Facial expressions as predictors of long-term outcomes following a traumatic event:

Comparing automated and manual coding systems.

Charlotte Cooper Pfeffer

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2023
Abstract

Facial expressions as predictors of long-term outcomes following a traumatic event: Comparing automated and manual coding systems.

Charlotte Cooper Pfeffer

Human faces provide a rich source of behavioral data. Following acute, potentially traumatic events, manual and automated coding systems of facial behavior may help identify individuals at risk for developing psychopathology. In the present study, OpenFace, an automated system, and FACS, a manual method, were compared as predictors of long-term functioning using facial behavioral data from clinical interviews collected one-month after a potentially traumatic event that brought participants into the emergency department of a Level-1 Trauma Center in New York City. We evaluated similarities and differences in facial emotions identified by FACS and OpenFace to determine their predictive accuracy in capturing Depression and PTSD 6-months and 12-months later. The findings suggest OpenFace is a more sensitive and precise measure of facial behavior than FACS.
Table of Contents

List of Tables and Figures ................................................................. iii
Acknowledgements ........................................................................ iv
Dedication ...................................................................................... v
Chapter 1: Introduction ................................................................... 1
  1.1 The Importance of Risk-Identification ....................................... 2
  1.2 Facial Behavior as a Transdiagnostic Indicator of Clinical Functioning .... 5
  1.3 Manual Coding Systems .......................................................... 9
  1.4 Automated Coding Systems .................................................. 12
  1.5 Comparing Automated and Manual Coding Systems .................. 15
Chapter 2: The Current Study .......................................................... 18
Chapter 3: Methods ......................................................................... 20
  3.1 Participants and Procedures .................................................. 20
  3.2 Measures .............................................................................. 22
  3.3 Manual Coding of Facial Expressions ..................................... 22
  3.4 Automated Coding of Facial Expressions ................................ 24
Chapter 4: Results ............................................................................ 26
  4.1 Data Analysis Overview ....................................................... 26
  4.2 Descriptive Statistics ............................................................ 26
  4.3 Correlations Between Facial Coding Methods ......................... 28
  4.4 Predicting Long-term Adjustment with Emotion Codes ............ 28
Chapter 5: Discussion ...................................................................... 36
  5.1 Summary and Explanation of Key Findings ............................. 36
5.2 Implications

5.3 Limitations

5.4 Future Directions

5.5 Conclusion

References

Appendix A: FACS Training Materials
List of Tables and Figures

Table 1: Descriptive Statistics for FACS and OpenFace (OF) Magnitude Scores and Psychopathology Measures…………………………………………………………………………………27

Table 2: Inter-Correlations between FACS and OpenFace (OF) Magnitude Scores and Psychopathology……………………………………………………………………………………………30

Table 3: Hierarchical Linear Regression of OpenFace Sadness on Depression at 6-months……31

Table 4. Hierarchical Linear Regression of OpenFace Sadness on Depression at 12-months…..32

Table 5. Hierarchical Linear Regression of OpenFace Anger on Post-traumatic stress disorder at 6-months………………………………………………………………………………………33

Table 6. Hierarchical Linear Regression of OpenFace Anger on Post-traumatic stress disorder at 12-months………………………………………………………………………………………33

Table 7. Hierarchical Linear Regression of OpenFace Fear on Post-traumatic stress disorder at 6-months………………………………………………………………………………………34

Table 8. Hierarchical Linear Regression of OpenFace Fear on Post-traumatic stress disorder at 12-months………………………………………………………………………………………35

Figure 1. Total Number of AUs Identified by FACS Coders Across All Participant Videos…..29
Acknowledgments

This dissertation would not have been possible without the exceptional mentorship and guidance by my doctoral advisor, Dr. George Bonanno. Thank you for your unparalleled mentorship over the years and for imparting your wisdom, skills, and knowledge. I would also like to thank the members of my dissertation committee, Dr. Amelia Aldao, Dr. Doug Mennin, Dr. Cindy Huang, and Dr. Philippa Connelly for generously sharing your empirical insights, expertise, and enthusiasm. Lastly, I would like to thank my dearest colleagues in the Loss, Trauma, and Emotion lab and members of my cohort for your support over the years.
Dedication

To Andrew and Harte.
Chapter 1: Introduction

Human faces reveal rich information about behavior. The study of facial expressions as indicators of functioning and psychopathology has a long history. Charles Darwin was one of the first to suggest that facial expressions communicate essential information about emotions, intentions, opinions and mental states (Darwin, 1872/1965). Since then, decades of research have demonstrated that emotional expression and valence can be used to predict diverse forms of psychopathology (Gaebel & Wölwer, 1992; Bonanno & Keltner, 1997; Gehricke & Shapiro, 2000; Renneberg et al., 2005). For example, facial expressions indicate emotion (Ekman, 1993; Russell, 1994) and pain (Craig, Hyde, & Patrick, 1991), regulate social behavior (Cohn & Elmore, 1988; DePaulo, 1992; Fridlund, 1994), and reveal pathology (Katsikitis & Pilowsky, 1988; Rinn, 1984). To capitalize on the wealth of behavioral data provided by facial expressions, several different coding systems have emerged.

The Facial Action Coding System (FACS) is a comprehensive, manual coding method that was for many years the gold standard in measuring facial behavior (Ekman & Friesen, 1978; Ekman, Friesen & Hager, 2002). More recently algorithm-based AI methods, such as OpenFace, have been developed to capture digital phenotypes and assess psychological outcomes (e.g., Huckvale, Venkatesh, & Christensen, 2019; Galatzer-Levy et al., 2020). Both methods have proved useful, but so far have not been compared as predictors of long-term functioning and psychopathology. Furthermore, at present there are no proven methods to predict the development of psychological functioning following an acute trauma (Shalev et al., 2019). The current study aims to address these deficits, by comparing the efficacy of both FACS and an automated coding system, OpenFace, as predictors of psychopathology following a potentially traumatic event. Identifying an accurate and efficient method to identify high-risk individuals
following admission to the emergency department after a trauma would allow for early targeted treatment interventions and reduce subsequent psychopathology in high-risk populations (Schultebraucks et al., 2020).

1.1 The Importance of Risk-Identification

The Federal Emergency Management Agency (FEMA) model, which funds mental health programs in response to disasters and terrorist attacks in the United States, assumes that most individuals will be affected by a potentially traumatic event. This approach emphasizes crisis counseling, support services, public education and outreach for all affected individuals (Brewin et al., 2008). For example, following the September 11th, 2001 terrorist attacks, the FEMA-funded Project Liberty provided crisis counseling to over 690,000 individuals in accordance with this model (Stuber et al., 2006). Unfortunately, six months later a survey revealed there was still a significant need for PTSD and depression treatment, particularly amongst individuals who still had no contact with mental health services (Stuber et al., 2006). Consequently, an enhanced services program was implemented two years later to address severe and persistent clinical presentations (Donahue, Lanzara, Felton, Essock, & Carpinello, 2006). Recipients of the enhanced services, which included a 10-12 session cognitive-behavioral intervention, exhibited a reduction in depression and grief symptoms but not PTSD symptoms (Donahue, Jackson, Shear, Felton, & Essock, 2006). These findings suggest an approach that aims to reach all affected individuals may inadvertently divert resources and treatment from direct victims with serious disorders. Without timely treatment interventions, this can lead to worsening chronic and treatment-resistant clinical presentations.

In contrast, the ‘screen and treat approach’ uses validated measures to assess all trauma-exposed individuals and provides evidence-based treatment to those who develop persistent
symptoms of psychopathology, monitoring outcomes using standardized instruments (Brewin et al., 2008). In the aftermath of terrorist bombings on the London transport system in 2005, rather than divert resources towards crisis counseling and public education, the public health response focused specifically on screening and advising directly affected individuals. Medical practitioners and hospitals referred survivors to a screening team, who collected information regarding the extent of their involvement in the bombings and current symptoms of psychopathology. Individuals who screened positive on mental health measures were invited for a more extensive assessment, and a clinical decision was made based on the self-reported trajectory of symptoms to refer the affected individual to immediate treatment or to continue monitoring for up to 9 months to ensure symptoms resolved naturally. All individuals who contacted the screening team, regardless of diagnostic status, were screened repeatedly to capture delayed-onset PTSD and PTSD that gradually worsened over time (Brewin et al., 2008).

The ‘screen and treat approach’ has proven to be an effective public health response by identifying more affected individuals and facilitating early intervention and successful treatment outcomes. Of the 596 individuals who completed the initial screening following the 2005 bombings, 370 were invited for further assessment and 255 were referred for treatment. By contrast, only 14 survivors were referred by family doctors and routine medical care for treatment. Moreover, preliminary data on a subsample of these individuals showed that by the end of treatment, mean scores on the Posttraumatic Diagnostic Scale (PDS: Foa et al., 1997) fell below clinical range of symptoms, indicating that the majority of patients referred to treatment had recovered (Brewin et al., 2008).

Ideally, at-risk individuals would be identified before developing clinical symptoms. Although most survivors of disasters and potentially traumatic events follow a resilient trajectory
(Bonanno, Brewin Kaniasty, & La Greca, 2010) a significant minority can go on to develop severe psychopathology, most commonly PTSD, major depressive disorder (MDD) and anxiety disorders (Norris et al., 2002). For those that develop subsequent psychopathology and need treatment, timely identification is crucial. For example, after an airplane crashed near Amsterdam in 2009, individuals who reported unmet psychological needs two months later were more likely to be related to a chronic PTSD trajectory than those who reported no treatment (Gouweloos-Trines et al., 2019). Early interventions with evidence-based treatment, such as trauma-focused cognitive behavioral therapy, may prevent chronic PTSD symptoms and effectively treat acute PTSD clinical presentations (Roberts, Kitchiner, Kenardy, & Bisson, 2009). Unfortunately, barriers to service utilization, including stigma towards mental health treatment and avoidance of trauma-related stimuli, often prevent people with acute distress from finding and utilizing treatment services. Moreover, general practitioners do not always recognize early symptoms of PTSD (Rosenbaum, 2004).

Following individuals who are identified early on to be at-risk may also help provide treatment later on to those who underestimate their need for support. After the 2001 New York Terrorist attacks, a substantial number of adults with diagnoses of PTSD and MDD did not believe they had a problem that required treatment (Boscarino, Adams, Stuber, & Galea, 2005). Specifically, African Americans were less likely to seek services and less likely to use medications post disaster compared to Whites (Boscarino et al., 2005). These findings align with previous population studies which have demonstrated racial and ethnic disparities in mental health care such as gaps in access, discrepancies in diagnostic practices, and the accessibility of treatment (Snowden, 2003). Amongst the nontreatment seekers who had PTSD or major depression, the most common reason reported for not seeking help was that they did not have a
problem that warranted treatment (Boscarino et al., 2005). These findings suggest a more proactive approach may be necessary to identify individuals at risk and refer to appropriate treatment when needed.

1.2 Facial Behavior as a Transdiagnostic Indicator of Clinical Functioning

At present, many researchers and clinicians rely on a diagnostic approach to identify and understand clinical presentations (Dalgleish, Black, Johnston, & Bevan, 2020). Formal taxonomic systems, such as the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the International Classification of Diseases (ICD), are used to organize signs and symptoms into hypothetical distinctions associated with specific psychiatric diagnoses. This framework can be problematic for several reasons. It often leads to the assumption that a diagnosis captures a discrete underlying disorder (Dalgleish et al., 2020). Unlike physical diseases which can have identifiable and discrete causes, mental health issues more often reflect a complex interaction between biological, behavioral, psychosocial, and cultural processes. Diagnostic symptom thresholds also impose an all-or-nothing, binary structure with distinct categorical entities (Regier, Kuhl & Kupfer, 2013). However, research suggests psychiatric symptoms are multidimensional and are better conceptualized along a continuous dimension (Brown, Campbell, Lehman, Grisham, & Mancill, 2001; Kessler, Chiu, Demler, Merikangas, & Walters, 2005). Likewise, because diagnoses are heterogenous, individuals with the same diagnosis can present with widely different symptoms (Galatzer-Levy & Bryant, 2013). One study demonstrated this heterogeneity by highlighting the 636,120 possible combinations of the symptoms listed in the DSM-V for the PTSD diagnosis (Galatzer-Levy & Bryant, 2013). The lack of specification in symptom profiles has implications for treatment response. For example,
patients with different subtypes of MDD have shown diverging responses to antidepressant treatments (ADTs) (Jaracz et al., 2015)

In order to improve diagnostic accuracy, risk identification, and tailor treatment selection, there has been growing support for an alternative transdiagnostic approach to understand the onset, maintenance, and treatment of clinical disorders. This perspective cuts across traditional diagnostic boundaries and focuses instead on the underlying processes of mental distress, observable behavior, physiology and functioning to identify psychopathology and target interventions (Dalgleish et al., 2020; Insel et al., 2010). A surge of new research has begun to explore digital phenotyping, a direct, moment-to-moment, objective measurement of clinical characteristics that uses digital data sources (Huckvale, Venkatesh, & Christensen, 2019). Digital phenotyping captures identifiable behavioral signatures of diverse central nervous system (CNS) disorders that correspond with physiological and cognitive variability (Cummins et al., 2015; Dagum, 2018; Jabbi and Keysers, 2008). Some of the many digital phenotypes that may be used to indicate psychiatric outcomes include vigilance, arousal, fatigue, agitation, psychomotor retardation, flat affect, inattention, compulsive repetition, and negative affective biases (American Psychiatric Association, 2013). Facial expressions of emotion are particularly promising, because they can be coded using open-source software (Amos, Ludwiczuk, & Satyanarayanan, 2016) and used to measure clinical functioning (Cohn et al., 2009; Gaebel & Wölwer, 1992; Girard et al., 2013).

While facial expressions alone do not capture specific psychological disorders, in combination with speech and movement they may be used transdiagnostically as signs of clinical functioning (Schultebrauks et al., 2020). For example, research has demonstrated that facial expressions convey information about emotional and social behaviors. One study examined the
social context differences in self-reported emotions and facial muscle activity of major depressed and non-depressed patients (Gehricke & Shapiro, 2000). Facial activity was measured with facial electromyography (EMG) as patients were asked to imagine happy and sad situations with and without visualizing other people. Depressed patients showed diminished facial muscle activity during happy and sad stimuli, compared to non-depressed patients, but there were no group differences in self-reported emotions (Gehricke & Shapiro, 2000). The discrepancy between self-reported emotion and facial muscle activity during happy and sad imagery could reflect psychomotor retardation in major depression. Interestingly, there were no group differences in frowning among depressed and non-depressed patients across social and solitary contexts. Depressed patients appeared more likely to hide their sadness than communicate it to others. These findings suggest measuring directly observable measures of behavior and physiology like diminished facial expressions and social disengagement in patients with active depressive symptoms may help identify those at risk for poor psychosocial functioning (Gehricke & Shapiro, 2000) and improve diagnosis and treatment selection (Insel, 2014; Galatzer-Levy & Bryant, 2013).

In addition to conveying social functioning, research has found that facial activity is associated with well-known symptoms of PTSD and major depressive disorder (MDD). Compared to healthy controls, patients with PTSD and MDD have demonstrated individual differences in the expression of facial features of emotion, vocal features, and content of speech (Schultebraucks et al., 2020; Cohn et al., 2009; Gaebel & Wölwer, 1992; Gehricke & Shapiro, 2000). Among combat veterans, decreased flexibility in emotion expression has been associated with higher levels of PTSD and depressive symptoms (Rodin et al., 2017). Likewise, PTSD patients show greater anger expressions compared to trauma-exposed but healthy
participants (Blechert, Michael, Wilhelm, 2013) as well as less positive affect (Kirsch & Brunnhuber, 2007). Numerous studies have demonstrated decreased facial expressivity in patients with MDD (Davies et al., 2017; Gaebel and Wölwer, 1992; Girard et al., 2013; Renneberg et al., 2005; Sloan, Strauss, Quirk, and Sajatovic, 1997). Reduced facial expressivity and movement coded from videos of patient interviews has been found to distinguish depressed patients with and without suicide risk (Heller et al., 2005; Galatzer-Levy et al., 2021). Additionally, facial emotion intensity provides a clear indication of flat affect, pain, and fatigue (Ekman, Freisen, & Ancoli, 1980; Kohler et al., 2008; Simon, Craig, Gosselin, Belin, & Rainville, 2008).

The clear association between facial expressions and psychopathology suggest behaviors, particularly those captured early on after a potential trauma, may provide unique insight into the development of subsequent psychopathology. While self-report measures, via clinical interviews or questionnaires, are undoubtedly useful for assessing psychological functioning, they are unable to assess behavioral observations that typically occur outside of awareness, during the acute phase of reactions to the stressor (Cohn et al., 2009). Moreover, clinical interviews and/or questionnaires often require a trained clinician to administer and score and even then, are subject to variations in skill-level and training. For example, a longitudinal study that examined the accuracy of PTSD and MDD screening instruments following an airplane crash in 2009, found that while the PTSD screener was stable and accurate over time the depression measure was inconsistent (Gouweloos-Trines, 2019). Specifically, a significant proportion of participants with MDD went undetected while others without clinical symptoms were classified as at-risk (Gouweloos-Trines, 2019). The findings highlight the risk of relying solely on self-report measures to accurately identify at-risk individuals.
1.3 Manual Coding Systems

The Facial Action Coding System (FACS) (Ekman & Friesen, 1978; Ekman, Friesen & Hager, 2002) is a comprehensive, anatomically based system that measures all discernable facial movement. FACS categorizes visually distinguishable facial activity, including a few head and eye positions, into 44 unique Action Units (AUs) designated by a numeric code and measures intensity of each AU on a five-point scale. For example, AU 12 codes contracts of the zygomatic major muscle while AU 6 codes contractions of the orbicularis oculi muscle. Muscles may contract in different ways producing visibly different actions; When the medial portion of the frontalis muscle contracts, the inner corner of the eyebrow raises producing AU 1, but when the lateral portion of the same muscle group contracts, the outer brow is raised, producing AU 2. As such, while FACS is anatomically based and AUs approximate individual facial muscle movements, it is not centered on a direct 1:1 correspondence between muscle groups and AUs.

Over 7,000 combinations of AUs have been observed (Ekman, 1982). Specific combinations of FACS action units map on to prototypic expressions of emotion (i.e., joy, sadness, anger, disgust, fear, and surprise) (Ekman & Friesen, 1978). However, FACS itself is purely descriptive and does not use emotion or other inferential labels. A separate system, known as EMFACS (Friesen & Ekman, 1984), can be used to measure emotional expressions by coding only emotion-relevant facial muscle movements derived from previous theory and research, or otherwise using the FACS interpretive Dictionary (Friesen & Ekman, undated, cited by Oster, Hegley, & Nagel, 1992). Nonetheless, in daily life expressions of emotion occur relatively infrequently. More times than not, emotion is communicated through small changes in features of the face, like furrowing the brows to indicate negative affect. By using AUs, which describe the smallest visibly discriminable changes in facial displays, alone or in combination
with others, FACS can generate detailed descriptions of facial expressions, in addition to making broader distinctions between positive and negative expression (Ekman 1993; Ekman & Friesen, 1978).

FACS has been used to substantiate the physiological presence of emotion in numerous studies (Ekman, Friesen, & Ancoli, 1980; Ekman, Levenson, & Friesen, 1983; Ekman, Davidson, & Friesen, 1990; Levenson, Ekman, & Friesen, 1990; Ekman, Friesen, & O’Sullivan, 1988). Its comprehensive design allows for new patterns related to emotional and situational states to emerge. For example, research implementing FACS discovered that smiles featuring both orbicularis oculi (AU6) and zygomatic major action (AU12) (hereafter, “Duchenne smiles”) were correlated with self-reports of enjoyment and different patterns of brain activity, whereas smiles that only engaged zygomatic major action (AU12) (non-Duchenne smiles) were not (Ekman, Davidson, and Friesen, 1990). These “non-Duchenne” smiles instead appear to function as social markers or deceptive signals to hide feelings and communicate appeasement to others (Bonanno et al., 2002; Bugental, 1986; Ekman, Friesen, & O’Sullivan, 1988). One consequence, is that Duchenne smiles have been associated with health and well-being (Harker & Keltner, 2001) while non-Duchenne smiles have not.

FACS has also been used to investigate characteristics of emotional expressions as they unfold over time (Reed, Sayette, & Cohn, 2007). Not only do dynamic characteristics of facial expressions provide critical information about the intensity of an emotion and whether it is genuine, but expressions may occur in rapid sequence and convey an altogether different emotional experience than what is conveyed by individual expressions in the sequence (Ekman, 1993). For example, smile controls refer to facial actions that counteract the upward pull of the smile so as to obscure it (Keltner, 1995). One study found individuals who had a history of
MDD as well as current depressive symptomatology were more likely than individuals who were asymptomatic to express smile controls (Reed et al., 2007). The results support earlier findings that individuals with depression or dysphoria are less likely than healthy controls to respond to positive stimuli with facial expressions associated with positive emotion (Gehricke & Shapiro, 2000). Furthermore, the findings indicate that affective reactions are determined by the presence of current symptomatology, and unfold differently compared to healthy controls. Such dynamic measures of facial expression may be useful for furthering our understanding of the mechanisms underlying depression.

Of particular interest to the current study, research based upon FACS has shown that facial expressions of both positive and negative emotions predict variations in long-term psychological adjustment (Bonanno & Keltner, 1997; Gottman & Levenson, 1992; Keltner, Moffitt, & Southamer-Loeber, 1995). For example, Bonanno and Keltner (1997) provided empirical support for the role of emotional expressions in adapting to grief following conjugal loss. Specifically, the authors demonstrated that the presence of Duchenne smiles during conjugal bereavement improved long-term functioning and predicted reduced grief severity up to 2 years after the loss of the spouse, even when controlling for initial grief. In a similar study, women who displayed strong Duchenne smiles in their college yearbook photos reported less distress on a daily basis up to 30 years later, in addition to greater overall emotional and physical well-being (Harker & Keltner, 2001). Research using FACS has also shown that facial expressions can predict the onset and remission of depression, schizophrenia, and other forms of psychopathology (Ekman & Rosenberg, 1997). Patients with depressive symptoms who showed more contempt or unfelt happiness at the time of hospital admission, were found to improve less than other patients over the course of treatment (Ekman, Matsumoto, & Friesen, 2005).
Furthermore, facial expressions predicted clinical improvement above and beyond widely used measures of clinical functioning assessed during initial intakes. Together these findings support the efficacy of using FACS during acute phases of disorders to refine diagnosis, increase reliability of diagnostic classification, and predict later outcomes.

Until recently FACS has served as a “gold standard” in research for measuring emotion, pain, and behavioral measures of psychopathology in facial expressions (Ekman & Rosenberg, 2005; Cohn, Ambadar, & Ekman, 2007). However, the use of FACS also presents notable challenges. Training is time-consuming and typically involves 100 hours or more to achieve acceptable levels of interobserver reliability in coding facial displays. Beyond being labor-intensive, manual methods like FACS are time-consuming to implement. One minute of behavior can take several hours to code. These characteristics discourage large sample sizes, prolong study completion times, and implicate generalizability. In an effort to alleviate this time burden, significant research has been devoted to the development of automatic systems for facial expression analysis.

1.4 Automated Coding Systems

Significant research has been dedicated to automated facial expression recognition. Due to its ability to detect subtle and ambiguous expressions as well as capture small changes in facial affect, automated systems grounded in Ekman and Friesen’s FACS (Ekman & Friesen 1978) are often preferred (Hamm et al., 2011). Existing approaches typically involve face detection and tracking followed by feature extraction and classification, but may vary in the exact methods used at each stage (Hamm, Kohler, Gur, & Verma, 2011). In the first stage of automated FACS system, facial landmarks and regions are detected using a deformable face model, such as Active Shape Model (ASM). These models are trained using manually measured
landmark locations from still images (Hamm et al., 2011). Then geometric and texture features are extracted for AU detection. Geometric changes capture certain AUs like the Inner Brow Raiser (AU1) and Outer Brow Raiser (AU2), which move the eyebrows when activated. Meanwhile texture features provide complementary information when geometric changes are less obvious, for example when the Brow Lowerer (AU4) is activated and vertical wrinkles are produced between eyebrows (Hamm et al., 2011). With this information extracted, machine learning is then used to predict the presence or absence of AUs. Algorithms input features and produce a binary outcome of whether a certain AU is present or not.

A growing body of research has found preliminary evidence for the efficacy of deep learning algorithms and automated systems of facial expression analysis in predicting psychopathology. One recent study tested for changes in digital markers of facial expressivity, negative affect measured in facial expression, pauses in speech, and head movement activity to measure response to antidepressant treatment (ADT) with selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine uptake inhibitors (SNRIs) (Galatzer-Levy et al., 2020). Treatment response was measured bi-weekly using clinical interviews and the Montgomery-Asberg Depression Rating Scale (MADRS) and remote smartphone-based video assessments. Videos captured from the smartphone front-facing camera were extracted for analysis conducted in a Python environment using open-source tools, producing a total of 18 variables including facial markers, voice markers, and movement markers. The authors found a consistent effect of ADTs (SSRIs/SNRIs) on digital markers of motor functioning (Galatzer-Levy, 2020). Specifically, facial and vocal activity increased over 4 weeks following the start of treatment, reflecting decreases in symptom severity as assessed by clinical administration of MADRS. Across conditions, the authors also found a decrease in anger expressions. In addition to
demonstrating the scalability of digital measurement of behaviors and automated analysis, the findings indicate SSRI/SNRI treatment may reduce depression severity by increasing facial expressivity and speech production through the graded increase of serotonin (Galatzer-Levy, 2020).

Similarly, Schultebraucks and colleagues (2020) examined direct digital measurement of facial features and their intensity, head and eye movement, and language to see whether they could accurately identify clinical functioning in a population at risk for MDD and PTSD. Facial features corresponding to AUs identified by FACS were labeled from raw MP4 video files using the OpenFace package in Python (Amos et al., 2016). The extracted raw features were used to compute a Facial Expressivity Score for each emotion (Happiness, Sadness, Anger, Disgust, Surprise, Fear, Contempt), and Peak Expressivity. The authors used a deep learning neural network approach to integrate multiple sources of information (e.g., face, movement, speech, and language) and found that higher fear-expressivity and anger-expressivity were important for the classification of PTSD while higher contempt-expressivity was predictive of MDD.

In another recent study, Galatzer-Levy and colleagues (2021) demonstrated that visual and auditory digital measurements can also be used as markers of suicide severity. Using visual and auditory measures of facial activity, head movement, and speech production from open-ended clinical interviews with psychiatric patients following a suicide attempt, the authors found that these proxy markers of underlying neurobiological functioning were correlated with Beck Scale for Suicide Ideation (BSS) scores. Specifically, as facial activity decreased, suicide risk severity increased. The results support the efficacy of deep learning algorithms and their application to open-ended clinical interviews to assess visual and auditory indicators of suicide severity in a clinical population.
Automated systems of facial analyses have also been used to measure cognitive functioning in individuals following traumatic events (Schultebraucks, Yadav, & Galatzer-Levy, 2021). After trauma exposure, cognitive impairment has been found to be related to symptoms of PTSD and other trauma-related psychiatric disorders like depression (Brandes et al., 2002; Mueller-Pfeiffer et al., 2010; Hammar & Årdal, 2009). Cognitive impairment is also a risk factor for subsequent psychopathology development (Samuelson et al., 2020). Using the behavioral information of facial, movement, speech, and voice data from video recorded qualitative interviews, in addition to traditional cognitive measures, researchers confirmed the accuracy of digital phenotyping in predicting cognitive performance (Schultebraucks et al., 2021). Specifically, the results demonstrated that together the digital measurements of behavior were consistent with traditional cognitive assessments and predicted cognitive functioning among trauma survivors (Schultebraucks et al., 2021).

1.5 Comparing Automated and Manual Coding Systems

Recent advances in machine learning and automatic facial detection require validation to establish the application of these methods in assessing risk factors of psychopathology. Cohn and colleagues (2009) compared automated facial image analysis, via active appearance models (AAMs), with manual FACS annotation to detect depression and changes in symptom severity within a large clinical sample. While both measures co-varied with depression, the accuracy and true positive rates for manual FACS coding were slightly higher (89%) than for AAM (79%). Several AUs accuracy exceeded this rate and had strong predictive power for depression; For example, AU 14’s true positive and true negative rates were 87% and 89%, respectively. These preliminary findings demonstrate that nonverbal affective behavior maps onto diagnosis and can be measured automatically to contribute to clinical evaluations.
In a more recent study, Girard and colleagues (2013) examined whether patterns of facial expression in depressed patients changed with symptom reduction over the course of treatment. The authors compared facial expressions analyzed from video using both manual and automated systems. Participants’ facial expressions were video recorded during semi-structured clinical interviews to provide additional ecological validity. The authors found that the frame-by-frame agreement between manual FACS coding and automated FACS coding was good, indicating the system’s ability to identify and classify individual AU events. The reliability, or the automatic systems’ ability to identify how often each AU occurred, was very good. Both systems had high reliability in the proportion of frames each AU was present. Intraclass correlations (ICC) scores were above 0.850 for each single AU and above 0.700 for most AU combinations. Manual FACS coding indicated that patterns of facial expression varied with depressive symptom severity (Girard et al., 2013). Compared with low symptom severity participants, high symptom severity participants had significantly lower overall AU 12 activity, significantly higher overall AU 14 activity, and significantly more AU 14 activity during smiling. These findings were consistent with the automatic FACS coding. With the exception of one non-significant finding, the automated system identified all of the differences that were significant or trending (Girard et al., 2013).

While both longitudinal studies provide support for the use of automated systems of coding facial expressions, it is notable that they did not compare the use of manual and automated systems of coding in predicting symptoms over time. Instead, each study mapped changes in facial expressions and symptom severity over the course of treatment (Cohn et al., 2009; Girard et al., 2013). Likewise, both studies focused solely on signs and symptoms of depression and did not evaluate other forms of psychopathology. Hence, while they compare
manual and automated systems in their ability to detect changes in depressed clinical presentations, it remains unclear whether automated coding systems of facial expressions can effectively *predict outcomes* above and beyond manual coding systems.
Chapter 2: The current study

Each year, almost 30 million patients seek services at an emergency department following a potentially traumatic stressor event (DiMaggio et al., 2017). While many of these patients experience at least some distress or psychiatric symptoms and then recover, around 10-20% go on to develop one or more disorders including anxiety, depression, and/or PTSD (Wiseman et al., 2015; Sullivan et al., 2017; Fakhry et al., 2017). These outcomes have long-term consequences for individuals, families, and the healthcare system. Chronic posttraumatic stress disorder is often severe and treatment resistant. Empirically validated treatments have proven successful in reducing risk for PTSD (Rothbaum et al., 2014; Shalev et al., 2016; Shalev et al., 2012). However, without established, timely and reliable risk assessment methods, early prevention strategies are often not implemented (Shalev, 2019). The emergency department presents a unique opportunity to access trauma survivors immediately after the event and intervene to prevent subsequent psychopathology in high risk-population, improve patient outcomes, and reduce long-term costs to health care systems (Schultebraucks et al., 2020).

Current methods of assessing psychopathology rely primarily on the verbal report of patients, their family, or caregivers without incorporating behavioral observations that have proven to be strong indicators of psychological disorders. Using visual data from clinical interviews introduces the possibility of increasing accuracy of clinical risk. While novel measurements such as OpenFace and other automated systems of analyzing visual facial data are promising, validation is required before such measures can be interpreted with confidence. This includes comparing these new systems to traditional forms of measurement, like FACS, in longitudinal research designs.
The primary aim of the proposed study is to compare the efficacy of both FACS and automated coding systems as predictors of subsequent psychopathology. Specifically, this study will (1) compare similarities and differences in facial emotions identified by FACS and OpenFace and (2) examine the real-world consequences of each facial coding measure individually and in multivariate regressions to determine their predictive accuracy as well as the extent that the different measures might capture independent/unique aspects of psychopathology. To the best of our knowledge, this will be the first study to evaluate their relative ability to predict longitudinal outcomes following a traumatic event.
Chapter 3: Methods

3.1 Participants and Procedures

Trauma survivors who were admitted to the emergency department (ED) of a Level-1 Trauma Center after experiencing a DSM-5 criterion A trauma were enrolled into a prospective longitudinal study cohort ($n = 221$) from 2012 to 2017 at Bellevue Hospital Center, New York City, NY. To be included in the study, participants had to be between 18 and 70 years old and fluent in English, Spanish, or Mandarin. Participants who had ongoing traumatic exposure such as domestic violence, evidenced homicidal or suicidal behavior, or who were incarcerated were excluded from the study. Further exclusion criteria included present or past psychotic symptoms, open head injury, coma, or evidence of a traumatic brain injury, or not reliable access to electronic mail or telephone. The present study consisted of a subsample who participated a clinical interview one-month after their accident and/or injury ($n = 79$). All procedures were reviewed, approved, and monitored by the NYU Institutional Review Board.

Participants who sustained a psychologically traumatic accident or assault were approached to participate in the study after being admitted to the Emergency Department (ED) of Bellevue Hospital Center (BHC). Preliminary information about potential participants was gathered from the Bellevue Hospital “White Board,” the trauma team’s rounding sheet, or at trauma surgery rounds. Patients who consented to participate were asked to complete a phone screen interview approximately seven days after their accident to determine full eligibility. In order to be eligible for the study, the patient had to meet full criteria for PTSD (PTSD checklist (PCL 5) score of 35 and over)) or be randomly selected to participate in the study despite not meeting criteria for full PTSD (a random 10% of patients who consent and who do not have full PTSD were invited to participate).
Participants who satisfied eligibility for the study, or who were part of the random 10% of consented patients selected, were invited to participate in a one-month assessment approximately 30 days after their accident/injury, and four subsequent online assessments of their psychological symptoms and neurocognitive functioning three, six, nine and twelve months after their ED admission. The online assessment consisted of self-report measures of psychological functioning, including Center for Epidemiologic Studies Depression Scale (CES-D), and the PTSD Checklist (PCL).

In addition to the online assessment, participants completed a brief qualitative interview conducted under laboratory conditions at Bellevue Hospital during the one-month assessment. Patients were asked to respond however they saw fit to the following five questions within 3-minute predetermined time limit for each question:

1) Tell me about your life before the event that brought you to the hospital.

2) Tell me about the event that brought you to the hospital.

3) Tell me about your hospital experience.

4) Tell me about your life since leaving the hospital.

5) What are your expectations about life in the future?

Interviewers asked brief pre-determined follow-up questions when patients stopped responding such as ‘tell me more about that’. Interviews were audio and video recorded with a high-resolution camera mounted behind the interviewer’s shoulder to provide a direct view of the participant’s face while responding to the interview questions. Videos from the qualitative interview were manually coded using the Facial Action Coding System (FACS: Ekman, Friesen, & Hager, 2002) or with OpenFace (Baltrusaitis et al., 2016) an open-source software package that has demonstrated validity next to expert human ratings of FACS.
3.2 Measures

**Posttraumatic Stress Disorder.** The PTSD Checklist for the DSM-5 (PCL-5) is a psychometrically sound and widely used scale for assessing PTSD in both clinical and research settings (Weathers, Litz, Keane, Palmieri, Marx, & Schnurr, 2013). The PCL-5 is a 20-item self-report measure based on DSM-5 PTSD criteria that assesses the presence, frequency, and severity of PTSD symptoms experienced in the past month on a 5-point Likert scale (0 = “not at all,” 4 = “extremely”). In a sample of motor vehicle accident and sexual assault victims the internal consistency for the PCL was high (α=.94) (Blanchard, Alexander, Buckley, and Forneris, 1996).

**Depression.** The Center for Epidemiological Studies Depression Scale (CES-D) is a 20-item self-report scale that measures affective and somatic symptoms of depression in the general population (Radloff, 1977). Items receive a rating on a four-point Likert scale (0 – 3), where higher scores indicate greater depressive symptomology. The internal consistency for CES-D ranges from .85 to .90 (Radloff, 1977).

3.3 Manual Coding of Facial Expressions

Prototypical emotional displays of fear, sadness, anger, genuine happiness (Duchenne smiles), contempt, and non-Duchenne smiles were determined by the presence of particular combinations of AUs as per the FACS investigator’s guide, and previous research (Diminich & Bonanno, 2014; Ekman & Friesen, 1978; Gruber, Dutra, Eidelman, Johnson, & Harvey, 2011; Gruber, Johnson, Oveis, & Keltner, 2008). Facial expressions were coded for the presence or absence of each AU activation for each second. Separate frequency, intensity, and duration scores for each emotion and for each participant were calculated. Because some participants spoke longer than others, frequency scores were adjusted. This was done by dividing the number
of seconds the participant spoke by maximum number of seconds (180), and then multiplying by
the total number of expressions. Intensity of expression was measured on a 5-point Likert scale
of (1 = minimal intensity to 5 = extreme intensity). A mean intensity score was then calculated
by averaging the mean scores of each expression of emotion across video segments, as well as a
mean duration score, measured in number of seconds. Finally, a magnitude score for each
emotion for each participant was computed to establish a single index for analyses. The
intensity, duration, and adjusted frequency scores for each emotion were standardized via z-score
transformation and summed to compute a magnitude score.

Coders were master’s level graduate students in clinical psychology who had undergone a
minimum of 100 hours of FACS training. After studying FACS materials (See Appendix A) and
scoring practice items FACS coders completed the FACS Final Test, an exam designed by Paul
Ekman that provides a standard of FACS compliance by ensuring that coders are not only
reliable with one another, but accurately coding AUs. The exam includes 34 short videotaped
excerpts from conversations and behaviors in the recording reflect actions typical of actual
research recordings. For example, people talk and move their heads, making identification of
AUs difficult. Furthermore, the test is recorded on a standard VHS video tape, which provides
additional challenges to AU identification. After completing the test, scores are analyzed and a
listing of scores with the correct answers and norms on how other coders have performed on the
final test are returned. Meeting a minimum level of agreement is required to pass the exam.
Upon passing the FACS Final Test, reliability was established across the coders involved in the
present study. They were blind to the hypotheses and purposes of the study, and footage was
coded without sound. To establish reliability, coders independently coded the same 4 segments
for 5 subjects (20 segments). Percentage of agreement was greater than 80% on all AUs. The remaining participant videos were then coded individually.

3.4 Automated Coding of Facial Expressions

OpenFace is an open-source facial behavior analysis tool kit used by machine learning and computer vision researchers, that allows for facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation (Baltrusaitis et al., 2016). The OpenFace software package (e.g., OpenFace, OpenCV) includes computer vision algorithms for automatic facial identification and feature extraction from videos that contain a single face. The algorithms are trained on multiple datasets involving participant images and videos and learn thousands of examples of AUs. In this study, the facial action unit recognition feature was used to analyze participant videos. First, all videos were segmented into individual video frames at 30 frames per second to be compatible with OpenFace software. Each frame was subsequently segmented into three matrices consisting of red, blue, green spectrum pixels. Values within these matrices indicate light to dark pixel values based on a corresponding color spectrum to produce frame by frame facial features. Facial features are variables that capture changes in color, brightness, and facial muscle activation. The frames and features were then analyzed by OpenFace prediction models, which produce AU presence and intensity predictions for each frame in a video. The OpenFace output generated was used to compute facial emotions (e.g., fear, contempt, sadness, anger, disgust, surprise, happiness) and intensity scores based on the AUs and level of activation present in each frame. Because the manually coded data was computed in seconds, the OpenFace data was then converted from frames back into seconds (30 frames per 1 second) for comparison purposes. Lastly, frequency, duration, and intensity scores for each emotion were computed. Frequency scores are a sum of all discrete instances of an expression onset across the
total number of seconds spoke (180 seconds). Duration reflects the average of all the expressions measured in seconds across the total time spoken. The intensity scores were computed by averaging the intensity of any emotional expression lasting up to 12 seconds. On the rare occasion that an emotional expression exceeded 12 seconds, a mean intensity was computed by averaging the first 13 seconds of the expression which was assumed to be a strong proxy measurement for the mean of every second in the expression. Frequency, duration, and intensity scores were standardized via z-score transformation and summed to compute a magnitude score.
Chapter 4: Results

4.1 Data Analysis Overview

The purpose of this study is to evaluate differences between manual and automated coding systems in predicting psychological outcomes following a traumatic event. To do this, we first explored the sensitivity of each coding method by looking at the frequency of emotions detected by each system. Next, we examined the correlations between the magnitude scores of FACS and OpenFace coded emotions (Fear, Anger, Sadness, Duchenne, non-Duchenne, Contempt) with baseline, 6-month, and 12-month depression and PTSD scores. Finally, we explored the ability of each facial emotion measure to predict psychopathology using individual linear regressions. Specifically, for FACS and OpenFace scores independently, we regressed emotion magnitude scores on 6-month and 12-month symptoms of depression and PTSD, while controlling for baseline symptoms on these same outcome measures. This allowed us to see if a coding system captured variability in psychological functioning after the traumatic event.

4.2 Descriptive Statistics

OpenFace identified Duchenne, Sadness, Anger, Fear, and Contempt in all 79 of the participants, AU 12 without 6 in 78 participants, and Disgust in 75 participants. FACS identified Duchenne in 72 participants, AU 12 without 6 in 70 participants, Sadness in 22 participants, Disgust in 5 participants, Fear in 1 participant, Contempt in 30 participants, and Anger in 0 participants (see Table 1).
<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>OF Duchenne</td>
<td>79</td>
<td>120.88</td>
<td>74.57</td>
</tr>
<tr>
<td>OF AU 12 w/o 6</td>
<td>78</td>
<td>81.21</td>
<td>61.62</td>
</tr>
<tr>
<td>OF Sadness</td>
<td>79</td>
<td>223.76</td>
<td>17.07</td>
</tr>
<tr>
<td>OF Disgust</td>
<td>75</td>
<td>87.80</td>
<td>77.36</td>
</tr>
<tr>
<td>OF Anger</td>
<td>79</td>
<td>219.32</td>
<td>27.54</td>
</tr>
<tr>
<td>OF Fear</td>
<td>79</td>
<td>221.77</td>
<td>21.66</td>
</tr>
<tr>
<td>OF Contempt</td>
<td>79</td>
<td>202.00</td>
<td>44.96</td>
</tr>
<tr>
<td>FACS Duchenne</td>
<td>72</td>
<td>7.15</td>
<td>8.43</td>
</tr>
<tr>
<td>FACS AU 12 w/o 6</td>
<td>70</td>
<td>6.89</td>
<td>5.97</td>
</tr>
<tr>
<td>FACS Sadness</td>
<td>22</td>
<td>5.52</td>
<td>1.03</td>
</tr>
<tr>
<td>FACS Disgust</td>
<td>5</td>
<td>2.39</td>
<td>0.12</td>
</tr>
<tr>
<td>FACS Anger</td>
<td>0</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>FACS Fear</td>
<td>1</td>
<td>4.83</td>
<td>--</td>
</tr>
<tr>
<td>FACS Contempt</td>
<td>30</td>
<td>5.10</td>
<td>0.94</td>
</tr>
<tr>
<td>DEP 1m</td>
<td>67</td>
<td>21.24</td>
<td>11.73</td>
</tr>
<tr>
<td>DEP 6m</td>
<td>41</td>
<td>18.90</td>
<td>11.48</td>
</tr>
<tr>
<td>DEP 12m</td>
<td>29</td>
<td>16.79</td>
<td>9.86</td>
</tr>
<tr>
<td>PTSD 1m</td>
<td>74</td>
<td>27.51</td>
<td>16.55</td>
</tr>
<tr>
<td>PTSD 6m</td>
<td>45</td>
<td>21.24</td>
<td>14.92</td>
</tr>
<tr>
<td>PTSD 12m</td>
<td>26</td>
<td>18.54</td>
<td>12.34</td>
</tr>
</tbody>
</table>

OF = Open Face, FACS = Facial Action Coding System, DEP = Depression (measured using the Center for Epidemiologic Studies- Depression (CES-D) scale), PTSD = Post-traumatic stress disorder (measured using the Post-traumatic stress disorder checklist (PCL-5)

-- Indicates missing data.
While AUs were identified by FACS in all videos, it was often not enough to meet criteria for an emotion to be coded. For instance, the emotion Fear is composed of AUs 1, 2, 4, 5, and 25, 26, or 27; some of the AUs were identified by FACS but not all, or not all within the same window of an expression. In other cases, FACS failed to capture AUs required for any emotion. For example, the emotion Disgust is indicated when AU 9 or AU 10 is present, but these action units were often missed by FACS and therefore Disgust was rarely detected. Notably, AUs involved in speech (AUs 10, 14, 17, 23, 24, 25 and 26) production appeared to be captured less frequently by FACS (see Figure 1).

4.3 Correlations Between Facial Coding Methods

We explored the correlations between magnitude scores of FACS and OpenFace coded emotions (Fear, Anger, Sadness, Duchenne, and Contempt) as well as 1-month, 6-month, and 12-month Depression and PTSD scores (see Table 2). Correlations were limited to the emotions identified by each coding system. Consequently, the FACS magnitude scores for Anger and Fear were omitted due to an insufficient number of responses. We hypothesized that there would be a relationship between the OpenFace and FACS emotion magnitude scores. Pearson’s bivariate correlation coefficient showed a negative linear relationship between the FACS magnitude score for Duchenne smiles and OpenFace magnitude score for Sadness (r=-.42, p<.01) and OpenFace magnitude score for Fear (r=-.29, p<.01). No other significant correlations were observed between FACS and OpenFace magnitude scores. There was also a negative linear relationship between OpenFace magnitude score for Anger and PTSD at 12 months (r=-.41, p<.05).

4.4 Predicting Long-term Adjustment with Emotion Codes

We examined whether facial expressions coded 1-month following a traumatic stressor predicted psychological outcomes at 6 months and 12 months later. We did this, in separate
**Figure 1.** Total Number of AUs Identified by FACS Coders Across All Participant Videos

<table>
<thead>
<tr>
<th>Emotion</th>
<th>AU 1</th>
<th>AU 4</th>
<th>AU 6</th>
<th>AU 11</th>
<th>AU 15</th>
<th>AU 25</th>
<th>AU 26</th>
<th>AU 27</th>
<th>AU 9</th>
<th>AU 10</th>
<th>AU 23</th>
<th>AU 24</th>
<th>AU 2</th>
<th>AU 10</th>
<th>AU 14</th>
<th>AU 17</th>
<th>AU 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sad</td>
<td>287</td>
<td>216</td>
<td>191</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>292</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>287</td>
<td>216</td>
<td>191</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>139</td>
<td>142</td>
<td>126</td>
<td></td>
<td></td>
<td></td>
<td>43</td>
<td>216</td>
<td>196</td>
<td></td>
</tr>
<tr>
<td>Disgust</td>
<td>216</td>
<td>139</td>
<td>142</td>
<td>126</td>
<td>7</td>
<td>29</td>
<td>43</td>
<td></td>
<td></td>
<td>138</td>
<td>196</td>
<td></td>
<td></td>
<td></td>
<td>112</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>216</td>
<td>138</td>
<td>196</td>
<td>112</td>
<td>21</td>
<td>23</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>44</td>
<td>89</td>
<td>138</td>
<td>142</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contempt</td>
<td>287</td>
<td>216</td>
<td>191</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* AUs = Action Units; Yellow denotes AUs involved in Speech; AUs are clustered by combination required to produce the Emotion (e.g., Sadness is composed of AU 1, 4, 6, 11, and/or 15)
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OF Duchenne</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>OF AU 12 w/o 6</td>
<td>.71***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>OF Sadness</td>
<td>-.21</td>
<td>-.10</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>OF Anger</td>
<td>.01</td>
<td>.02</td>
<td>.57***</td>
<td>.11</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>OF Fear</td>
<td>- -0.04</td>
<td>.02</td>
<td>-.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>OF Contempt</td>
<td>.13</td>
<td>.20</td>
<td>.42***</td>
<td>-.02</td>
<td>.30**</td>
<td>.54**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>FACS Duchenne</td>
<td>.10</td>
<td>-.08</td>
<td>-.42***</td>
<td>.06</td>
<td>-.22</td>
<td>-.29</td>
<td>-.19</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>FACS AU 12 w/o 6</td>
<td>-.13</td>
<td>.01</td>
<td>.04</td>
<td>-.23</td>
<td>.02</td>
<td>-.08</td>
<td>.51***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>FACS Sadness</td>
<td>.30</td>
<td>.32</td>
<td>-.27</td>
<td>.11</td>
<td>-.31</td>
<td>.01</td>
<td>.11</td>
<td>.41</td>
<td>-.40</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>FACS Anger</td>
<td>.57</td>
<td>.00</td>
<td>.73</td>
<td>.78</td>
<td>.31</td>
<td>.74</td>
<td>.45</td>
<td>-.34</td>
<td>.28</td>
<td>-.33</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>FACS Fear</td>
<td>-.09</td>
<td>-.03</td>
<td>.11</td>
<td>.01</td>
<td>-.02</td>
<td>.28</td>
<td>.06</td>
<td>.01</td>
<td>-.37**</td>
<td>.27</td>
<td>--</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>FACS Contempt</td>
<td>-.04</td>
<td>.05</td>
<td>.15</td>
<td>.06</td>
<td>-.02</td>
<td>.06</td>
<td>.06</td>
<td>-.07</td>
<td>-.04</td>
<td>-.42</td>
<td>--</td>
<td>-.35</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>DEP 1m</td>
<td>-.12</td>
<td>-.10</td>
<td>-.18</td>
<td>.00</td>
<td>-.27</td>
<td>-.13</td>
<td>.21</td>
<td>.27</td>
<td>-.02</td>
<td>-.40</td>
<td>--</td>
<td>-.09</td>
<td>.53**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>DEP 6m</td>
<td>-.18</td>
<td>-.22</td>
<td>-.17</td>
<td>.02</td>
<td>-.15</td>
<td>-.08</td>
<td>.05</td>
<td>-.04</td>
<td>-.10</td>
<td>.18</td>
<td>--</td>
<td>-.40</td>
<td>.44**</td>
<td>.64**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>PTSD 1m</td>
<td>-.07</td>
<td>-.07</td>
<td>.02</td>
<td>-.07</td>
<td>-.02</td>
<td>.19</td>
<td>-.06</td>
<td>-.15</td>
<td>-.24</td>
<td>-.16</td>
<td>.83**</td>
<td>.40*</td>
<td>.25</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>PTSD 6m</td>
<td>-.22</td>
<td>-.22</td>
<td>-.12</td>
<td>.16</td>
<td>-.27</td>
<td>-.07</td>
<td>.19</td>
<td>.23</td>
<td>-.16</td>
<td>-.39</td>
<td>--</td>
<td>-.20</td>
<td>.71**</td>
<td>.78**</td>
<td>.52**</td>
<td>.64***</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>PTSD 12m</td>
<td>-.09</td>
<td>-.05</td>
<td>.07</td>
<td>-.25</td>
<td>-.41*</td>
<td>.07</td>
<td>-.05</td>
<td>-.16</td>
<td>.00</td>
<td>.59</td>
<td>--</td>
<td>-.28</td>
<td>.68***</td>
<td>.57*</td>
<td>.73***</td>
<td>.56**</td>
<td>.72***</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**OF** = OpenFace, **FACS** = Facial Action Coding System, **DEP** = Depression (measured using the Center for Epidemiologic Studies-Depression (CES-D) scale), **PTSD** = Post-traumatic stress disorder (measured using the Post-traumatic stress disorder checklist (PCL-5))

† Variables for FACS Anger (n=1) and FACS Fear (n=0) were excluded due to insufficient number of responses.

-- Indicates missing data.

* $p<.05$, **$p<.01$, ***$p<.001$. 

30
analyses, by regressing depression and PTSD scores at 6 months and 12 months against emotion magnitude scores, while controlling for 1-month outcome scores on the same measure. Analyses were limited by the low number of observations in outcome measures (e.g., 26 participants completed PTSD measures at 12 months; See Table 1.) which restricted the number of predictors allowed in a model (Hair, 2011). Owing to this limitation, we focused only on key variables most conceptually relevant to Depression (i.e., Sadness) and PTSD (i.e., Fear or Anger). Because FACS failed to detect these critical emotions (i.e., Fear, Anger, or Sadness), we were unable to compare the two coding systems and limited our prediction models to OpenFace emotion magnitude scores only.

For 6-month depression symptoms, we entered 1-month depression symptoms on the first step, and the OF magnitude score for Sadness on the second step. Entering Sadness on the second step increased the model’s R Square to a marginally significant degree ($F_{change}, p = .07$) and accounted for an additional 6.7% variance in 6-month depression symptoms. The overall model for 6-month depression was significant, $F(2,34) = 9.01, p < .001, R^2 = .31$. In the final model, Sadness marginally significantly predicted reduced 6-month depression ($\beta = -.26, p = .07$) (See Table 3).

| Table 3. Hierarchical Linear Regression of OpenFace Sadness on Depression at 6-months |
|---------------------------------|-----|-------|-------|-------|-------|-------|
| Variable                        | $\beta$ | $t$   | $R^2$ | $\Delta R^2$ | $F (df)$ | $\Delta F (df)$ |
| Step 1                          |       |       |       |       |       |       |
| Depression 1m                   | .53   | 3.68*** | .26   | .28   | 13.55 (1, 35)*** | 13.55 (1, 35)*** |
| Step 2                          |       |       |       |       |       |       |
| Depression 1m                   | .57   | 4.05*** | .31   | .07   | 9.01 (2, 34)*** | 3.50 (1, 34) |
| OF Sadness                      | -.26  | -1.87  |       |       |       |       |

OF = OpenFace Magnitude scores, Depression measured using the Center for Epidemiologic Studies-Depression (CES-D) scale;

* $p<.05$, ** $p<.01$, *** $p<.001$. 
We ran the same analyses using OF sadness against 12-month depression symptoms. We entered 1-month depression symptoms on the first step and the OF magnitude score for sadness on the second step. Entering Sadness on the second step resulted in a non-significant increase in the model’s R Square (F change, $p = .19$) and accounted for an additional 6% variance in 12-month depression symptoms. The overall model for 12-month depression was significant, $F(2,23) = 3.97, p < .05, R^2 = .26$. However, Sadness failed to reach significance ($beta = -.25, p = .19$) (See Table 4).

Table 4. Hierarchical Linear Regression of OpenFace Sadness on Depression at 12-months

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$R^2$</th>
<th>$\Delta R^2$</th>
<th>$F$ (df)</th>
<th>$\Delta F$ (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression 1m</td>
<td>.44</td>
<td>2.43*</td>
<td>.16</td>
<td>.20</td>
<td>5.91 (1, 24)$^*$</td>
<td>5.91 (1, 24)$^*$</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression 1m</td>
<td>.48</td>
<td>2.65*</td>
<td>.19</td>
<td>.06</td>
<td>3.97 (2, 23)$^*$</td>
<td>1.83 (1, 23)</td>
</tr>
<tr>
<td>OF Sadness</td>
<td>-.25</td>
<td>-1.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OF = OpenFace Magnitude scores, Depression measured using the Center for Epidemiologic Studies-Depression (CES-D) scale;

*p < .05, **p < .01, ***p < .001.

We repeated these analyses using OpenFace magnitude scores for anger and fear against 6 and 12-month PTSD. First, we examined anger against 6-month PTSD, by entering 1-month PTSD symptoms on the first step, and the OF magnitude score for Anger on the second step. Entering Anger on the second step increased the model’s R Square (F change, $p = .05$) and accounted for an additional 5.3% of the variance in 6-month PTSD symptoms. Overall, the variables statistically significantly predicted 6-month PTSD, $F(2,40) = 17.39, p < .001, R^2 = .47$. Anger predicted significantly less 6-month PTSD ($beta = -.23, p = .05$) (See Table 5).
We examined whether these findings remained at 12-months, and found that Anger increased the model’s R Square (F change, $p = .03$) and accounted for an additional 13.9% variance in 12-month PTSD symptoms. The variables significantly predicted 12-month PTSD, $F(2,22) = 9.01$, $p < .001$, $R^2 = .45$. Anger continued to predict significantly less PTSD at 12 months ($\beta = -.37$, $p = .03$) (See Table 6).

### Table 5. Hierarchical Linear Regression of OpenFace Anger on Post-traumatic stress disorder at 6-months

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$R^2$</th>
<th>$\Delta R^2$</th>
<th>$F$ (df)</th>
<th>$\Delta F$ (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD 1m</td>
<td>.64</td>
<td>5.36</td>
<td>.40</td>
<td>.41</td>
<td>28.78 (1, 41)$^*$</td>
<td>28.78 (1, 41)$^*$</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD 1m</td>
<td>.63</td>
<td>5.41</td>
<td>.44</td>
<td>.05</td>
<td>17.39 (2, 40)$^{***}$</td>
<td>3.94 (1, 40)$^*$</td>
</tr>
<tr>
<td>OF Anger</td>
<td>-.23</td>
<td>-1.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OF = OpenFace Magnitude scores, PTSD = Post-traumatic stress disorder (measured using the Post-traumatic stress disorder checklist (PCL-5)

* $p<.05$, ** $p<.01$, *** $p<.001$.

We examined whether these findings remained at 12-months, and found that Anger increased the model’s R Square (F change, $p = .03$) and accounted for an additional 13.9% variance in 12-month PTSD symptoms. The variables significantly predicted 12-month PTSD, $F(2,22) = 9.01$, $p < .001$, $R^2 = .45$. Anger continued to predict significantly less PTSD at 12 months ($\beta = -.37$, $p = .03$) (See Table 6).

### Table 6. Hierarchical Linear Regression of OpenFace Anger on Post-traumatic stress disorder at 12-months

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$R^2$</th>
<th>$\Delta R^2$</th>
<th>$F$ (df)</th>
<th>$\Delta F$ (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD 1m</td>
<td>.56</td>
<td>3.23</td>
<td>.28</td>
<td>.31</td>
<td>10.41 (1, 23)$^{**}$</td>
<td>10.41 (1, 23)$^*$</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD 1m</td>
<td>.53</td>
<td>3.37</td>
<td>.40</td>
<td>.14</td>
<td>9.01 (2, 22)$^{***}$</td>
<td>5.56 (1, 22)$^*$</td>
</tr>
<tr>
<td>OF Anger</td>
<td>-.37</td>
<td>-2.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OF = OpenFace Magnitude scores, PTSD = Post-traumatic stress disorder (measured using the Post-traumatic stress disorder checklist (PCL-5)

* $p<.05$, ** $p<.01$, *** $p<.001$.

Lastly, we regressed PTSD scores at 6 months and 12 months against the OpenFace magnitude score for Fear, while controlling for 1-month PTSD symptoms. Entering Fear on the
second step against 6-month PTSD scores increased the model’s R square (F Change, \( p = .47 \)) and accounted for an additional 1% (R Square change) variance in 6-month PTSD symptoms. Overall, the variables statistically significantly predicted 6-month PTSD, \( F(2,40) = 14.44, p < .001, R^2 = .42 \). Fear was not significantly predictive of 6-month PTSD (\( beta = -.08, p = .50 \)) (See Table 7).

Table 7. Hierarchical Linear Regression of OpenFace Fear on Post-traumatic stress disorder at 6-months

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \beta )</th>
<th>( t )</th>
<th>( R^2 )</th>
<th>( \Delta R^2 )</th>
<th>( F (df) )</th>
<th>( \Delta F (df) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td>( .41 )</td>
<td>( .41 )</td>
<td>28.78 (1, 41)**</td>
<td>28.78 (1, 41)**</td>
</tr>
<tr>
<td>PTSD 1m</td>
<td>.64</td>
<td>5.36***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td>( .42 )</td>
<td>( .01 )</td>
<td>14.44 (2, 40)**</td>
<td>0.47 (1, 40)</td>
</tr>
<tr>
<td>PTSD 1m</td>
<td>.64</td>
<td>5.34***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OF Fear</td>
<td>-.08</td>
<td>-.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OF = OpenFace Magnitude scores, PTSD = post-traumatic stress disorder (measured using the post-traumatic stress disorder checklist (PCL-5)

*\( p < .05 \), **\( p < .01 \), ***\( p < .001 \).

Likewise, we regressed 12-month PTSD scores against OpenFace Fear. Entering Fear on the second step against 12-month PTSD scores increased the model’s R square (Sig F Change = .12) and accounted for an additional .4% (R Square change) variance in 12-month PTSD symptoms. Overall, the variables statistically significantly predicted 12-month PTSD, \( F(2,22) = 5.07, p = .01, R^2 = .32 \). However, in the final model, Fear again was not significantly predictive of 12-month PTSD (\( beta = .06, p = .73 \)).
Table 8. Hierarchical Linear Regression of OpenFace Fear on Post-traumatic stress disorder at 12-months

<table>
<thead>
<tr>
<th>Variable</th>
<th>β</th>
<th>t</th>
<th>R²</th>
<th>ΔR²</th>
<th>F (df)</th>
<th>ΔF (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD 1m</td>
<td>.56</td>
<td>3.23**</td>
<td>.31</td>
<td>.31</td>
<td>10.41 (1, 23)**</td>
<td>10.41 (1, 23)**</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD 1m</td>
<td>.57</td>
<td>3.16**</td>
<td>.32</td>
<td>.004</td>
<td>5.07 (2, 22)*</td>
<td>0.12 (1, 22)</td>
</tr>
<tr>
<td>OF Fear</td>
<td>.06</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OF = OpenFace Magnitude scores, PTSD = Post-traumatic stress disorder (measured using the Post-traumatic stress disorder checklist (PCL-5)

*p<.05, **p<.01, ***p<.001.
Chapter 5: Discussion

5.1 Summary and Explanation of Key Findings

Human faces provide a rich source of behavioral data. Following acute, potentially traumatic events, manual and automated coding systems have been used to identify individuals at risk for developing psychopathology. In the present study, OpenFace and FACS were compared as predictors of long-term functioning using behavioral data from clinical interviews collected one-month after a potentially traumatic event that brought participants into the emergency department of a Level-1 Trauma Center in New York City. The overarching aim of the study was to establish a method that could identify high-risk individuals and, ultimately, contribute to the facilitation of early, tailored treatment interventions to reduce psychopathology in high-risk populations. To do this, we set out to compare similarities and differences in facial emotions identified by FACS and OpenFace and determine their predictive accuracy in capturing Depression and PTSD 6-months and 12-months later. The current study highlights unique differences between the coding methods, and yields potentially important implications for future research and risk-identification.

The findings from the current study indicate OpenFace is a more sensitive measure of facial behavior than FACS. For all participants, OpenFace recognized each emotion (Duchenne, non-Duchenne, Sadness, Disgust, Anger, Fear and Contempt) at greater frequencies than FACS (see Table 1). FACS captured Duchenne and non-Duchenne smiles consistently across participants (72 and 70 respectively) but at much lower rates (see Table 1). Furthermore, FACS rarely identified Sadness (N=22) and Contempt (N=30) and never recognized Disgust (N=5), Anger (N=0), or Fear (N=1). Because FACS appeared to have failed to capture key emotion variables (e.g., Sadness, Fear, and/or Anger) most related to the outcome measures, we were
unable to explore its predictive accuracy further. Nonetheless, we examined the predictive accuracy and extent to which OpenFace emotions predicted psychopathology 6- and 12-months after a potentially traumatic event. Somewhat counter-intuitively, our findings suggest expressing sadness and anger one month after a potentially traumatic event may be adaptive and have a salubrious effect on future psychological outcomes. Specifically, conveying sadness predicted marginally less depression 6 months after a potentially traumatic event, and expressing anger predicted less PTSD at 6 and 12 months.

There are several possible explanations for the finding that expressing anger predicted fewer symptoms of PTSD 6 and 12 months following a potentially traumatic event. Research suggests anger is characteristic of most anxiety disorders, but particularly pronounced in PTSD (Olatunji, Ciesielski, & Tolin, 2010). Although it is unclear whether anger is a cause or consequence of PTSD, one longitudinal study among trauma-exposed individuals found PTSD symptoms predicted subsequent anger levels but anger did not predict ensuing PTSD symptoms (Orth, Cahill, Foa, & Maercker, 2008). Likewise, data suggests anger may emerge as a result of PTSD (Olatunji et al., 2010). These findings suggest the timing of when anger is expressed matters; Anger expressed early on may be adaptive, but later may reflect PTSD symptom severity. For instance, research examining the functional role of anger during a simulated-peer rejection paradigm amongst first-year college students found that expressing anger was adaptive in an appropriate situational context (e.g., discussing a negative memory) and predicted lower sympathetic arousal in that context (Coifman et al., 2016). Lower sympathetic arousal has been associated with increases in working memory and greater parasympathetic activity (Giuliano, Gatzke-Kopp, Roos, & Skowron, 2017), which are indicators of psychological health and regulatory resources (Beauchaine & Thayer, 2015). Likewise, prior research has found that
individuals experiencing anger make relatively optimistic judgements and choices (Lerner & Keltner, 2001), compared to individuals experiencing fear which motivates pessimistic judgements and choices. This in turn could impact a myriad of behavioral health decisions that psychological outcomes.

This evidence had resonance with the findings from the current study which suggested there is benefit to expressing negative emotion while discussing one’s life before, during, and after a potentially traumatic event. Other research has also suggested that expressing negative emotions (e.g., anger, fear, sadness, disgust) can be functionally adaptive under certain conditions. Specifically, if negative emotions are contextually sensitive and adaptive with respect to a particular outcome that they can influence, they may have a beneficial impact on psychological health and adjustment (Coifman, Flynn, & Pinto, 2016). For instance, research has demonstrated that sadness expressed during a loss-evoking context can predict greater treatment adherence, through enhanced problem solving, deliberation, and reflection (Coifman et al., 2016). This is especially true for emotions expressed through facial expressions as opposed to those self-reported (Coifman et al., 2016). In fact, most research that supports the association between negative emotions and poor adjustment in community samples relies on self-reported assessments which are subject to response bias, demand characteristics, and influenced by individual differences in patterns of emotion conceptualization (Kashdan et al., 2015) and attention to emotional experience (Kashdan et al., 2006). However, when emotions are elicited and measured in real-time, data suggests context-sensitive negative emotions can predict adaptive outcomes.

In an effort to understand why FACS did not identify facially expressed emotions at much lower frequency than OpenFace and identified some key emotions (disgust, anger, fear) at
very low frequencies, we re-examined the raw data. First, we confirmed that OpenFace was coding facial behavior, by reviewing the recorded clinical interviews and establishing that participants were indeed producing facial movement when emotions were identified. Then we re-examined the raw data to examine why FACS did not code these emotions. Across all clinical interviews, the FACS coding recognized AUs when facial activity was present. This suggests that FACS coders were appropriately coding, at least some, AUs when there was movement in participants’ faces. Nonetheless, emotions were still not identified by FACS at the same frequency as they were by OpenFace.

One potential explanation, is that FACS did not recognize, and therefore underscored, AUs in the clinical interviews. When we examined the frequency of AUs coded across all conditions, we discovered that certain AUs were coded much less frequently than others. For example, AU 11 was coded in 21 different conditions across participants whereas AU 1 was coded in 287. This may reflect how frequently the AU occurs (AU 1 may be expressed more than AU 11), or it could indicate that AU 11 was underrecognized and therefore underscored by FACS. If a lot of facial activity is present at a single moment, in rapid sequence, it may be difficult to discern which AUs are being expressed. Specific combinations of AUs are required for an emotion to be considered present. For instance, Sadness is indicated when a combination of AUs 1, 4, 6, 11, and 15 is coded. In order for the threshold for Sadness to be met, the combination of AUs must be present within a 5-second window of time. It is possible that FACS identified some but not all of the AUs required for the emotion to be present. This would explain why even though FACS appeared to code AUs, emotions were not identified as frequently as they were by OpenFace.
Another factor that could contribute to the under-identification of AUs by FACS is the act of speaking. Speech production recruits’ areas of the face, including the lips, cheeks, and chin, that produce changes in the face involving the same muscles that produce AUs. To code facial expressions, human FACS coders must distinguish whether the facial activity is a result of speech or if it is a functionally different movement that should be coded. Some facial movements are clearly unnecessary for the production of speech, such as AUs that occur in the forehead or eyes. However, other AUs that occur around the mouth can be masked by movements that produce speech. One way to distinguish an AU from a speech movement is by the intensity at which it occurs. Most facial activity involved in speech occurs at a low intensity, therefore anything accentuated beyond what is necessary for speech may be codable. Another way to distinguish movements is by its timing; an AU that begins before speech or that continues to be visible in the face for longer than is needed for speech should be coded. Similarly, AUs should be coded when there is a pause in speech. A goal of FACS scoring is to disregard movements that function only to produce speech, while coding movements that may co-occur with speech.

In our study, AUs that involve the lips (AUs 10, 14, 16, 17, 18, 20, 23, 24, 25, 26 and 28) and are most problematic to score during speech, appeared to be coded less frequently (See Figure 1.) As participants were speaking throughout the clinical interview, it is plausible that speech interfered with coding, resulting in fewer identified emotions by FACS. For instance, Anger is composed of a combination of AUs 4, 5, 6, and 23 or 4, 5, 7, and 24. Both AU 23 and 24 are notably difficult to identify during speech. Hence, even if the other AUs were present and recognized by FACS, Anger would not be coded as present if AU 23 or 24 were obscured by
speech. This could partly explain why, in the current study FACS failed to capture as many emotions as OpenFace.

Training for FACS coding requires a rigorous curriculum based on a manualized coding system (See Appendix A). After studying FACS materials and scoring practice items, all FACS coders must satisfactorily complete the FACS Final Test, which provides a standard of FACS compliance. In theory, this ensures that trainees are not only reliable with one another, but accurately coding AUs. The FACS Final Test includes 34 short, video excerpts from conversations. The behaviors in the videos are diverse, reflecting actions typical of actual research recordings. For example, people talk and move their heads, making identification of AUs difficult. Furthermore, the test is recorded on a standard VHS video tape, which provides additional challenges to AU identification. After completing the test, coded FACS scores are analyzed and a listing of scores with the correct answers and norms on how other coders have performed on the final test are returned. Meeting a minimum level of agreement is required to pass the exam. Although, FACS coders learn to recognize AUs from still images and videos of human faces, there are clear limitations to the manual coding system. In addition to requiring extensive training, it is also subjective and thus prone to bias (Hamm et al., 2011). And while FACS coders are trained to distinguish AUs from speech production, this possible confound still remains an issue (Ekman & Friesen, Investigator’s Guide).

By contrast, OpenFace algorithms developed for real-time AU detection are created using datasets that consist of thousands of high-resolution AU labelled images (Baltrušaitis, Mahmoud, & Robinson, 2015). These datasets include videos of participants who exhibit spontaneous facial expressions in response to emotion-elicitation tasks. The sensitivity of OpenFace AU recognition systems depend largely on the dataset used to train the algorithm (Baltrušaitis et al., 2015). For
instance, training datasets must be balanced, and include a range of natural expressions, individual differences and pose variations, to refine AU detection. Research has found that using combined training datasets to develop OpenFace algorithms can further enhance AU detection (Baltrušaitis et al., 2015). While developing OpenFace algorithms still face challenges, such as estimating neutral facial expressions when the person is not known, consistent with the current findings, it appears to be superior to hand coding methods because machine learning algorithms repeatedly learn AUs through exposure to thousands of AU examples, resulting in a more precise coding system with predictive validity.

5.2 Implications

The current findings have important implications for the measurement of facial behavior and screening for risk identification. Leveraging the current findings to improve coding systems would allow for more effective risk identification. This in turn, could help providers tailor treatment to individuals, reduce costs of care, and manage resources more efficiently. It is possible that people who are identified as at-risk and receive an additional assessment will never develop clinically significant symptoms. It remains undetermined whether there are potential negative effects to screening healthy people, including pathologizing normal psychological reactions and disrupting adaptive coping processes (Gouwelloos-Trines, 2019). However, not screening may result in untreated PTSD which may lead to chronic PTSD, low quality of life, substantial health care costs, and co-morbid psychological health issues including substance abuse (Bichescu et al., 2005; Priebe et al., 2009).

5.3 Limitations

Our study has several limitations worth noting. First, although the total study cohort included 221 participants, only 79 participated in the clinical interview 1-month after their
accident and/or injury, and an even smaller subsample had outcome data at 6 and 12 months. This not only impacts the generalizability of our findings, but restricted the analyses we were able to do. Specifically, because the outcome data was limited to a small group of participants, we elected to examine one relevant predictor at a time. This prohibited us from comparing the predictive accuracy of different emotions. The small sample size may have also inadvertently reduced the power of the analyses resulting in false negative findings. For instance, in our small sample OpenFace Fear did not predict PTSD at 6 or 12 months. It is possible that with a larger sample, we would have observed different results. Relatedly, although our study suggested that FACS failed to identify most emotions, owing to the overall small sample size of our study, we were unable to directly compare the two forms of measurement. Given the clinical context of the study, it seems unlikely that there would be so few facial expressions, as indicated by FACS. However, without an absolute gold standard for measuring when a facial expression is present or not, we cannot definitively say whether OpenFace is a more sensitive measure and has greater predictive accuracy than FACS. Finally, the outcomes measures we used, the Center for Epidemiological Studies-Depression (CES-D) scale and Post-traumatic stress disorder checklist (PCL-5), were self-report measures. As such, they are subject to response biases and may capture some inaccuracies.

5.4 Future Directions

Despite these limitations, our findings have important implications for the measurement of facial behavior. Future research should continue to explore the limitations of FACS to verify that this system fails to capture emotions at the same rate or frequency as OpenFace and importantly why this is the case. Specifically, exploring the impact of speech on AU recognition would allow us to appreciate some of the barriers to manual coding systems. Likewise,
understanding which AUs are coded more infrequently and why, could enhance FACS training and overall sensitivity. Larger longitudinal studies are also required to confirm the predictive accuracy of OpenFace in the present study and allow for the comparison of several emotions as predictors of psychological outcomes. Lastly, further exploring our finding that anger predicted less PTSD 6 and 12 months later and looking at the expression of anger longitudinally to determine when it becomes detrimental to our health would benefit clinical treatment.

5.5 Conclusion

While most survivors of potentially traumatic events follow a resilient trajectory, a significant minority will later develop severe psychopathology. Facial behavior captured in the acute aftermath of a potential trauma, can help guide the identification of at-risk individuals for early treatment interventions. Until recently, manual coding systems like FACS were the gold standard in research for measuring facial expressions. However, these methods are time-consuming, labor-intensive, and ultimately subjective. Recent advances in automatic facial detection appear to overcome the inherent limitations of manual coding, but require validation to establish their efficacy in assessing risk factors of psychopathology. Until the present study, it remained unclear whether automated coding systems of facial expressions could effectively predict outcomes above and beyond manual coding systems. The current findings suggest automatic coding systems are in fact more precise and accurate than manual methods at measuring facial behaviors. Specifically, OpenFace identified more facial expressions and had greater predictive accuracy than FACS. These findings have important implications for risk assessment and targeted treatment interventions.
References


# Appendix A

<table>
<thead>
<tr>
<th>Upper Face AUs</th>
<th>Lower Face AUs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AU</strong></td>
<td><strong>Name</strong></td>
</tr>
<tr>
<td>1</td>
<td>Inner Brow Raise</td>
</tr>
<tr>
<td>2</td>
<td>Outer Brow Raise</td>
</tr>
<tr>
<td>4</td>
<td>Brow Lowerer</td>
</tr>
<tr>
<td>5</td>
<td>Upper Lid Raise</td>
</tr>
<tr>
<td>6</td>
<td>Cheek Raise</td>
</tr>
<tr>
<td>7</td>
<td>Lids Tight</td>
</tr>
<tr>
<td>43</td>
<td>Eye Closure</td>
</tr>
<tr>
<td>45</td>
<td>Blink</td>
</tr>
<tr>
<td>46</td>
<td>Wink</td>
</tr>
<tr>
<td>70</td>
<td>Brows Not Visible</td>
</tr>
<tr>
<td>71</td>
<td>Eyes Not Visible</td>
</tr>
<tr>
<td><strong>Head Positions</strong></td>
<td><strong>Miscellaneous AUs</strong></td>
</tr>
<tr>
<td>51</td>
<td>Turn Left</td>
</tr>
<tr>
<td>52</td>
<td>Turn Right</td>
</tr>
<tr>
<td>53</td>
<td>Head Up</td>
</tr>
<tr>
<td>54</td>
<td>Head Down</td>
</tr>
<tr>
<td>55</td>
<td>Tilt Left</td>
</tr>
<tr>
<td>56</td>
<td>Tilt Right</td>
</tr>
<tr>
<td>57</td>
<td>Forward</td>
</tr>
<tr>
<td>58</td>
<td>Back</td>
</tr>
<tr>
<td><strong>Eye Positions</strong></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>Eyes Left</td>
</tr>
<tr>
<td>62</td>
<td>Eyes Right</td>
</tr>
<tr>
<td>63</td>
<td>Eyes Up</td>
</tr>
<tr>
<td>64</td>
<td>Eyes Down</td>
</tr>
<tr>
<td>65</td>
<td>Walleye</td>
</tr>
<tr>
<td>66</td>
<td>Crosseye</td>
</tr>
<tr>
<td><strong>Lip Parting and Jaw Opening</strong></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Lips Part</td>
</tr>
<tr>
<td>26</td>
<td>Jaw Drop</td>
</tr>
<tr>
<td>27</td>
<td>Mouth Stretch</td>
</tr>
</tbody>
</table>
**Action Units 25, 26, 27 – Lips Part, Jaw Drop, Mouth Stretch**

These three Action Units are considered together because they concern mouth opening, which involves separation both of the lips and of the teeth\(^1\). AU 25 specifies how far the lips are parted. AU 26 specifies how far the jaw drops when the muscles that act to close the jaw relax. Thus AU 26, like AU 43, reflects a change in appearance from a neutral position that is produced by muscular relaxation rather than contraction: the more relaxation, the more the appearance changes. AU 27 measures the forced opening and stretching of the mouth by muscles that act in opposition to muscles that close the jaw. Whereas AU 26 describes the limited opening of the oral cavity (i.e., teeth parting) that can be produced by relaxing the muscle that closes the jaw, AU 27 reflects other muscles that contract and pull the mandible down to open the mouth wide. Although the muscle that relaxes to form the appearances of AU 26 might relax even more as the jaw is pulled open by muscles underlying AU 27, you never score AU 26 with AU 27. It is also possible that a small opening of the jaw to about the same extent as might be produced by relaxation of the muscles that close the jaw, is due primarily to the contraction of the muscles that lower the jaw, not relaxation. This distinction can be a subtle one that you make by observing the quality and timing of the jaw movement, as explained below. Although AUs 25, 26, and 27 imply something about how open the mouth is (i.e., separation of the lips and of the teeth), they primarily describe different behaviors rather than the degree of mouth opening.

AUs 25, 26, and 27 have optional intensity scoring that helps to refine the description of degree of mouth opening. Scoring the intensity of these AUs provides a relatively complete description of the degree of mouth opening. The intensity criteria for scoring AU 25 are not absolute, but rather are relative to the separation produced by the lowering of the jaw alone, as indicated by the score for AUs 26 and 27. Many different actions can affect the separation of the lips and the intensity score for AU 25. The intensity scores for AUs 26 and 27 are more closely tied to specific muscular actions than the more descriptive intensity scores for AU 25. Appropriate use of the optional intensity scoring for AUs 25, 26, and 27 can provide a descriptive picture of mouth opening in studies that need such information.

**A. Appearance Changes due to AUs 25, 26, 27**

**AU 25**

1. The lips part, which may expose the inner mucosal area of the lips more.
2. The teeth and gums may be exposed.
3. The oral cavity may be exposed, depending on the action of AUs 26 and 27.

Henceforth, whenever AU 25 is added to an Action Unit or combination simply to denote that the lips are parted, the appearance changes associated with AU 25 will not be repeated. You should refer back to this section if you want to review the changes.

Inspect image 25; Section C for these AUs has several videos of AU 25 illustrating intensity scoring.

**AU 26**

1. The mandible is lowered by relaxation so that separation of the teeth can at least be inferred.
2. If the lips part, space between the teeth may be seen; score 25+26.
3. Mouth appears as if jaw has dropped or fallen with no sign of the jaw being pulled open or stretching of the lips due to opening the jaw wide.
4. The time course of the action is relatively slow as the muscle relaxes.
5. It is possible for the mandible to be lowered and yet for the lips to remain closed. Often when this happens, you see the lowering action and the lips pressed by AU 24 (described in Chapter 7), or the lower lip pushed up by AU 17. It

---

1. In the previous edition of the Manual, the issues of degree of mouth opening and the muscular basis for these AUs were intertwined. The following, more extensive description separates and clarifies these issues.

57
is also possible for the surface friction of the two lips or stiffness of the lips to maintain the shut-lip position when only the mandible is lowered. Also look carefully for the presence of 17 and/or 24. Score 26 alone to indicate a jaw drop without parted lips.

AU 27

1. The mandible is pulled down.
2. Can open the mouth quite far as the mandible is pulled down, changing the shape of the mouth opening from an oval with the long axis in the horizontal plane to one in the vertical direction.
3. Mouth does not appear as if it has fallen open but as if it is actively pulled down forcibly or stretched open widely.
4. The lips may be stretched vertically by the extent of the opening of the mouth.
5. Flattens and stretches cheeks.
6. Changes shape of skin on the chin boss and the appearance under the chin.
7. It is possible to detect an AU 27 without the lips being parted when the mandible appears pulled down. Score closed lip jaw lowering as 27 alone, otherwise and typically, the AU 27 is scored with a 25 to indicate the lips parting.

B. How to do AUs 25, 26 and 27

Relax your face, keep your teeth together without clenching them. If you relax your lips and other muscles in the mouth area, your lips may part. This action is AU 25 alone. Some mouths are such that mere relaxation does not part the lips. If you find your lips do not part when you relax them, relax your jaw so that the teeth part slightly and let a breath help break the stickiness of the lips so that the lips part. Try to do this so slowly and slightly that you cannot visually detect in your mirror the dropping of your jaw or any blowing through the lips. If your lips part, this action is 25 alone. If you can detect the jaw drop, the score is 25+26; if you can detect the blowing, you might score 25+33 (blow). There are several actions that easily part the lips and increase their separation, such as pulling the upper lip up with AU 10 or the lower lip down with AU 16. In these cases, you do not score 25 alone, but add 25 to the muscular action, i.e., 10+25 or 16+25.

To perform AU 26, relax your mouth and let your jaw fall open; do not pull or force your jaw open, just relax the muscle that clamps your teeth together and let your jaw fall open so that your teeth are separated. You have made AU 26. If you are doing this correctly, there is no muscular tension in your lips, and no stretching of your lips. Once your jaw closing muscle is relaxed enough to permit the jaw to drop wide enough, your lips should part, scored 25+26. It is easier to detect a 26 if the lips part than if they remain closed. How soon the lips part with a drop of the jaw varies in different people, and as indicated above, some people need no jaw drop to part their lips. Examine whether your jaw needs to drop in order to part your lips, and if so, how far it must drop. Notice also how far you can drop your jaw merely by relaxing, not pulling it.
open. You should find that this jaw drop is limited; beyond this limited extent, AU 27 is scored because a muscle must act to pull the jaw open further. Most people can drop their jaw with AU 26 so that the tongue or index finger can fit between the teeth, but not much more, when their head is in a normal upright position. Notice what happens at the corners of your lips as the mouth and lips move from being closed, to the relaxed drop of the jaw of AU 26, and beyond as you pull your mouth open. Relax your jaw to perform an AU 26, then nod your head up and down – the relaxed jaw drop is greater when the head is back than when it is forward.

To perform AU 27, pull your jaw down, far down, opening your mouth wide open as if a physician were going to examine your tonsils. You have made a large AU 27, scored 25+27 because of the lip parting. Note that while you have not tensed your lip muscles, they are stretched somewhat by the extent of opening of your mouth. The shape of your mouth opening is also stretched in the vertical direction. Notice what happens to the position of the red parts of the lips relative to the teeth when the jaw is dropped, keeping the lips relaxed, i.e., no other actions moving the lips. You should see that the center part of the upper lip does not change its position relative to the upper teeth, but the lateral parts near the corners are pulled down relative to the upper teeth as the mouth is stretched open. The center and corners of the lower lip, on the other hand, move higher relative to the lower teeth as you stretch your jaw down, which tends to pull the teeth down more than the lower lip. You need to be familiar with the effect of jaw dropping and stretching alone on the position of the lips relative to the teeth in order to score the intensity of AU 25 with 27. Can you cover your mouth opening with your lips when the stretching open of the jaw is maximum? Try to perform the minimum AU 27 starting with your teeth together, then rapidly pull or snap your jaw down to open your mouth about as much as you were able to relax it open when performing AU 26. If you need some motivation to help you do this movement, imagine you were trying to yell “Hey!” to someone as a warning. Look in the mirror to see how the time course of this rapid, jerky jaw movement differs from that of the much slower relaxed opening of AU 26, which results in much the same appearance at its end point. This quality of movement is what you must use to distinguish an AU 26 from an AU 27 when the jaw drop is no more than the maximum that can be produced by merely relaxing the jaw closing muscle. A jaw drop beyond the maximum that can be produced by relaxing (space between the teeth about that which permits the tongue or index finger to fit) must be scored an AU 27, regardless of the rapidity of the movement.

When you have performed and examined the appearances of 25, 26, and 27, try combinations of 25 with 26 and 27, i.e., drop or stretch your jaw open with relaxed lips and notice how far the lips part. Then pull the lips apart even farther with other AUs, such as 10 and 16. Determine what combination of AUs produces the greatest absolute lip separation. Starting with the teeth together, try separating the lips as much as possible using AUs such as 10 and 16, then keep the lips separated and drop and stretch your jaw, noticing the range of lip separation possible. Stretch your mouth wide open to various degrees and then bring your lips closer together than the separation caused merely by the jaw lowering. Separation of both the lips and the teeth contribute to mouth opening. These are the distinctions that intensity scoring for AU 25 is intended to capture.

C. Intensity Scoring for AUs 25, 26, and 27

Intensity scoring for AUs 25, 26, and 27 is intended to enhance the description of mouth opening that is incidentally reflected by the AUs themselves. This intensity scoring is optional because many people will determine that the behavioral distinctions needed for their research are captured by the AU scores alone, without intensity. If your research requires a more complete description of mouth opening, scoring the intensity of all three AUs, 25, 26, and 27, is recommended because the representation of mouth opening depends on knowing the degree of both lip separation and jaw lowering. For many studies, the simple decision whether the lips are parted by 25 or not may be sufficient, in which case, you need not learn the intensity scoring for 25, only what constitutes lip separation. Other studies may need scores for jaw separation, but not lip separation, while other studies need no intensity scoring of these AUs.

The intensity scores for AU 25 are descriptive in that they indicate how far the lips are parted, regardless of the muscular actions, if any, that produce the parting. Several AUs can part the lips while other AUs can unite or bring them closer together, and these actions and their combinations move the lips into many shapes, positions, and separations, which we are not attempting to describe comprehensively. The assignment of intensity to AU 25 is defined to reflect the lip separation relative to the jaw opening (space between the teeth) rather than an absolute distance between the lips. This measurement of lip parting relative to jaw lowering necessitated a change in the use of our five-point intensity scale that may at first seem odd, but it serves the purpose of describing mouth opening. These factors make the intensity scoring for AU 25
slightly more complex than other AUs, and the text for its description more lengthy. In brief, as you read the criteria for the intensities of AU 25, keep in mind that the use of the A-B-C-D-E intensity scale is different when the jaw is closed versus lowered. When the jaw is closed, 25A represents a minimal lip separation, 25E, a maximum separation, etc., similar to scoring the intensity of other AUs. When the jaw is lowered, however, our use of the intensity scale changes. Simply score 25C if the lips are separated and this separation is no more than a trace different from that which might be caused by the lowered jaw alone. If the separation of the lips is less than what the lowered jaw would cause itself, use the code ‘B’ to indicate at least slightly less, and the code ‘A’ to indicate at least severely less separation. On the other hand, if the separation of the lips is more than what the lowered jaw would cause itself, use the code ‘D’ to indicate at least slightly more, and the code ‘E’ to indicate at least severely more separation. You will read how to determine these scores in the criteria detailed below. We suggest that novice FACS coders not learn all the details of intensity scoring for AU 25 on their first reading of the Manual because understanding and applying the criteria is easier after one becomes familiar with all the AUs that act to separate or bring the lips together. Instead, the new FACS coder should begin by understanding what constitutes lip separation, and score 25 without intensity, then after completing the rest of the Manual, return to study the scoring criteria for AU 25, if its intensity is to be scored.

The intensity scores for AUs 26 and 27, like those for most other AUs, indicate an aspect of the underlying muscular action, relaxation of one in the former, contraction of others in the latter. AUs 26 and 27 do not form a continuum describing the degree of jaw lowering – they describe different ways the jaw may be lowered. The intensity scoring for AUs 26 and 27 describes the degree of jaw lowering produced by the two jaw-opening mechanisms.
<table>
<thead>
<tr>
<th>Example Image</th>
<th>FACS Score</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td>0</td>
<td>The face is not showing any action, often called a neutral face. The face is not actually at rest because the eyes are open, the jaw is closed, etc., but no AU can be scored. This neutral face is the baseline for scoring AUs in the example images of this person.</td>
</tr>
<tr>
<td><img src="image2.png" alt="Image" /></td>
<td>4D</td>
<td>Severe vertical wrinkles and skin bunching between the eyebrows, severe drawing together of the eyebrows, and severe to extreme eyebrow lowering put this 4 at the D intensity. The arch in the lowest horizontal furrow of the forehead has straightened because of the lowering and pulling down of the glabella. There is no evidence of AU 5 when looking closely at only the eyelid. The dark shadow between the lips should not entice the coder to score 25 because 25 requires certainty, but they are certainly closed.</td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td>10B</td>
<td>The upper lip is raised slightly and the nasolabial furrow has deepened slightly to form the characteristic AU 10 pouch.</td>
</tr>
<tr>
<td><img src="image4.png" alt="Image" /></td>
<td>25B</td>
<td>The teeth clearly show, and the lips are separated by about 2mm (not more than slightly greater than 2mm), thus meeting the B level on both criteria. Nothing suggests that the jaw has dropped even though the upper teeth are not clearly visible.</td>
</tr>
</tbody>
</table>
Summary of Scoring Steps

I. Initial Scoring

List your initial scoring on line I, using parentheses, e.g., ( ref ), ( or ), ( R or B ), ( ? ), ( C or D ), for uncertain scores and no parentheses for definite scores. Do step I quickly, noting your first impressions. In step II you will carefully consider each AU.

II. Checking for Omissions

Check the Upper Face AUs on the back side of the Score Sheet to discover AUs which you failed to consider in step I.

Make a check mark for scores listed in step I, a minus sign – for not applicable scores, and a triangle for new scores, definite or uncertain. Add these new scores, definite and ones in parentheses, to line II of the Score Sheet. This step will require viewing the facial behavior several times.

III. Reorganizing the Initial Scores

Rewrite the list of scores to incorporate the step I and step II scores. Enter the list of scores on line III.

IV. Checking Alternative Rules and Reference Sections

Check each AU, definite or uncertain, in scoring list III in the Alternative Rules Table 2-1. You can reorder the scores under IV on the Score Sheet to facilitate this check. Write on the Score Sheet any Alternatives that apply. Note any revised intensity criteria found in the Reference sections. Rewrite the scoring list to incorporate any ( or ) expressions required by Alternative Rules. Enter this revision on line IV. You may wish to consult Subtle Difference Tables at this point.

V. Verifying Intensity Criteria and Unilaterality Questions

Score intensity for any AUs where intensity is not optional. Settle any ( L/R or Bi ) questions.

Facial Action Coding System 80 Interrelationship Among Scoring Steps

VI. Final Decisions

If questions remain, check ( or ) expressions in Subtle Difference Tables. Recast scores into sets to show your possible choices including definite scores and one of the ( or ) AUs, and refer to Subtle Difference tables and Sections A, B, and C. Imitate the complete sets of possible scores on your own face. Enter the result below Final Scoring Upper Face.
Score Sheet #2: Example Photo 1

Facial Action Coding System: Score Sheet

Designed by Paul Ekman and Wallace V. Friesen

Lower Face
I. Initial Scoring: _______________________________________

II. Omission Check: ______________________________________

III. Reorganized Scoring: __________________________________

IV. Reference Check:
- AUs in Numerical Order: ________________________________
- Alternative AUs: _______________________________________
  Reference Check: _______________________________________
- Results for Step IV: ____________________________________

V. Revised Scoring: _______________________________________

Head/Eye Position: _______________________________________

Upper Face
I. Initial Scoring: $1+2 + (5B \text{ or } 5C)$

II. Omission Check: $7 \text{ ref}$

III. Reorganized Scoring: $1+2 + (5B \text{ or } 5C) + (7 \text{ ref})$

IV. Reference Check: (especially: 4 with 9; 6 with 9, 10, 12, & 13; 7 with 6, 12, & 13)
- AUs in Numerical Order: $1+2+5+7$
- Alternative AUs: ____________ Reference Check: $7 \text{ wr/5}$
- Results for Step IV: ____________________________________

V. Revised Scoring: $1C + 2C + 5C$

Final Scoring Upper Face: $1C + 2C + 5C$

Final Scoring Lower Face: ________________________________

Final Head/Eye Positions: ________________________________

Final Full Face Score: ________________________________

(Score 73 if Entire Head/Face is out of view)

Coder's Name: _______________________ Date: ________________ Time: ________________

Stimulus: ___________________________ Segment: ________________ Item: __________________

Location: Beginning: ___________________ End ___________________

© Copyright 2001 Paul Ekman, Wallace V. Friesen, & Joseph C. Hager - Permission to reproduce this two page scoresheet is granted.
Scoring Example 1:

Compare the image Example 1 with the neutral face. Examine line I of Score Sheet 2. The initial scoring is a definite 1+2 that accounts for the brow raise, and uncertainty about whether 5 was 'B' or 'C' in intensity. The Omission Check reveals a possible AU 7, and "(7 ref)" is added to the scores on line II. The reorganized scoring on line III incorporates the new, possible AU 7. The AUs are listed in order under IV for the Alternative Rules and Reference Check. No applicable rules are listed so the word "none" is written next to "Alternative AUs." The Reference for AU 7 shows that the criteria for 7B changes in 5+7. No reorganizing of scores is needed so line Results for Step IV is empty. For the intensity of AUs 1 and 2, the intensity guidelines in 1+2secC on page 54 indicate that the observed marked or pronounced raising of the brow is scored C. The raise of the outer brow is close to being severe or at the D level, but given the arched permanent shape of the brows in neutral, the D level is too high. These scores are added to line V Revised Scoring.

Presence of 7? If you look up in the Reference: Action Units that Change the Intensity Criteria for AU 7 on page 30, you will find that there is a separate entry for 7B in a 5+7 combination. Since this is a still image the following criteria for 7B are used:

- the raising of the lower eyelid skin over the bottom of the eyeball causes a slight bulge to appear in the lower lid since this skin is stretched over the bottom of one eyeball.
- slight pouching of the lower eyelid skin as it is pushed up.

There is almost no change in the lower eyelid, but what does appear is partially contrary to what 7 might produce: more of the lower rim of the iris is exposed than in neutral and the eyelashes in the lower lid indicate that the trace of stretching in the lower lid is lateral, not medial as in 7. The outer edges of the lower eyelid may be pulled up a trace, consistent with AU 7 or 5, but the action of AU 5 accounts better for these three identified changes than 7. The revised scoring on line V has eliminated the (?) expression.

Intensity of 5? All that remains is the decision about whether 5 is 'B' or 'C.' Look up the intensity scoring for AU 5 in 5secC on page 25. The neutral image show the upper lid covering part of the iris, and the face being scored shows virtually all of the iris in the left eye and perhaps even a bare hairline of sclera (B level), but in the right eye slightly more than a hairline of sclera shows above the iris (C level). Using the intensity in the more intense eye, the scoring should therefore be 5C, which is added to line V. The final scoring 1C+2C+5C is entered on the Final Scoring Upper Face line on the Score Sheet.

Scoring Example 2:

Compare the image Example 2 with the neutral face. Examine line 1 on Score Sheet 3. The Initial Scoring is a definite 1, to account for the raising of the inner brow, and uncertainty about the presence of AU 43, to account for the narrowed eye aperture. The Omission Check suggests that AUs 4 and 7 may be present, so they are added on line II. The coder remembers that the intensity criteria for 4 in addition to 1 are different from those for 4 alone, and so enters the note (1+4 ref) instead of (4 ref) on line II. Line III shows the reorganized scoring integrating the scores from lines 1 and II. AUs 7 and 43 or both could account for the narrowed eye aperture, indicated in the notation (7 or 43) on line III. The check for Alternatives reveals no alternative AUs for 1, 4, or 7.

Presence of AU 7? The Reference for AU 7 lists no changed criteria for 7 in addition to 1 or 1+4 (and both must be considered since at this point the coder does not know whether the final scoring will be 1 or 1+4) so that the criteria for 7 alone are used. There is evidence of AU 7 beyond the narrowed eye aperture (ignoring for now that this appearance might be produced by either AU 43 or her downcast eyes). There is a trace of increased wrinkling and increased bagging in the lower eyelid, and a trace of medial drawing together of skin, especially in the changes at the inner corners of the eye, and an indication of tension in the trace of change in the shape of the lower eyelid. These signs are not produced by 43. The intensity guidelines for AU 7 indicate the evidence for 7 is insufficient to meet the criteria for 7B in this face, but the coder is certain that 7 is present based on the appearance changes enumerated above, so scores 7A.

Presence of 43? The coder next checks the criteria for 43. The Reference for AU 43 lists no changed criteria for 1+4+43, or 1+43, and a minor change only for 7+43E (and the eyes are definitely not 43E), so the 43 alone criteria are applied. The
evidence for 43 is marginal because only a trace more of the upper eyelid is exposed and the eye aperture is only slightly less wide at most (ignore for now that she looks down, a complicating action). Thus, AU 43, if scored, would be A level based on the intensity guidelines. The Subtle Differences table on page 67 lists a contrast for 7 vs. 43 vs. 7+43 that is relevant to deciding between 7 and 43. (Also, 7secC talks about the criteria for scoring AU 43.) Having made the case for 7 (with no 5), a 43 is almost redundant, and in this case we decide not to include the score because it is not necessary to do so and the still image cannot satisfy the .25 second criterion for 43A. (Later, you will learn that the downward gaze direction is a better explanation of these signs, and so 43A would not be a wise score anyway.)

**Presence of 47?** The last intensity criteria to be checked are for 4 in a 1+4. The evidence for AU 4 is apparent in the slight vertical wrinkling and skin bunching in the glabella, the marked pulling together of the eyebrows, and the shapes of wrinkles and eyebrows. Reading the guidelines for intensity, you can verify that evidence is sufficient to score 4B.

The definite AU 1 showing marked wrinkling is scored 1C using the intensity guidelines for AU 1. The revised scoring in line V shows that the scoring for this example is 1C+4B+7A.
Score Sheet #3: Example Photo 2

Facial Action Coding System: Score Sheet

Designed by Paul Ekman and Wallace V. Friesen

Lower Face
I. Initial Scoring:

II. Omission Check:

III. Reorganized Scoring:

IV. Reference Check:
   - AUs in Numerical Order:
   - Alternative AUs: Reference Check:
   - Results for Step IV:

V. Revised Scoring:

Head/Eye Position:

Upper Face
I. Initial Scoring:
   1 + (43 ref)

II. Omission Check:
   (1+4 ref) (7 ref)

III. Reorganized Scoring:
   (1 or 1+4 ref) + (7 ref or 43 ref)

IV. Reference Check: (especially: 4 with 9; 6 with 9, 10, 12, & 13; 7 with 6, 12, & 13)
   AUs in Numerical Order: 1 + 4 + 7 + 43
   Alternative AUs: none Reference Check:
   Results for Step IV:

V. Revised Scoring: 1C + 4B + 7A

Final Scoring Upper Face: 1C + 4B + 7A

Final Scoring Lower Face: Final Head/Eye Positions:

Final Full Face Score: (Score 75 if Entire Head/Face is out of view)

Coder's Name: Date: Time:

Stimulus: SAMPLE Image 2 Segment: Item:

Location: Beginning End

© Copyright 2001 Paul Ekman, Wallace V. Friesen, & Joseph C. Hager - Permission to reproduce this two page scoresheet is granted.