences from 1 (strongly disagree) to 5 (strongly agree).

After playing a game session, each player filled in a Game Experience Questionnaire (GEQ [41]). The GEQ consists of

four separate modules, which can be used individually. We used the in-game and social presence modules to evaluate the participant’s experience during the game, and to evaluate empathy, negative feelings and behavioural involvement with the other players, respectively. Players filled in the GEQ as many times as the number of game sessions they participated in. This gave MP-BGAAD 248 GEQ tests, which can be combined with the HEXACO-60 tests and in-game moments.

IV. METHODOLOGY

In this section and the next, we report some baseline approaches we have tested for the analysis of affect during board game play. We will first explain how the dataset is used to create features. Then, we will discuss how these features are used for automated detection of players’ expressiveness.

A. Feature extraction

We have used the OpenFace 2.2 [45], [46]<sup>4</sup> tool to locate faces in the video frames, and to extract facial landmark locations, head pose, eye gaze and facial expressions.

In each video, we have two players sitting side-by-side. As their relative positions do not change, tracking the nose landmark locations is sufficient to determine whether the output of OpenFace belongs to the left or right person in view. During a play session, it sometimes happens that a player reaches for something far away. The player then might appear in the recording of the other two players. To eliminate these unwanted faces that OpenFace detects, we check for clusters of outliers, corresponding to incidental face detections. To determine whether this process correctly labels the players with their assigned identities (from 1 to 4), we selected random frames and manually checked the outputs. From the selected frames, 100% was correctly labeled.

OpenFace provides a confidence score for each detection, which we used to exclude false or problematic detections from the feature set. The details of how this threshold affects the performance is presented in Table IV. Confidence thresholding gives us an improved feature set, but with missing frames and noise. To counteract these two problems we filter our feature set with a Savitzky-Golay smoothing filter [47]. We select this filter’s window length (15) and polynomial order (3) empirically.

<sup>4</sup><https://github.com/TadasBaltrusaitis/OpenFace>.



Models	Hyper-parameters		OpenFace confidence thresholds				
			0	0.25	0.5	0.75	
Random Forest	with class weights	25 trees	depth:10	.5221	.5255	.5269	.5244
			depth:25	.3998	.4002	.4071	.3977
			depth:50	.3290	.3252	.3318	.3271
		50 trees	depth:10	.5250	.5278	.5285	.5271
			depth:25	.4020	.4021	.4075	.4040
			depth:50	.3132	.3126	.3139	.3114
		75 trees	depth:10	.5272	.5290	.5288	.5286
			depth:25	.4036	.4049	.4097	.4053
			depth:50	.3277	.3266	.3286	.3249
		100 trees	depth:10	.5276	.5290	.5293	<b>.5299</b>
			depth:25	.4031	.4066	.4102	.4040
			depth:50	.3199	.3180	.3193	.3165
	no class weights	25 trees	depth:10	.3584	.3610	.3615	.3682
			depth:25	.3847	.3886	.3914	.3888
			depth:50	.3814	.3939	.3875	.3942
		50 trees	depth:10	.3659	.3760	.3776	.3761
			depth:25	.3872	.3897	.3916	.3921
			depth:50	.3730	.3856	.3829	.3860
		75 trees	depth:10	.3693	.3825	.3840	.3801
			depth:25	.3992	.4023	.3997	.4034
			depth:50	.3887	.4014	.3988	.4018
		100 trees	depth:10	.3662	.3801	.3850	.3796
			depth:25	.3947	.3991	.3994	.4008
			depth:50	.3848	.3950	.3912	.3957
ELM	10 hidden units	tanh	.2730	.2713	.2720	.2682	
		rbf	0.01	.0003	.0003	.0003	.0007
			0.1	.0003	.0003	.0003	.0007
	50 hidden units	tanh	.3681	.3730	.3730	.3715	
		rbf	0.01	.3020	.2852	.2859	.2770
			0.1	.2849	.2727	.2749	.2698
100 hidden units	tanh	<b>.3737</b>	.3686	.3705	.3719		
	rbf	0.01	.3277	.3310	.3317	.3312	
		0.1	.3287	.3304	.3322	.3307	
K Nearest Neighbors	K = 3		.2779	.3021	.3054	<b>.3110</b>	
	K = 5		.2708	.2968	.2987	.3049	
	K = 9		.2542	.2782	.2825	.2867	
	K = 15		.2371	.2622	.2646	.2666	
	K = 31		.2166	.2377	.2396	.2430	
Decision Tree	with class weights	depth:5	.4986	.5087	<b>.5089</b>	.5063	
		depth:15	.4154	.4248	.4253	.4362	
		depth:30	.3501	.3428	.3500	.3533	
	no class weights	depth:5	.4092	.3980	.4082	.4092	
		depth:15	.3958	.3916	.3963	.3885	
		depth:30	.3485	.3568	.3515	.3510	
Random			.2131				
All non-neutral			.2385				

TABLE IV  
5 FOLD CROSS VALIDATION F1 SCORES ON THE TRAINING SET.

The processed data are then used to extract some features. These features are calculated over each small segment of a video, which are created with a sliding window approach. The window length (50 frames) and stride length (16 frames) are selected based on the best inter-rater agreement calculated in Section III-D. The features are divided into three categories: head movement (24), gaze movement (8) and action units (19), respectively.

- **Head movement:** OpenFace provides us with the location of the head in millimeters with respect to the camera. The location is given in 3D coordinates. We calculate the

Action Unit	Corresponding action
AU-04	Lowering of the brow.
AU-05	Raising of the upper eye lid.
AU-06	Raising of the cheeks.
AU-07	Tightening of the eye lid.
AU-09	Wrinkle in the nose.
AU-15	Lowering of the lip corner.
AU-20	Stretching of the lip.
AU-23	Tightening of the lips.
AU-26	Dropping of the jaw.

TABLE V  
THE FACIAL ACTION UNITS USED IN THE ANALYSIS.

absolute movement of the head. The velocity and acceleration are calculated as the first and second derivative of the position with respect to time. OpenFace also provides the rotation of the head, in radians. These values can be seen as pitch, yaw, and roll. We calculate the absolute rotation to determine velocity and acceleration. For every segment, the mean and variance are calculated for the 3D coordinates of movement and pitch, yaw and roll for rotation. This provides us with 24 features for head movement.

- **Gaze movement:** OpenFace outputs the angle of the gaze by taking the average of the gaze vectors of both eyes. This creates two gaze angles in the horizontal and vertical direction. Similar to head movement, we calculated the mean and variance of the velocity and acceleration per segment. The result is eight features for the gaze.
- **Action Units:** OpenFace provides us with a subset of action units (AU), used to describe facial movements such as AU 45, which corresponds to the blink event, as well as an intensity value between 1 and 5. In the case of AU45, a value of 5 corresponds to a fully closed eye. Straightforward thresholding the smoothed intensity of AU45 as a function of time gives us the number of peaks per segment to determine the number of blinks. The other AUs that are used are shown in Table V. The mean and variance of the intensity are calculated for each AU.

### B. Classifying Emotional Moments

We use the same sliding window segmentation used in feature extraction to match our frame-level annotations to the extracted features. Out of many classifiers available, we use Random Forests [48], Extreme Learning Machines (ELM) [49], K Nearest Neighbors, and Decision Trees [50] for our segment level classification task.

Although we have several class labels as explained in Section III-D, the distribution of the classes is extremely imbalanced. The neutral class dominates the others, with 86.05% of all the video segments labelled as neutral. Consequently, we combine the minority labels into a single class called ‘non-neutral’ for the baseline experiments. We perform binary classification to classify the non-neutral segments. Since our focus is correctly classifying the non-neutral segments to perform further analysis on them, we select the F1 score, which is harmonic mean of precision and recall, as our evaluation metric.

Models	Hyper parameters		OF conf. thresh.	F1	Precision	Recall
ELM	100 units	tanh	0.25	.42	<b>.57</b>	.33
K Nearest Neighbors	K = 3		0.75	.34	.46	.26
Decision Tree	class weights	depth:5	0.5	.52	.42	.68
Random Forest	class weights	100 trees depth:10	0.75	<b>.54</b>	.50	.60
Random				.24	.16	.50
All non-neutral				.28	.16	<b>1.0</b>

TABLE VI  
TEST SET RESULTS.

V. EXPERIMENT AND RESULTS

Our experimental results, presented in this section, will serve as a baseline for future evaluations. We randomly split our dataset into 70% training set and 30% test set based on the game sessions, so that all the videos of any play session are only in one of the sets. That way we enable session-based group dynamics analysis such as social interactions and roles. Currently, we only use player based features and do not look into any social cues. We expect future research on this dataset to focus on extracting higher-level features.

Table IV shows our findings on the training set with 5-fold cross-validation comparing different hyper-parameters and OpenFace confidence thresholds. We selected the best hyper-parameters and OpenFace confidence threshold for each classifier to be used in the test set experiments. We present scores for two dummy generators in the last two lines of our tables for comparison. The first generator randomly guesses between neutral or non-neutral states, and the second one always classifies segments as non-neutral, our target class. Both in the training set and the test set experiments, the latter generator gets better F1 scores than the former.

The hyper-parameters we try in Table IV are class weights, tree counts, and maximum tree depths for Random Forest classifiers. We use class weights and maximum tree depth for Decision Trees as well. Class weights are selected inversely proportional to the class distribution in the training set. This gives higher priority to the minority class and the classifiers with class weights get the highest scores. ELM hyper-parameters are the hidden unit counts, activation functions and the in case of RBF, the kernel width.

The final results on the test set are presented in Table VI. As can be seen from the table, the best OpenFace confidence thresholds for all four classifiers are found to be different than zero. It means that some levels of thresholding cleans out the extracted data correctly. Unfortunately, ELM overfits to the neutral class and because of that its performance is lower than Random Forests and Decision Trees. ELMs are known to be affected by imbalanced data and require combination with proper data selection steps to overcome this issue [51]. In our experiments, we have not used any data selection steps and thus the different ELMs we used in our experiments performed

poorly compared to the other classifiers.

VI. CONCLUSION

We have introduced the Multi-Person Board Game Affect Analysis Dataset, MP-BGAAD, consisting of video recordings of players playing different types of board games engaging in multi-player interactions. Self-reported personality tests of all the players and the game experience questionnaires filled after every game session make this dataset open to many research directions.

We have presented some baseline scores for our frame-level affect annotations on the videos. Our test set experiments show that out of all four classifiers, the random forest with class weights to boost minority class predictions gets the highest baseline score, followed closely by a class weighted shallow decision tree. Our results show that state-of-the-art feature extraction tools and straight-forward machine learning techniques cannot get high accuracy results on our challenging dataset. We believe that these challenges will enable new research on the analysis of affect, social interaction, personality-game behaviour relationship, and game behaviour-game experience connection.

One of our future aims is extracting bodily motion features and to create multimodal classifiers. After that, we would like to create a new set of annotations which would facilitate research on social interactions and group dynamics.

ACKNOWLEDGMENT

The authors would like to thank Bilkent University for hosting eNTERFACE'19 and especially Hamdi Dibeklioğlu and Elif Sürer for organizing it.

REFERENCES

- [1] E. G. Shapiro, S. J. Hughes, G. J. August, and M. L. Bloomquist, "Processing of emotional information in children with attention-deficit hyperactivity disorder," *Developmental Neuropsychology*, vol. 9, no. 3-4, pp. 207-224, 1993. [Online]. Available: <https://doi.org/10.1080/87565649309540553>
- [2] A. I. Matorin and J. R. McNamara, "Using board games in therapy with children." *International Journal of Play Therapy*, vol. 5, no. 2, p. 3, 1996.
- [3] D. Frey, "Recent research on selective exposure to information," in *Advances in experimental social psychology*. Elsevier, 1986, vol. 19, pp. 41-80.
- [4] E. T. Nickerson and K. B. O'Laughlin, "It's fun—but will it work?: The use of games as a therapeutic medium for children and adolescents," *Journal of Clinical Child Psychology*, vol. 9, 1980.
- [5] R. A. Gardner, *The psychotherapeutic techniques of Richard A. Gardner*. Creative Therapeutics Cresskill, NJ, 1986.
- [6] P. M. Blom, S. Bakkes, C. T. Tan, S. Whiteson, D. Roijers, R. Valenti, and T. Gevers, "Towards personalised gaming via facial expression recognition," in *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2014.
- [7] R. W. Picard, *Affective computing*. MIT press, 2000.
- [8] A. A. Salah, T. Gevers, N. Sebe, and A. Vinciarelli, "Challenges of human behavior understanding," in *International Workshop on Human Behavior Understanding*. Springer, 2010, pp. 1-12.
- [9] A. A. Salah and T. Gevers, *Computer analysis of human behavior*. Springer, 2011.
- [10] F. Noroozi, D. Kaminska, C. Corneanu, T. Sapinski, S. Escalera, and G. Anbarjafari, "Survey on emotional body gesture recognition," *IEEE Transactions on Affective Computing*, 2018.
- [11] R. A. Khan, C. Arthur, A. Meyer, and S. Bouakaz, "A novel database of children's spontaneous facial expressions (liris-cse)," *arXiv preprint arXiv:1812.01555*, 2018.

- [12] G. C. Littlewort, M. S. Bartlett, L. P. Salamanca, and J. Reilly, "Automated measurement of children's facial expressions during problem solving tasks," in *Face and Gesture 2011*. IEEE, 2011, pp. 30–35.
- [13] P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera," *IEEE transactions on intelligent transportation systems*, vol. 4, no. 4, pp. 205–218, 2003.
- [14] J. Schwarz, C. C. Marais, T. Leyvand, S. E. Hudson, and J. Mankoff, "Combining body pose, gaze, and gesture to determine intention to interact in vision-based interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 3443–3452.
- [15] P. V. Rouast, M. Adam, and R. Chiong, "Deep learning for human affect recognition: Insights and new developments," *IEEE Transactions on Affective Computing*, 2019.
- [16] A. Schirmer and R. Adolphs, "Emotion perception from face, voice, and touch: comparisons and convergence," *Trends in Cognitive Sciences*, vol. 21, no. 3, pp. 216–228, 2017.
- [17] P. P. Filintisis, N. Efthymiou, P. Koutras, G. Potamianos, and P. Maragos, "Fusing body posture with facial expressions for joint recognition of affect in child-robot interaction," *arXiv preprint arXiv:1901.01805*, 2019.
- [18] A. Psaltis, K. Kaza, K. Stefanidis, S. Thermos, K. C. Apostolakis, K. Dimitropoulos, and P. Daras, "Multimodal affective state recognition in serious games applications," in *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2016, pp. 435–439.
- [19] P. Ekman and W. V. Friesen, *Manual for the facial action coding system*. Consulting Psychologists Press, 1978.
- [20] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial action coding system: The manual on cd rom," *A Human Face, Salt Lake City*, pp. 77–254, 2002.
- [21] H. Dibeklioglu, A. A. Salah, and T. Gevers, "Are you really smiling at me? spontaneous versus posed enjoyment smiles," in *European Conference on Computer Vision*. Springer, 2012, pp. 525–538.
- [22] C. Corneanu, F. Noroozi, D. Kaminska, T. Sapinski, S. Escalera, and G. Anbarjafari, "Survey on emotional body gesture recognition," *IEEE Transactions on Affective Computing*, 2018.
- [23] I.-O. Stathopoulou and G. A. Tsihrintzis, "Emotion recognition from body movements and gestures," in *Intelligent Interactive Multimedia Systems and Services*. Springer, 2011, pp. 295–303.
- [24] P. R. Kleinginna and A. M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition," *Motivation and emotion*, vol. 5, no. 4, pp. 345–379, 1981.
- [25] K. Mulligan and K. R. Scherer, "Toward a working definition of emotion," *Emotion Review*, vol. 4, no. 4, pp. 345–357, 2012.
- [26] E. Shouse, "Feeling, emotion, affect," *M/c journal*, vol. 8, no. 6, p. 26, 2005.
- [27] P. Desmet, "Measuring emotion: Development and application of an instrument to measure emotional responses to products," in *Funology*. Springer, 2003, pp. 111–123.
- [28] B. A. Schouten, R. Tieben, A. van de Ven, and D. W. Schouten, "Human behavior analysis in ambient gaming and playful interaction," in *Computer Analysis of Human Behavior*. Springer, 2011, pp. 387–403.
- [29] G. Zhu, C. Xu, Q. Huang, W. Gao, and L. Xing, "Player action recognition in broadcast tennis video with applications to semantic analysis of sports game," in *Proceedings of the 14th ACM International Conference on Multimedia*, ser. MM '06. New York, NY, USA: ACM, 2006, pp. 431–440.
- [30] M. Perše, M. Kristan, J. Perš, and S. Kovačič, "A template-based multi-player action recognition of the basketball game," in *In: Janez Pers, Derek R. Magee (eds.), Proceedings of the ECCV Workshop on Computer Vision Based Analysis in Sport Environments, Graz, Austria, 2006*, pp. 71–82.
- [31] T. Tsunoda, Y. Komori, M. Matsugu, and T. Harada, "Football action recognition using hierarchical lstm," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 99–107.
- [32] E. Hudlicka, "Affective game engines: Motivation and requirements," in *Proceedings of the 4th International Conference on Foundations of Digital Games*, ser. FDG '09. New York, NY, USA: ACM, 2009, pp. 299–306. [Online]. Available: <http://doi.acm.org/10.1145/1536513.1536565>
- [33] J. Broekens, E. Hudlicka, and R. Bidarra, *Emotional Appraisal Engines for Games*. Cham: Springer International Publishing, 2016, pp. 215–232. [Online]. Available: [https://doi.org/10.1007/978-3-319-41316-7\\_13](https://doi.org/10.1007/978-3-319-41316-7_13)
- [34] A. Cruz, B. Bhanu, and N. Thakoor, "Facial emotion recognition in continuous video," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Nov 2012, pp. 1880–1883.
- [35] A. Kapoor and R. W. Picard, "Multimodal affect recognition in learning environments," in *Proceedings of the 13th Annual ACM International Conference on Multimedia*, ser. MULTIMEDIA '05. New York, NY, USA: ACM, 2005, pp. 677–682. [Online]. Available: <http://doi.acm.org/10.1145/1101149.1101300>
- [36] D. A. Salter, A. Tamrakar, B. Siddiquie, M. R. Amer, A. Divakaran, B. Lande, and D. Mehri, "The tower game dataset: A multimodal dataset for analyzing social interaction predicates," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, Sep. 2015, pp. 656–662.
- [37] J. Rehg, G. Abowd, A. Rozga, M. Romero, M. Clements, S. Sclaroff, I. Essa, O. Ousley, Y. Li, C. Kim *et al.*, "Decoding children's social behavior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3414–3421.
- [38] X. Sun, J. Lichtenauer, M. Valstar, A. Nijholt, and M. Pantic, "A multimodal database for mimicry analysis," in *Affective Computing and Intelligent Interaction*, S. D'Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds. Springer Berlin Heidelberg, 2011, pp. 367–376.
- [39] S. Lemaignan, C. E. R. Edmunds, E. Senft, and T. Belpaeme, "The pinsoro dataset: Supporting the data-driven study of child-child and child-robot social dynamics," *PLOS ONE*, vol. 13, no. 10, pp. 1–19, 10 2018. [Online]. Available: <https://doi.org/10.1371/journal.pone.0205999>
- [40] M. C. Ashton and K. Lee, "The hexaco-60: A short measure of the major dimensions of personality," *Journal of personality assessment*, vol. 91, no. 4, pp. 340–345, 2009.
- [41] K. Poels, Y. de Kort, and W. IJsselstein, *D3.3 : Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games*. Technische Universiteit Eindhoven, 2007.
- [42] C. E. Schaefer and S. Reid, "Game play," *New York: John Wiley and Sons*, 1986.
- [43] J. P. Zagal, J. Rick, and I. Hsi, "Collaborative games: Lessons learned from board games," *Simulation & Gaming*, vol. 37, no. 1, pp. 24–40, 2006.
- [44] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.
- [45] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–10.
- [46] T. Baltrušaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 59–66.
- [47] W. H. Press and S. A. Teukolsky, "Savitzky-golay smoothing filters," *Computers in Physics*, vol. 4, no. 6, pp. 669–672, 1990.
- [48] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001. [Online]. Available: <https://doi.org/10.1023/A:1010933404324>
- [49] G.-B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107–122, Jun 2011. [Online]. Available: <https://doi.org/10.1007/s13042-011-0019-y>
- [50] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [51] U. Mahdiyah, M. I. Irawan, and E. M. Imah, "Integrating data selection and extreme learning machine for imbalanced data," *Procedia Computer Science*, vol. 59, pp. 221 – 229, 2015, international Conference on Computer Science and Computational Intelligence (ICSCSI 2015).



**Arjan Schimmel** Received his B.Sc (2016) Computing Science, with a game technology track, and M.Sc (2019) Game and Media Technology from Utrecht University, the Netherlands. His research interests is in applying modern computing technologies, like machine learning, to human behavior and interactions.



**Metehan Doyran** is a PhD candidate and a Teaching Assistant at the Information and Computing Sciences Department of Utrecht University, the Netherlands. He works at Social and Affective Computing group under Prof. Albert Ali Salah's supervision. He received his B.Sc. (2015) and M.Sc. (2018) degrees in Computer Engineering from Boğaziçi University, Turkey. His B.Sc. final project was "Dynamic Role Allocation of Soccer Robots" and he reached with his team Cerberus to quarter finals in RoboCup SPL 2015 in China. His M.Sc. thesis was "Indoor Visual

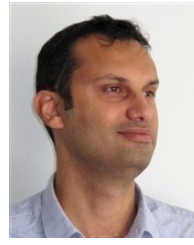
Understanding with RGB-D Images Using Deep Neural Networks". Currently his main research interest is using artificial intelligence and computer vision techniques for human behavior analysis.



**Sander Bakkes** Sander Bakkes is an assistant professor at Utrecht University Dept. of Information and Computing Science and is affiliated with Utrecht Center for Game Research. He received his Ph.D. degree in artificial intelligence in video games. His research areas are adaptive interactive environments, game personalisation, applied gaming, automated game design, procedural content generation, player experience modelling, and artificial Intelligence.



**Pinar Baki** was born in Trabzon, Turkey in 1995. She received her B.Sc. from the Computer Engineering Department of Boğaziçi University. She is currently doing her M.Sc. at the same department. For her masters thesis, she is trying to detect the mood states of bipolar disorder from multimodal data. She is also a research engineer at Arçelik, studying age, gender and emotion analysis from speech data.



**Heysem Kaya** completed his PhD thesis on computational paralinguistics and multimodal affective computing at Computer Engineering Department, Boğaziçi University in 2015. He has published over 40 papers in international journals and conference proceedings. His works won four Computational Paralinguistics Challenge (ComParE) Awards at INTERSPEECH conferences between 2014 and 2017; as well as three ChaLearn Challenge Awards on video-based personality trait recognition (ICPR 2016) and explainable computer vision & job candidate screening (CVPR 2017) competitions. His team was the first runner up in video-based emotion recognition in the wild challenge (EmotiW 2015 at ICMI). His research interests include mixture model selection, speech processing, computational paralinguistics, explainable machine learning and affective computing. He serves in editorial board of SPIIRAS Proceedings, as well as reviewer in more than 20 journals including IEEE Trans. on Affective Computing, Neural Networks and Learning Systems, Multimedia, Image and Vision Computing, Computer Speech and Language, Neurocomputing, Speech Communication, Digital Signal Processing and IEEE Signal Processing Letters. He is a faculty member in the Social and Affective Computing Group of the Department of Information and Computing Sciences, Utrecht University.

up in video-based emotion recognition in the wild challenge (EmotiW 2015 at ICMI). His research interests include mixture model selection, speech processing, computational paralinguistics, explainable machine learning and affective computing. He serves in editorial board of SPIIRAS Proceedings, as well as reviewer in more than 20 journals including IEEE Trans. on Affective Computing, Neural Networks and Learning Systems, Multimedia, Image and Vision Computing, Computer Speech and Language, Neurocomputing, Speech Communication, Digital Signal Processing and IEEE Signal Processing Letters. He is a faculty member in the Social and Affective Computing Group of the Department of Information and Computing Sciences, Utrecht University.



**Kübra Ergin** was born in Turkey in 1993. She received the Industrial Product Design B.S. degree from Istanbul Technical University, İstanbul, Turkey, in 2018. After graduating, she started to work as a User Experience Designer at sahibinden.com, which is a classified web site, İstanbul, Turkey until 2019. Currently, she is working as a User Experience Designer in Accenture Industry X.0, İstanbul, Turkey. She is interested in cognitive science, cognitive design methods, digital design and digital interactive art installations. Her main purpose is to combine machine learning methods with aesthetic production processes.

machine learning methods with aesthetic production processes.



**Ronald Poppe** received a Ph.D. in Computer Science from the University of Twente, the Netherlands (2009). He was a visiting researcher at the Delft University of Technology, Stanford University and University of Lancaster. He is currently an assistant professor at the Information and Computing Sciences department of Utrecht University. His research interests include the analysis of human behavior from videos and other sensors, the understanding and modeling of human (communicative) behavior and the applications of both in real-life settings. In

2012 and 2013, he received the most cited paper award from Image and Vision Computing. In 2017, he received a TOP grant from the Dutch Science Foundation.



**Batkan Türkmen** received his M.Sc. degree in computer engineering from Boğaziçi University, Turkey in 2019 under the supervision of Prof. Albert Ali Salah. He obtained his B.Sc. degree in computer engineering from Bilkent University, Turkey in 2015. His research interests lie in the field of affective computing and deep learning to address human behavior analysis.



**Albert Ali Salah** is a Full Professor of Social and Affective Computing at Utrecht University, the Netherlands. He works on multimodal interfaces, pattern recognition, computer vision, and computer analysis of human behavior. He has over 150 publications in related areas, including the edited books Computer Analysis of Human Behavior (2011) and Guide to Mobile Data Analytics in Refugee Scenarios (2019). Albert has received the inaugural EBF European Biometrics Research Award (2006), Boğaziçi University Foundation's Award of Research Excellence (2014), and the BAGEP Award of the Science Academy (2016). He serves as a Steering Board member of eNTERFACE and ACM ICMI, as an associate editor of IEEE Trans. on Cognitive and Developmental Systems, IEEE Trans. Affective Computing, and Int. Journal on Human-Computer Studies. He is a Senior Member of IEEE, member of ACM, and senior research affiliate of Data-Pop Alliance.

search Excellence (2014), and the BAGEP Award of the Science Academy (2016). He serves as a Steering Board member of eNTERFACE and ACM ICMI, as an associate editor of IEEE Trans. on Cognitive and Developmental Systems, IEEE Trans. Affective Computing, and Int. Journal on Human-Computer Studies. He is a Senior Member of IEEE, member of ACM, and senior research affiliate of Data-Pop Alliance.



**Almıla Akdağ Salah** is a digital humanities scholar. Her research focuses on developing methodologies to close/distant reading of cultural objects with machine learning tools. Currently she works on two projects: one investigating the history of sex and body in the western culture, the second analyzing the effect of breathing on trauma. She is an associate professor of Industrial Design at İstanbul Şehir University, and works as an adjunct faculty at Utrecht University, Department of Information and Computing Sciences.