Sequential Patterning of Facial Actions in the Production and Perception of Emotional Expressions

WITH, Stéphane, KAISER WEHRLE, Susanne

Abstract

Despite the fact that the facial expressions of emotions are naturally dynamic social signals, their communicative value has typically been studied using static photographs. In this paper, we focus on the perception of emotions from naturally occurring, dynamic facial displays produced in social interactions. In describing their impressions of 200 video records of spontaneous emotional expressions produced during a face-to-face emotional sharing task, observers were found to agree on five emotional factors: enjoyment, hostility, embarrassment, surprise, and sadness. FACS coding and sequential analysis using a pattern detection algorithm showed that recordings rated high on one emotional factor were characterized by unique sequences of facial actions coordinated with eye and/or gaze actions. Our results suggest that the dynamic unfolding of facial displays and their combination with additional nonverbal signals may play an important and still under-investigated role in emotion perception in face-to-face interactions.

Reference


DOI: 10.1024/1421-0185/a000062
Sequential Patterning of Facial Actions in the Production and Perception of Emotional Expressions

Stéphane With and Susanne Kaiser

University of Geneva

Author Note

Stéphane With, Department of Psychology, University of Geneva; Susanne Kaiser, Department of Psychology, University of Geneva.

This research was made possible by a grant from the Swiss Center for Affective Sciences.

Correspondence concerning this article should be addressed to Dr. Stéphane With, Department of Psychology, University of Geneva, 40, Boulevard du Pont-d’Arve, CH-1205 Geneva, Switzerland.

E-mail: stéphane.with@unige.ch
Abstract

Despite the fact that the facial expressions of emotions are naturally dynamic social signals, their communicative value has typically been studied using static photographs. In this paper, we focus on the perception of emotions from naturally occurring, dynamic facial displays produced in social interactions. In describing their impressions of 200 video records of spontaneous emotional expressions produced during a face-to-face emotional sharing task, observers were found to agree on five emotional factors: enjoyment, hostility, embarrassment, surprise, and sadness. FACS coding and sequential analysis using a pattern detection algorithm shows that recordings rated high on one emotional factor were characterized by unique sequences of facial actions coordinated with eye and/or gaze actions. Our results suggest that the dynamic unfolding of facial displays and their combination with additional nonverbal signals may play an important and still under-investigated role in emotion perception in face-to-face interactions.

Keywords: facial expressions, emotions, FACS, sequential patterns, THEME
Sequential Patterning of Facial Actions in the Production and Perception of Emotional Expressions

The established empirical evidence demonstrating that seven emotion labels, namely, anger, sadness, happiness, disgust, contempt, surprise, and fear, can be reliably and cross-culturally attributed to specific facial configurations is based on the results of emotion “recognition” studies using static photographs of posed expressions presented in isolation and pre-selected for maximum discriminability (Barret, Lindquist, & Gendron, 2007). The generalization of such “recognition” data to the empirical study of emotional signal processing in the context of interpersonal interactions can be questioned on several grounds. First, most of the standardized facial stimuli used in those experiments (see Matsumoto & Ekman, 1988) are produced by actors following strict guidelines about how to pose for each expression according to a set of predefined prototypes (Ekman, 2007). By contrast, naturally occurring facial expressions have been shown to be of weaker intensity, less clear cut, and their interpretation more elusive and ambiguous than posed ones (Motley, 1993; Motley & Camden, 1988; Russell, 1997; Yik, Meng, & Russell, 1998). Moreover, several studies using the Facial Action Coding System (FACS; Ekman, Friesen, & Hager, 2002) to specify the characteristics of spontaneously produced facial displays report very low frequencies of the prototypical expressions predicted to signal basic emotions (Camras et al., 2002; Matias & Cohn, 1993; Scherer & Ellgring, 2007). Second, in face-to-face interactions, facial displays are only one component of an integrated multi-signal communicative system in which additional elements act as contextual cues that may modulate the meaning of facial displays.

Emotional signals are best characterized as a complex combination of rapidly changing and overlapping individual facial actions coordinated with additional nonverbal cues like gaze and head orientation, position, and movements. The few studies that have
investigated the combination of facial displays with such additional signals suggest that head orientation (Hess, Adams, & Kleck, 2007), body postures (Aviezer et al., 2008), head positions (Krumhuber, Manstead, & Kappas, 2007), and gaze orientation (Adams & Kleck, 2005) all have a modulating impact on the meaning derived from facial displays. Finally, when it comes to collecting relevant empirical data about how facial emotions are perceived, traditional “recognition” studies using static portrayals are unable to control for the possibility that the dynamic components of facial expressions provide unique information that may lead to differential effects on emotion perception. Starting with original findings on the importance of duration of facial displays for the detection of deceptive signals (Ekman, Friesen, & O’Sullivan, 1988), evidence has started to accumulate concerning the importance of the dynamic parameters of facial expression for the recognition of subtle facial expressions of emotions (Ambadar, Schooler, & Cohn, 2005; Bould & Morris, 2008) and judgments of genuineness (Krumhuber & Kappas, 2005) and trustworthiness (Krumhuber, Manstead, Cosker et al., 2007). All this points to the possibility that the perception of emotions from static and dynamic facial stimuli might involve distinct cognitive processes and that until recently behavioral scientists may have seriously underestimated the importance of motion dynamics for making sense of subtle or otherwise ambiguous facial expressions that permeate real-life situations. To date, attention to the dynamic components of facial expressions has been limited to the study of prototypical expressions of emotions.

Studies that use prototypical expressions to assess the meaning of facial expressions, regardless of whether they are presented as static portrayals or as a dynamic sequence, typically follow a top-down approach: A certain facial expression is presented and participants are to attribute an emotion term to it. This method assumes that concordance in the attributions of an emotion label to a predefined set of expressions
allows the conclusion that the attributed emotion is the reason for this facial expression. We feel that a bottom-up approach is needed if we are to move beyond confirmatory research designs and attempt to discover new behavioral patterns that generate further hypothesis testing and model development. Componential theories of emotions differ from the prototypical view of facial expressions. One central claim of the componential view is that single elements of facial expressions might convey meaningful information at more molecular levels than full-blown prototypes (Smith & Scott, 1997). Componential accounts of facial expressions are derived from appraisal theories of emotions (for a review, see Scherer, 2001). Appraisal theories claim that emotions are elicited and differentiated by conscious and/or non-conscious evaluations of events that are relevant to an individual’s goals and needs. Amongst the appraisal theories, Scherer’s multi-componential model of emotions is particularly interesting because it predicts that the temporal sequence in which individual facial actions unfold are a reflection of an individual’s ongoing cognitive processing of emotionally relevant stimuli (Wehrle, Kaiser, Schmidt, & Scherer, 2000). If this is indeed the case, an individual facial action could be likened to a word which when positioned at an appropriate location in a string of words contributes to the production of a meaningful sentence. This opens up the possibility that observers attribute different emotional values to morphologically similar facial actions depending on their sequential organization during emotionally expressive sequences. This theoretical account of the links between facial displays and emotions does not predict a single facial prototype for a limited set of basic emotions and can thus explain the existence of multiple patterns for a single emotion family including the frequent occurrence of single actions and small combinations of muscular groups. Our general hypothesis is that individual facial actions are temporally structured in a way that is both perceptible and emotionally meaningful to an observer. Additionally, we want to
extend the investigation of the temporal sequence in which facial actions unfold to how they are coordinated in time with head and eye motion, positions and interpersonal orientation. Such coordinated dynamic patterns, if they can be shown to exist (see Delplanque et al., 2009, for suggestive evidence), may be perceived as communicating specific emotional meanings beyond what has been described in traditional studies of emotion recognition from facial expressions.

**Method**

**Emotional Expressions Database**

First, we needed to develop a database of dynamic facial expressions that were natural enough and emotional enough to be used for purposes of this study. We used an emotion-sharing task to elicit affect. To elicit emotional memories, we used a set of propositions corresponding to situation parameters predicted to elicit the emotions of anger, fear, guilt, sadness, and contempt according to appraisal theories (see Scherer, 2001). We used specific appraisal profiles to elaborate items to guide participants in recollecting and selecting five distinct narratives of autobiographical emotional memories\(^1\). At no time were the participants told the type of emotion expected; the only constraint was that the retrieved event was to correspond to the suggested features detailed in the appraisal items. We counterbalanced the order of the five emotional appraisal profiles in order to neutralize the potential effects of fatigue and habituation.

**Participants**

Participants were 16 Caucasian females recruited by ads placed in various university facilities (\(M = 31\) years, \(SD = 14\)). Of the 16 participants, 13 completed the protocol. Three participants were unable to produce all the required narratives. We did not use the videos of three additional participants due to technical problems: The first was not included because the audio recording did not function; the second because the
participant often moved out of the camera’s frame: and the third because the camera unexpectedly stopped during the recording session.

**Procedure**

We videorecorded the interviews in our lab at the University of Geneva. Participants were seated on a chair and their faces were videotaped by a hidden camera located in a cupboard in front of them. The interviewer was seated at a 45° angle to the right at a distance of one and a half meters from the participant. The role of the interviewer was to present the instructions for the tasks and to introduce the items for the guided recollection of personal memories. Once the participants were ready to begin their storytelling, the interviewer would start listening without engaging in conversation. Support in the form of backchannels signals or minimal empathic statements were allowed when appropriate. The procedure was repeated four times until each participant had produced five narratives. At the end of the procedure, the participants were debriefed about the purpose of the study and permission to use the recordings for scientific purposes was requested. Each participant received CHF 25 as remuneration for their participation.

**Selection of Judgment Stimuli**

All the analyses reported in this paper are based on a core set of 200 files extracted from the 50 face-to-face semi-structured interviews conducted with the participants who took part in the emotional narrative task. We instructed two undergraduate psychology students, not otherwise involved in the study, to independently view the 50 videos and time-mark the beginning of sequences in which participants seemed to be experiencing an emotion. Segments on which the judges agreed yielded an initial database of 350 clips. In order to avoid giving more weight to the expressive style of some participants, we randomly selected 20 video clips per participant (corresponding
to the number of video clips obtained from the least expressive participant), ending up with a core set of 200 video sample files. The rule for extracting a segment of video was to include the full propositional unit accompanying, preceding, or following the emotional sequence. The mean duration of an emotional sequence is 5.88 s (minimum: 1.8 s; maximum: 15 s).

**Rating Task**

To collect data on the emotional message value communicated by the 200 expressive sequences in our core set of videos, we used a multi-scalar rating task. Our rationale for selecting descriptive adjectives was that they correspond to affective states that observers can recognize from non-verbal cues. Most of the previous research on the meaning of facial expressions has been conducted within a basic emotion framework. According to this perspective, the emotional states that can be recognized from facial expressions include: *surprise* (a valence-neutral expression), *anger, disgust, sadness, fear, and contempt* (negative expressions), and happiness (a positive expression) (Ekman, 1999). Ekman contends that all positive states share the same facial configuration. However, recent research has suggested that differentiated positive emotion states may differ in their facial expressions (Mortillaro, Mehu, & Scherer, 2010; Shiota, Campos, & Keltner, 2003). Therefore, we decided to break the traditional “happy” category down into several categories referring to distinct positive emotions (*joyful, enthusiastic, relieved, affectionate, entertained, proud*). We also added the adjectives “*embarrassed*” and “*nervous*” to the traditional list of “basic” emotions. This is justified on the basis of Keltner’s (1995) report that, looking at a combination of facial expressions and specific eyes and head movements, observers can accurately identify such states. Finally, besides emotions, observers have been found to agree that facial expressions can also signal cognitive states (Scherer & Grandjean, 2008). Thus, in addition to the more traditional
emotion terms, we tentatively included a few adjectives that have a more cognitive overtone to their meaning. These include: *perplexed*, *ironic*, and *disappointed*. For each video recording, participants judged the relevance of a total of 17 adjectives to report their impressions of the affective states being communicated — *embarrassed*, *disgusted*, *ironic*, *proud*, *surprised*, *nervous*, *entertained*, *sad*, *scornful*, *joyful*, *affectionate*, *angry*, *enthusiastic*, *anxious*, *perplexed*, *disappointed*, and *relieved*. Relevance ratings were done using a bipolar scale labeled: “Not at all” and “A lot” with no gradient categories in between. We implemented the task in a computer program written in MATLAB so that participants could work by themselves in the laboratory. Prior to testing, we standardized all stimuli for size by converting each video sequence to the size of 1200 x 1200 pixels.

Participants were seated in front of a Dell PC wired to a 17-in screen. We kept the viewing angle and distance constant by seating participants on a chair fixed at a 26-in distance from the front of the screen. Before the actual rating task started, the instructions were presented on the screen and a mock trial was launched to make sure that the participants had understood and could comply with the instructions. After the participant’s questions were answered and the participant understood the task, the program started. The experiment ended after the last item had been rated. By default, the cursor appeared at the center of the first scale at the top of the screen. The cursor did not move to the next scale item until the participant had given his response by moving the cursor to the desired location on the line and entering his response by a mouse click. To neutralize order-of-presentation effects, both the position of the adjectives on the screen and the order of the video files were presented randomly by the program for each trial.

**Participants in the Rating Study**

Forty-five individuals participated in the judgment study (*M* = 26.5 years, *SD* = 5.89). All were Caucasians, native French-speaking women recruited through ads posted
in various university facilities. Participation was limited to women because it is well-established that they are more accurate judges of the emotional meaning of nonverbal cues than men (Hall & Matsumoto, 2004). To limit the impact of cognitive overload and fatigue on the experimental results, participants were not asked to rate all 200 video samples. Instead, they were randomly assigned to one of three rating blocks composed of 15 individuals each. Each group was to rate a unique subset of the original core set of stimuli. We ensured group equality by including at least six videos from each of the ten encoders across the three rating groups so that no encoder would be over-represented in any group. Furthermore, to control expression equality across rating groups, we selected portrayals that were most dissimilar in terms of types of action units presented in the portrayals of each encoder. Five videos, not part of the core set, were added to the three blocks. This step was taken in order to ensure that the judges used the scales consistently across the three groups. Group One included 72 videos, Group Two 72, and Group Three 71. The Cronbach’s \( \alpha \) computed on the sum of the 17 scale scores for the five video samples common to the three rating blocks was high (\( \alpha = 0.94 \), standardized, \( \alpha = 0.95 \), inter-item correlations = 0.20). Therefore, we felt justified pooling the ratings of the three groups for further analysis.

**Judgment Study Results**

**Reliability Validation**

To assess internal reliability, we computed Cronbach’s alphas for each adjective scale. Setting an inferior threshold of \( \alpha = 0.70 \), the analysis supported the internal reliability of 14 adjective scales. Three adjectives (viz., proud, relieved, and affectionate) for which internal reliability was not supported were excluded from subsequent analysis.
Principal Components Factor Analysis

To reduce the number of remaining terms, we performed a principal-components factor analysis with Varimax rotation on the ratings with the terms as variables and the participants and video records as cases. Using eigenvalues > 0.90, the analysis yielded a five-factor solution explaining > 78% of variance. We assigned each adjective to one factor only, the highest factor loading determining the factor to which an adjective was assigned. The loadings are listed in Table 1.

Insert Table 1 about here

The adjectives with the highest loadings on the first factor are “entertained,” “enthusiastic,” and “joyful.” As these adjectives refer to pleasant affects, we labeled the first factor “enjoyment.” The adjectives with the highest loadings on the second factor are “angry,” “scornful,” and “disgusted.” As these adjectives all connote rejecting or opposing something/someone, we labeled the second factor “hostility.” As the adjectives “nervous” and “embarrassed” loaded highest on the third factor, we labeled it “embarrassment.” Similarly, we named the fourth factor “surprise” (highest loadings: “surprised” and “perplexed”) and the fifth factor “sadness” (highest loadings: “sad” and “disappointed”).

K-Means Clustering

To form groups of video samples that were rated most similarly on the five factors extracted from the previous PCA, we performed a K-means clustering on the means of the factor scores of the video ratings. In order to maximize the probability of characterizing each group of video samples by one factor, we imposed a five-group
solution on the analysis. Figure 1 shows that this attempt proved adequate since all the clusters do indeed relate strongly to one specific factor.

Our judgment study provides evidence that observers can attribute reliable emotional meaning to the expressive displays constitutive of the stimuli database.

**Emotional Expressive Sequences Coding**

All facial activity in the video stimuli was annotated according to the Facial Action Coding System (FACS, Ekman et al., 2002). FACS is a comprehensive and anatomically based coding system designed to measure all visually distinguishable facial activity on the basis of action units (AUs). The manual annotation was performed frame-by-frame using the Anvil software (Kipp, 2004). Annotation is performed on multiple tracks by inserting time-anchored codes that can be further specified by different attribute values. During the annotation process, each individual AU was coded in separate runs. Each facial AU was assigned a duration, delimited by its onset and offset times. Asymmetries and laterality of movement were also scored. In addition to the FACS action units, we provided additional codes for gaze orientation and head position, movement, and social orientation. In this context, “social orientation” refers to either the eyes or the head being directed at or away from the interviewer. Gaze social orientation was scored with respect to two modalities: The participant could be scored either as “looking at” or “looking away from” the interviewer. We determined several gaze codes: looking straight, looking to the side, looking up or down, blinking, and eyes closed (gaze orientation was not coded when the participant was blinking or had her eyes closed). Head social orientation was coded either as being oriented towards (“head on”) or away
from ("head away") the interviewer. Head movement codes included: chin descending or ascending on a vertical axis (head descending; head ascending); diagonal movements upwards or downwards (head ascending and turning; head descending and turning); horizontal movements either to the right or to the left (head turning); lateral head tilts (head tilting). Two actions performed with the head are also coded: nods of the head (as in saying “yes”); shakes of the head (as in saying “no”). A last category of head codes involved head positions that were maintained for at least 2 s. These position codes were: head down; head raised, head turned away, and head tilted laterally.

**Coding Reliability Assessment**

Facial action coding was performed by two trained FACS coders. The reliability of FACS scoring was assessed on two levels of analysis (Sayette, Cohn, Wertz, Perrott, & Parrott, 2001): 1) agreement on the occurrence of individual FACS action unit scoring; and 2) temporal precision of individual AU scoring for onsets and offsets. For the “non-FACS” codes, we computed indexes of intra-individual reliability for the first author. This was done by re-scoring 30% of the data set, with a one-year interval between the two sessions. First, Cohen’s kappa coefficients were computed for the occurrences of single action units. All facial action units (except four: AUs 11, 13, 18, and 22) showed good to excellent kappas ranging from 0.64 to 1. The four AUs with kappas below 0.40 were dropped from subsequent analysis. Second, to assess the precision of scoring, we used tolerance time windows of 0 and 5 frames, which correspond to a reliability of $1/25^{th}$ and $1/6^{th}$ of a second, respectively. Coders were considered to agree on the occurrence of an AU if they both identified it within the same time window. Using a $1/6^{th}$-second tolerance window, all the upper and lower face AUs showed good to excellent reliabilities for “onset” scoring ($\kappa = 0.63$-$0.97$). The results are comparable for “offset” scoring ($\kappa = 0.59$-$0.89$). The vast majority of additional non-FACS scores showed good
to excellent coefficients for occurrences ($\kappa = 0.70-0.97$) as well as for the beginning of an event ($\kappa = 0.60-0.90$) and the end of an event ($\kappa = 0.62-0.91$). Two exceptions are acceptable scores for the onset of *Head Tilting* ($1/25^{th}$ - $\kappa = 0.49$ and $1/60^{th}$ – $\kappa = 0.58$) and the offset of *Head Raising* at $1/25^{th}$ of a second ($\kappa = 0.58$). The reliability analysis indicates satisfactory agreement between coders at an exact frame resolution for all lower and upper face AUs that have been shown to play a role in emotional expressions. The scoring of additional nonverbal categories was found to be stable with respect to time and was used in further analysis.

**Data Analysis**

**The T-Patterns Model**

The data analysis pursued two aims. The first was to detect probabilistic patterns in the stream of facial, gaze, and head actions in our data set. Our interest was in testing the hypothesis that the sequence in which behavioral events unfold during emotionally expressive displays is not random, but structured in time following reliable and identifiable patterns. The second aim was to assess whether specific sequential patterns of nonverbal actions could be linked to the uniques clusters of videos sharing a similar judgment profile. An adequate pattern detection algorithm takes into account the structure of the behavioral stream being studied. Traditional sequence analysis is based on component order alone to detect regularities in time series. This can be problematic when, as is the case here, the number of variables is high and the transitions between them are rapid. We use the T-pattern developed by Magnusson (2000, 2006). A T-pattern is essentially a combination of events. Here the events are the activation of specific FACS action units. More specifically, the events in a T-pattern occur in the same temporal order and with fairly similar time durations between each event. The T-pattern concept is
shown in Figure 2. Imagine that each letter on the first line represents a distinct FACS action unit.

Insert Figure 2 about here

The underlying line stands for the time period in which the successive actions unfold. Note how, with only six different facial actions, it already becomes difficult to visually detect the repetition of a simple \([(ab)(cd)]\) sequence at first glance (see sequence in second row). In each of the T-patterns, the components occur in a particular order and the temporal distance between one component and the next lies within a defined time window that is characteristic for the particular pattern. If these time distances between components become too short or too long, the pattern disappears. These constraints are essential for the detection of patterns of rapidly changing expressive actions. They are often impossible to detect on the basis of component order alone. This is due to the fact that a varying number of random behaviors can occur between their components. The T-pattern detection algorithm is implemented in the THEME 5 software (Noldus) that was used to perform our analysis.

**Setting Parameters to Detect T-Patterns**

Setting search parameters for the detection of T-patterns has a critical influence on the kind and number of patterns that THEME detects. The first decision is to determine the number of times a pattern has to occur to be detected. We decided to be conservative and to retain only patterns that occurred at least 15 times in any given cluster of videos. A significance level of 0.001 was set as the accepted probability to determine how far from random expectation critical interval relationships could occur for patterns to be kept or dropped. The next decision was related to the minimum percentage
of video samples within a judgment cluster in which a pattern should occur to be
detected. By default, in case of a high rate of occurrences of a pattern that is present in
only one or two samples of a cluster, it will still be detected. In order to report only the
patterns that are present across a maximum number of records, we set up a 60% samples
threshold for patterns detection.

**T-Patterns Search Results**

The T-pattern search algorithm was applied separately to the concatenated
annotation files constitutive of each judgment cluster. Each code specified in the coding
scheme was fed into the analysis. The examination of the general characteristics of the T-
patterns found in the five clusters yielded the following information. First, the number of
T-patterns found in a cluster is linearly related to the number of files composing a cluster.
Nonetheless, the proportion of T-patterns involving at least one FACS action decreases
when the number of sample files increases (see Table 2).

Insert Table 2 about here

We propose two non-mutually exclusive possible explanations for this
phenomenon. The first concerns the type of coding involved. FACS codes are “event”-
type codes whereas a majority of the additional codes are “state” codes. By definition,
“event” codes are scored based on their frequency of occurrence and vary from one file to
the other. On the other hand, “state” codes are scored positive on all the sample files.
Only the frequency of transition states from one modality of the variable to another varies
across the annotation files. Second, the frequency of transition states is sensitive to the
time scale most characteristic of a specific variable. Some variables, like eye or head
movements, can be so rapid and pervasive that their frequency of occurrence increases
dramatically, compared to less frequent and longer lasting facial actions, with the number
of files involved in a cluster. We think that these two factors combined are the main reason why the proportion of T-patterns with FACS codes decreases when cluster size increases. Considering that we decided to keep only the patterns that included at least one facial action code, we were able to retain, depending on the cluster, from 2% to 24% of all the T-patterns originally detected by THEME. Because we are interested in patterns that are sequential in nature, we reduced the data set further by excluding all T-patterns containing transition lags between two events that were equal to zero. Lags of zero indicate a simultaneous “onset” of codes on two or more annotation tracks. Across the five clusters, 493 T-patterns were detected that fit this requirement.

**T-Patterns Validation**

The detection of T-patterns is tested possibly millions of times when exploring for patterns in a single data set. Obviously, many would thus be significant even if the data were random. To address this issue, each search in the experimental data was followed by a search in a shuffled version of the same data, that is, after the time points in each series in the real data have been randomly redistributed over the observation period. This way the size of the data set remained the same. By repeatedly shuffling and then searching for patterns in the same data set, an occurrence distribution with a mean and a standard deviation is obtained for each pattern length. Table 3 shows the number of T-patterns detected in the annotation files for the five clusters and the mean number of random patterns detected after 10,000 randomizations of the same data. The significance of the difference between the number of independent patterns found in the experimental data compared to that found in the simulated data is computed on the basis of the distance, expressed here in standard deviations, between the number of “experimental” patterns and the distribution of the simulated values around their mean.

Insert Table 3 about here
The large number of standard deviations reported in Table 3 strongly suggests that the T-patterns found in the five clusters of videos are not the result of chance effects for a *p*-value of 0.001.

**T-Patterns Composition**

When considering the composition of the T-patterns found in the five clusters of expressive sequences, two complementary questions need to be addressed. First, are the event types (facial, gaze, or head actions) belonging to the composition of T-patterns in one cluster different from those in the others? Second, does the position of the codes in the chain of events defining the T-patterns differ from one cluster to the other? In Table 4, we report the proportion of facial actions that are found in independent T-patterns across the five clusters.

Insert Table 4 about here

Four AUs were found to play a role in the composition of T-patterns of all judgment profiles: AU1 and AU2 acting together to lift the eyebrows, usually producing long horizontal wrinkles across the forehead; AU5 which raises the upper eyelids, exposing the white sclera above the iris; and AU17 which raises the chin. By contrast, four AUs were shown to be specific to a single cluster. The first was AU6, which was only found in the enjoyment cluster. In the emotion expression literature, AU6, which raises the cheeks and produces “crow’s feet-like” wrinkles extending radially from the outer corners of the eye aperture is commonly referred to as the “Duchenne marker,” in memory of French anatomist Duchenne de Boulogne who first highlighted it as a distinctive sign of felt enjoyment when associated with a smile. The second display is a
unilateral lip raise (AU10) only found in the expressions that communicate a sense of hostility. According to Ekman & Friesen (2003), AU10 is a distinctive sign of disgusted expressions. The third AU to be associated with a single cluster was a unilateral smiling action (AU12U), which was present in the embarrassment cluster. Finally, AU16, which designates a downward action of the lower lip that uncovers the lower teeth, was only found in the hostility cluster. The 13 remaining action units were common to at least 2 clusters in varying proportions. Our analysis also showed that, in T-patterns, facial actions were aligned with nonverbal actions of the head and gaze. The modalities of gaze actions were more evenly distributed across the clusters than the facial actions. Each code in this category was found in the composition of T-patterns of at least three of the five clusters (see Table 5).

Insert Table 5 about here

Concerning the head codes, shaking the head and maintaining the head in a raised position were found exclusively in the hostility cluster. By contrast, lowering the head and maintaining the head in a diverted position relative to the interviewer was most pervasive in the embarrassment T-patterns.

The examination of the distribution of the various event types in the T-patterns across the five clusters of video files showed both commonalities and specificities. In order to address the question of whether the sequential organization of event types in the T-patterns can be linked with the categorization task performed by the observers on the basis of their impression judgments, we examined the number of shared T-patterns across the five clusters. Of the 493 patterns considered, only 4 (< 1%) were found to share an identical structure across cluster boundaries.
We conclude that the T-patterns detected by THEME demonstrate the existence of a reliable sequential organization of the individual components of the emotional expressions contained in the core set of emotional expressions. These expressive patterns are composed of facial actions often coordinated in time with gaze and head actions. These behavioral sequences appear to be highly related to the way independent observers characterize the emotional meaning communicated in the video records, as is evidenced by the lack of substantial overlap of T-patterns across the judgment clusters.

**T-pattern Illustrations**

For illustration purposes, we will now present three distinct T-patterns extracted from the enjoyment, hostility, and embarrassment clusters. These patterns represent repetitive instances of behavioral sequences that have been statistically validated by THEME. To illustrate how the message value of similar facial actions can be modulated by the preceding and following actions in an expressive sequence, we selected patterns that all shared a smiling action (AU12 in the FACS system). Patterns were implemented in the FACSGen tool for illustration (Roesch, Tamarit, Reveret, Grandjean, Sander, & Scherer, 2011). The first T-pattern illustrated in Figure 3 comes from the enjoyment cluster. It starts with an orientation of the gaze towards the interviewer, followed by a cheek raise action (AU6). The next event in the sequence is a raising of the eyebrows, which is directly followed by the onset of a smile. The sequence ends when AU6 recedes while the smile continues.

Insert Figure 3 about here

The second T-pattern is specific to the hostility cluster (see Figure 4). Here, the sequence starts with a smile, followed by AU14, which tightens the corners of the mouth,
pulling the corners somewhat inwards, and narrowing the lip corners. The next event is the offset of the smile followed by the start of an action that tightens the lips, making the red parts of the lips appear narrower (AU23). The sequence ends with the offset of AU14.

Insert Figure 4 about here

The third T-pattern was detected in the embarrassment cluster. It is interesting because it only contains one facial action in a behavioral sequence in which head and gaze actions are dominant. The pattern starts with the participant looking at the interviewer, then the head starts to go down and the eyes turn away to the left side. At this point the smiling action starts and the sequence ends with the head turning away to the left on its horizontal axis.

Insert Figure 5 about here

**Discussion**

The present study demonstrated that the emotional message value of dynamic expressions of emotions can be linked to statistically verified and recurrent temporal patterns of behavioral events composed of facial, gaze, and head actions. The expressive actions in these T-patterns are not necessarily linked by a direct sequential contingency but by relatively invariant temporal relations not identifiable by traditional linear analyses. The total number of T-patterns was significantly correlated with the total number of behaviors coded, but the relative percentages of T-patterns containing facial actions were not. Thus, the detection of T-patterns with facial actions do not appear to be an artefact of more extensive behavioral coding in those judgment clusters that contained
more video files. These findings represent a new and unique contribution to the study of
dynamic emotional expressions as they occur in face-to-face interactions. The behavioral
codes aligned in a T-pattern reveal an underlying sequential and rhythmic organizational
structure in which facial displays are coordinated in time with head and gaze movements,
positions, and social orientation. Moreover, in each cluster of videos rated as conveying
either enjoyment, hostility, embarrassment, surprise, or sadness, THEME detected several
specific behavioral sequences not found in the other clusters. This suggests that the
dynamic unfolding of nonverbal actions in emotional expressions may play a substantial
role in emotion perception and that there may well exist more than one expressive
phenotype for the efficient communication of categorically distinct emotional states.

The findings of this study were generated with several notable limitations. The
first is related to the fact that the judgment study was conducted with the sound in the
videos turned on. Consequently, verbal statements may have acted as important clues to
guide observers in their ratings of the video samples. Nevertheless, even if speech content
was not adequately controlled for as a source of information to classify the emotional
expressions, one would still have to account for the lack of any substantial overlap of the
T-pattern structures across the different clusters. Another limitation concerns the nature
of the T-patterns themselves. By definition, the strength of the T-pattern approach is that
it reduces the signal-to-noise ratio by disregarding event types that are present in a
behavioral stream but that do not fall into the boundaries of the critical intervals
computed by the algorithm. Therefore, we cannot exclude the possibility that some
specific event types that are not part of T-patterns but nevertheless occur more often than
not within spitting distance could have acted as important, yet undetected cues for
expression classification. Also, our undertaking is limited by the lack of inclusion of male
participants in the production and perception phases of the study. This limits the
generalizability of our findings. Finally, by sub-sampling the video clips to be used in the rating study by converging on the least expressive participant, we limited our dataset. Therefore, we may have missed additional information contained in these videos.

At this time, we can conclude that a definitive association between the behavioral sequences detected by THEME and the perception of the emotional message value conveyed by the dynamic emotional expressions used in this study exists, but the nature of that link needs to be further specified. Currently, we are in the process of transposing the T-patterns into the FACSGen software (Roesch et al., 2010) in order to produce dynamic movies that will be used in judgment studies in order to assess the exact communicative impact of the behavioral sequences highlighted in this research. We are convinced that the data presented in this manuscript constitute a first critical step to move away from simple static prototypical expression recognition data to the understanding of how emotional meaning is attributed to dynamic facial expressions as they occur in face-to-face interactions.
References


Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. Scherer, A. Schorr, & T. Johnston (Eds.), Appraisal processes in


Footnote

\footnote{1 Appraisal items used in this research can be obtained from the first author.}
Table 1

*Factors Loadings of the Emotion Rating Scales on the Factors*

<table>
<thead>
<tr>
<th>Adjectives</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry</td>
<td>0.12</td>
<td>0.87</td>
<td>-0.01</td>
<td>0.08</td>
<td>0.10</td>
</tr>
<tr>
<td>Anxious</td>
<td>0.28</td>
<td>0.12</td>
<td>0.41</td>
<td>-0.21</td>
<td>-0.65</td>
</tr>
<tr>
<td>Disappointed</td>
<td>0.26</td>
<td>-0.43</td>
<td>-0.08</td>
<td>0.10</td>
<td>0.71</td>
</tr>
<tr>
<td>Disgusted</td>
<td>0.16</td>
<td>0.84</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.32</td>
</tr>
<tr>
<td>Embarrassed</td>
<td>0.06</td>
<td>0.29</td>
<td>0.82</td>
<td>0.03</td>
<td>-0.11</td>
</tr>
<tr>
<td>Entertained</td>
<td>0.83</td>
<td>0.13</td>
<td>0.13</td>
<td>0.12</td>
<td>0.31</td>
</tr>
<tr>
<td>Enthusiastic</td>
<td>0.87</td>
<td>0.10</td>
<td>-0.24</td>
<td>-0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>Ironic</td>
<td>-0.48</td>
<td>-0.14</td>
<td>0.22</td>
<td>0.25</td>
<td>0.41</td>
</tr>
<tr>
<td>Joyful</td>
<td>0.93</td>
<td>0.12</td>
<td>0.03</td>
<td>0.05</td>
<td>0.16</td>
</tr>
<tr>
<td>Nervous</td>
<td>-0.04</td>
<td>-0.07</td>
<td>0.89</td>
<td>-0.13</td>
<td>-0.05</td>
</tr>
<tr>
<td>Perplexed</td>
<td>0.24</td>
<td>0.07</td>
<td>0.24</td>
<td>0.79</td>
<td>-0.15</td>
</tr>
<tr>
<td>Sad</td>
<td>0.30</td>
<td>0.11</td>
<td>0.15</td>
<td>0.16</td>
<td>0.82</td>
</tr>
<tr>
<td>Scornful</td>
<td>0.03</td>
<td>0.84</td>
<td>-0.15</td>
<td>0.09</td>
<td>0.15</td>
</tr>
<tr>
<td>Surprised</td>
<td>-0.08</td>
<td>0.06</td>
<td>-0.05</td>
<td>0.85</td>
<td>0.22</td>
</tr>
</tbody>
</table>
Table 2

Number and Percentages of T-Patterns in Rating Clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Sample files</th>
<th>Patterns Tot.</th>
<th>FACS Tot.</th>
<th>FACS Sequential</th>
<th>% of FACS Tot.</th>
<th>% of Tot. Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoyment</td>
<td>14</td>
<td>134</td>
<td>32</td>
<td>25</td>
<td>75</td>
<td>18</td>
</tr>
<tr>
<td>Surprise</td>
<td>26</td>
<td>241</td>
<td>38</td>
<td>27</td>
<td>71</td>
<td>11</td>
</tr>
<tr>
<td>Hostility</td>
<td>54</td>
<td>3903</td>
<td>260</td>
<td>224</td>
<td>86</td>
<td>6</td>
</tr>
<tr>
<td>Embarrassment</td>
<td>51</td>
<td>2069</td>
<td>116</td>
<td>100</td>
<td>86</td>
<td>5</td>
</tr>
<tr>
<td>Sadness</td>
<td>55</td>
<td>7605</td>
<td>143</td>
<td>117</td>
<td>81</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>13952</td>
<td>589</td>
<td>493</td>
<td>84</td>
<td>4</td>
</tr>
</tbody>
</table>

Note. Sample files = number of video files in a cluster; Patterns Tot. = total number of T-patterns found in a cluster; FACS Tot. = total number of T-patterns containing FACS codes; FACS Sequential = Number of T-patterns containing FACS codes where transition lags between two events are > 0 (Lags of zero indicates a simultaneous onset of two or more events in a T-pattern); % of FACS Tot. = Percentage of Sequential FACS T-patterns with respect to the total number of T-patterns containing FACS codes; % of Tot. Patterns = Percentage of Sequential FACS T-patterns with respect to the total number of T-patterns found in a cluster.
Table 3: Comparison of T-patterns in the experimental dataset and after randomization

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number-pat. exp.</th>
<th>Length-max exp.</th>
<th>Length-mean exp.</th>
<th>Length-mode exp.</th>
<th>Mean-number rando</th>
<th>Length-max rando</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoyment</td>
<td>134</td>
<td>6</td>
<td>3.70</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>23.10</td>
</tr>
<tr>
<td>Hostility</td>
<td>3903</td>
<td>7</td>
<td>3.99</td>
<td>4</td>
<td>78</td>
<td>3</td>
<td>9.95</td>
</tr>
<tr>
<td>Embarrassment</td>
<td>2069</td>
<td>6</td>
<td>3.40</td>
<td>4</td>
<td>0.5</td>
<td>2</td>
<td>687.00</td>
</tr>
<tr>
<td>Surprise</td>
<td>241</td>
<td>5</td>
<td>3.00</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>57.53</td>
</tr>
<tr>
<td>Sadness</td>
<td>7605</td>
<td>9</td>
<td>4.50</td>
<td>4</td>
<td>19</td>
<td>2</td>
<td>32.28</td>
</tr>
</tbody>
</table>

Number-pat. exp.: number of T-patterns in the experimental dataset; Length-max exp.: maximum length of pattern in experimental dataset; Length-mean exp.: length mean of patterns in experimental dataset; Length-mode exp.: length mode of patterns in experimental dataset; Mean-number rando: mean number of patterns found after randomisation; Length-max rando: maximum length of pattern after randomisation; Standard deviation: between the number of "experimental" patterns and the distribution of the simulated values around their mean.
Table 4

Percentages of FACS Action Units in Independent T-Patterns Across the Five Rating Clusters

<table>
<thead>
<tr>
<th>Codes / Clusters</th>
<th>Enjoyment</th>
<th>Hostility</th>
<th>Embarrassment</th>
<th>Surprise</th>
<th>Sadness</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU1+2 Inner and outer brow raise</td>
<td>28</td>
<td>30</td>
<td>32</td>
<td>26</td>
<td>22</td>
</tr>
<tr>
<td>AU4 Brow lowerer</td>
<td>0</td>
<td>17</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>AU5 Upper lid raiser</td>
<td>26</td>
<td>31</td>
<td>30</td>
<td>63</td>
<td>13</td>
</tr>
<tr>
<td>AU6 Cheek raiser</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AU7 Lid Tightener</td>
<td>0</td>
<td>23</td>
<td>19</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>AU9 Nose Wrinkler</td>
<td>0</td>
<td>28</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AU10 Upper lip raiser</td>
<td>4</td>
<td>7</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AU10U Unilateral upper lip raiser</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AU12 Lip corner puller</td>
<td>56</td>
<td>12</td>
<td>38</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>AU12U Unilateral lip corner puller</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AU14 Dimpler</td>
<td>8</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AU14U Unilateral dimpler</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>AU15 Lip corner depressor</td>
<td>0</td>
<td>15</td>
<td>17</td>
<td>0</td>
<td>42</td>
</tr>
<tr>
<td>AU16 Lower lip depressor</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AU17 Chin raiser</td>
<td>12</td>
<td>25</td>
<td>23</td>
<td>33</td>
<td>22</td>
</tr>
<tr>
<td>AU20 Lip stretcher</td>
<td>8</td>
<td>15</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AU23 Lip tightener</td>
<td>0</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>AU24 Lip pressor</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>AU25 Lips part</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>7</td>
</tr>
<tr>
<td>AU26 Jaw drops</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>22</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 5

Percentages of Gaze Codes and Head Codes in Independent T-Patterns Across the Five Rating Clusters

<table>
<thead>
<tr>
<th>Codes / Clusters</th>
<th>Enjoyment</th>
<th>Hostility</th>
<th>Embarrassment</th>
<th>Surprise</th>
<th>Sadness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blink</td>
<td>12</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Eyelids droop</td>
<td>0</td>
<td>3</td>
<td>23</td>
<td>19</td>
<td>40</td>
</tr>
<tr>
<td>Look at</td>
<td>32</td>
<td>4</td>
<td>0</td>
<td>22</td>
<td>8</td>
</tr>
<tr>
<td>Look away</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>4</td>
<td>52</td>
</tr>
<tr>
<td>Look down</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>7</td>
<td>39</td>
</tr>
<tr>
<td>Lower head</td>
<td>0</td>
<td>4</td>
<td>21</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Head raise</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Head raise and turn</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Head raised</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Head on</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Head turned away</td>
<td>4</td>
<td>4</td>
<td>12</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Head tilting side</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Head shake</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 1. K-means clustering. Plot of means on factors for each cluster.
Figure 2. Exemplifying a T-pattern encompassing four hypothetical events.
Figure 3. Enjoyment cluster. Sequence of event types in T-pattern: a) look at; b) AU6, c) AU1+2; d) AU12; e) stop AU6.
Figure 4. Hostility cluster. Sequence of event types in T-pattern: a) AU12; b) AU14; c) Stop AU12; d) AU23; e) Stop AU14.
Figure 5. Embarrassment