Profiling Facial Expressions of Emotion: Insights from Their Emotional, Semantic, and Contextual Similarities

Laura Manno

100191204

School of Psychology, University of East Anglia

September 2023

A thesis submitted for the degree of

Doctor of Philosophy

Under the supervision of

Dr. Mintao Zhao, Prof. Andrew Bayliss, Dr. Stephanie Rossit

Declaration

I hereby declare that the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Laura Manno

September 2023

Acknowledgments

I would like to express all my gratitude to everyone who has supported and guided me throughout my journey towards completing this PhD. First of all, my Supervisor, Mintao Zhao, for his continuous guidance, encouragement, and expertise. To my extended Supervisory Team, Stephanie Rossit and Andrew Bayliss, and the dedicated technicians, Lizzie Watson and Richard Carey-Knight, for their insightful feedback and invaluable suggestions, which have greatly enriched the quality of my work.

To my family, whose unconditional and warm love has always reached me, even from miles and miles away. Especially to Mum and Dad, because your belief in me has been a constant source of strength, and because no matter the choices I make or the number of times I stumble, I always feel that you are proud of me.

To Tracuzzi, you took care of me when I didn't have the strength to take care of myself, and wherever I go, if you are with me, I am at home. And of course, to Brie, (yes, friends, my cat tops the list), for being my silent and fluffy emotional support. No matter the level of stress, a single look at you makes everything melt away.

To the friends I made during this journey, despite my initial determination not to make new friends, it was impossible not to love you all. Agatha, Paula, Ghizzy, Prerna, and Alvaro, thank you for inspiring me with your stories, experience, work, and ideas. Thank you for the shared laughter, tears, the fun we had together and all your support and sweet words.

A special thanks to *The Crew*. To Nicho, so similar and at the same time so different from me, the one that truly understands the importance of the "useless". Thanks for indulging in all my nonsense, thanks for teaching me how to turn bad experiences into things to laugh about, and for always knowing what to say to cheer me up, but most importantly… thanks for being so "orange"! To Greata, it's not easy to find someone who is happy about your happiness. Thank you for being that person and my constant support. Thanks, because even though I always complain about the same staff, you keep listening and replying to me as if

it were the first time. To Luca, literally the first face I encountered at UEA, you patiently guided me when I was far away from everyone and everything, when I was afraid of not being enough. Thanks, because you bought me cakes when I was sad, and, especially, for appreciating my 15-minute voice messages. To Alessia, for always being there in the front row when the going gets tough, we didn't even have time to ask for help and you were already assembling our furniture. Thanks for teaching me the art of not killing plants, and for always being smiling and full of joy. To Francesco, because even if you put all your effort into being grumpy you are the sweetest of us all. Our unicorn games were twice as entertaining because of you. Thank you for supplying whisky for all of us for three years (honestly, you were a “buy one get two” with Nicho, but you have been a good deal after all). To Becky, it’s too late for you…you are Italian now! Thanks for patiently waiting for me to finally be confident enough to speak to you in English and thanks for becoming my friend. Least but not last, to Alice, thank you for the countless hours spent philosophizing about love, work, friendship, everything. Overthinking with you is an absolute blast!
Finally, to all my "San gioacchini e acquisiti," we used to say that we are not friends, because you can choose your friends. I didn’t have the chance to choose you…you all happened! and as a family, we learned to love each other despite our differences. I still struggle with the idea of being so far away from each one of you. Thank you for understanding my absence and for always being my family.

To all the incredible friends and people that I don’t have space to mention (Olita, Cavallo, Clara, Stefania, Rigo, Brunella….is a long list) this thesis represents a testament to all the support and encouragement I received from you during my life. Thanks for being a part of this incredible journey.

Abstract

Facial expressions of emotions serve as a fundamental form of nonverbal communication, offering a rich tapestry of information about individuals' emotions, intentions, and thoughts. While classical theories of emotion perception have laid a solid foundation for understanding facial expressions, they may oversimplify what we perceive from facial expressions of emotions by reducing it into a few basic emotions or two emotion dimensions, potentially missing the richness and complexity of real-life emotional experience. In this dissertation, I explore a more sensitive approach to the study of facial emotion processing by conducting six behavioural and cross-cultural studies employing multiple-dimensional profiling tasks and Representational Similarity Analysis (RSA). This allowed me to (1) investigate whether people extract complex, high-dimensional emotional content from facial expressions of emotions; (2) determine if high-dimensional representations of facial emotions outperform classical categorical emotion models in predicting perceptual similarities between facial emotions; (3) uncover what stimulus- and observer-based factors underlie perceptual similarity between facial emotions; (4) examine the impact of participants' cultural background, emotion intensity, facial motion, and context on the profiling and perception of facial emotions; and finally, (5) explore how human facial emotion processing may differ from machine learning approaches to emotion perception.

Throughout the six studies reported here, participants engaged in a series of profiling task where they reported their perception of facial emotions along multiple emotional, semantic, and contextual dimensions, generating unique profiles for each facial emotion under different conditions. Response profiles were compared across different cultural backgrounds (Chinese vs. British participants), facial motion (Static vs. Dynamic), emotion intensity (High vs. Low), and emotional contexts (Physical vs. Social scenarios; congruent vs incongruent). Participants also performed a direct similarity rating task to produce a measure of perceptual similarity between facial emotions. Finally, I obtained other measures of similarities based on different sources of information (i.e., Physical, Categorical, Profiling and

Intensity similarity) and performed RSA and multiple regression analysis to identify underlying factors that contribute to perceptual similarity of facial emotions.

The results showed that (1) facial emotion perception is complex and multi-dimensional, integrating rich emotional content, fine-grained semantics, and relevant contextual information; (2) multi-dimensional emotion profiles outperformed traditional categorical emotion models in predicting perceptual similarities between facial expressions; (3) perceptual similarity is influenced by both physical stimulus-based cues and high-level perceiver-based emotion perception; (4) participants' cultural background, emotion intensity, facial motion, and emotional contexts significantly impact facial emotion perception; and finally, (5) Machine learning models, while achieving human-level emotion categorization, may not capture the richness and complexity of human emotional experience, as reflected in emotion profiles.

These findings underscore the importance of recognizing the complex nature of human emotional experience and the effectiveness of an emotion profiling paradigm in revealing the rich and diverse information perceived from natural facial expression of emotion. Theoretically, the present results challenge the prevailing view that emotion perception is universal, discrete, and best described as a single semantic label. Instead, these results provide further support for the emerging view that perception of emotion is multiple dimensional, blended and varies in a gradient way. Methodologically, the profiling paradigm used in this project is not only able to reproduce many classical findings in emotion research (e.g., difference across cultures, facial motion, and emotional context), it also uncovers novel and fine-grained differences regarding emotional, semantic, and contextual information conveyed by facial expressions of emotions. Practically, theories and models of facial emotion perception play a pivotal role in various aspects of daily life, from machine learning based face processing to therapeutic interventions. By incorporating a more holistic and multi-dimensional perspective into emotion perception, it may help these practical settings design better and sensitive tools, techniques, and interventions that captures the complex nature of human emotion experience with facial expressions of emotions.

Contents

Chapter 1

Theoretical Background

1.1 Introduction

Facial expressions are a powerful form of nonverbal communication that provides a wealth of information about others' emotional state, intentions, and thoughts, guiding our social interaction and affecting our behavior (Sander et al., 2007; Scarantino, 2017). While verbal language is a fundamental component of human communication, it only represents a fraction of the information conveyed during human social interactions. The majority of messages are transmitted through individuals' nonverbal behavior and the significance of distinct communicative cues, either verbal or nonverbal, can vary depending on the specific context (Friedman, 1978; Mehrabian, 1971). Facial expressions, nonverbal vocal cues, gestures, body postures, interpersonal distance, touching, and gaze are the "silent languages" that accompany our every interaction (Matsumoto et al., 2013). Among these, facial expressions are complex signals that we seem to easily process every day. Even when not directly engaged in a face-to-face interaction, we continuously extract meaningful cues from the faces of people around us (Latané & Darley, 1968; Mumenthaler & Sander, 2012; Parkinson & Simons, 2009), which trigger inferential processes and affective reactions (Van Kleef, 2009). Reading others' faces, we infer their emotions, goals, desires, relational orientation, behavioural intentions, and appraisal of the ongoing situation, which then leads us to adaptively change our behavior (Keltner & Ekman, 2000; Keltner & Haidt, 1999; Oosterhof & Todorov, 2008; Shuman et al., 2015). Similarly, contagion processes (e.g., mirror neuron activity, mimicry), make us experience affective reactions comparable to the facial emotions observed, leading us to adjust our behavior accordingly (Bastiaansen et al., 2009; Herrando & Constantinides, 2021; Prochazkova & Kret, 2017). Whether we are aware of it or not, our behavior is influenced by processing others' facial expressions and our brain has a remarkable ability to recognize facial emotions quickly and accurately, reflecting its fundamental importance in social cognition and communication (Krautheim et al., 2020; Muukkonen & Salmela, 2022).

Various theories have been developed to offer different perspectives on the nature and function of facial expressions. According to traditional widely accepted models, facial emotions are mostly conceptualized into a limited set of discrete basic categories (Ekman, 1992), or represented through a combination of two, or a few more, dimensions such as valence (positive vs. negative) and arousal (low vs. high) (Russell, 1980). Emotions are mostly considered either as innate biological-based processes, automatically and unconsciously expressed across cultures (Ekman & Friesen, 1971), or learned concepts shaped by social norms, cultural values, and individual experiences (Russell et al., 2003). Nonetheless, while the approaches and methods adopted in these classic theoretical frameworks (e.g., stimuli depicting prototypical facial expressions, categorization tasks, etc.) are particularly suited to reduce complexity and provide a common basis for understanding emotions, may not be sufficiently sensitive to reveal the richness and complexity of human emotional experience in daily life (Cowen et al., 2019; Jack et al., 2018).

In our daily life, emotional experiences are often complex. Individuals find it challenging to accurately discern and articulate their own emotions and, when asked to report their mood, they typically do not rely on one single emotion category (Moore & Martin, 2022; Saarni, 1999; Trampe et al., 2015). Emotions are represented within a semantic space that include plentiful of terms that refer to a rich variety of emotional states (Barret, 2009; Sabini & Silver, 2005; Shaver et al., 1987). They are not experienced as isolated and distinct entities but rather complex and intertwined phenomena which are often ambiguous and overlapping. Studies investigating the subjective experience associated with emotions reported how these are highly intercorrelated both within and between participants reporting them (Russell & Carroll, 1999; Watson et al., 1999). Similarly, facial expressions do not usually convey a pure, single affective state, and, while we tend to agree on the main emotion category conveyed by a given facial expression, the extent to which other emotions are perceived may differ across individuals and cultures (Fang et al., 2018). We regularly express a broader range of emotions beyond the seven “universal” ones (Keltner et al., 2016; Laukka et al., 2013; Sauter & Scott, 2007), and the same emotion can be expressed via a variety of within-category expressions (Sauter, 2010; Shiota et al., 2017; Szameitat et al., 2009). Conversely, the same facial movements or expressions can acquire different meanings depending on the expressive behaviour simultaneously occurring in other modalities (Aviezer et al., 2008), the context in which they occur (Greenaway et al., 2018; Wieser & Brosch, 2012), as well as the cultural background of both the perceiver and the person exhibiting the expressions (Elfenbein, 2013; Elfenbein et al., 2007). Therefore, to understand how emotions are experienced in real-life, it is essential to develop methods and approaches that can effectively quantify the complexity of emotional experience and are sensitive to the variations across cultures and individuals.

Recent studies have broadened the scope of emotional communication research beyond facial expressions, delving into modalities like vocalization, bodily movement, posture, tactile contact, gaze, and autonomic responses (Cordaro et al., 2016; Dael et al., 2012; Graham & LaBar, 2007; Grewe et al., 2009; Hertenstein et al., 2009; Laukka et al., 2014; Lobmaier et al., 2008). Emotional behavior is increasingly viewed as a multimodal dynamic pattern, challenging the oversimplified view of emotions solely driven by evolution or culture. As a result, recent research has embraced a nuanced approach to emotional experience, exploring within-category variations in expression and perception across cultures. This departure from traditional dichotomies acknowledges the complexity of emotions, shaped by biological, cultural, and contextual factors. Thanks to the growing use of data-driven approaches and computational modeling, the investigation of emotional expression has extended well beyond the study of prototypical facial expressions associated with the six basic emotions (Cowen et al., 2019; Cowen & Keltner, 2020; Jackson et al., 2019). Innovative methodologies have facilitated the comparison between traditional categorical and dimensional approaches, leading to the recognition of integrative models where both dimensions and categories contribute to explaining emotional experience (Cordaro et al., 2018; Gendron et al., 2020; Jack et al. 2016, 2018). These models consider the influence of culture, prior beliefs, and contextual information on individual emotional responses (Brooks et al., 2019; Brooks & Freeman, 2018; Greenaway et al., 2018; Wieser & Brosch, 2012; Snoek et al., 2023).

Nonetheless, most of the tasks commonly used to collect behavioural responses still reflect the assumptions of classical theories of emotions, potentially neglecting precious information. Methods sensitive to complex and nuanced emotional experience may open the way toward a deeper understanding of facial emotion processing, resulting in models that can better account for the rich nature of facial emotions in real life. While extensive research has been conducted on the process of discerning facial emotions, predominantly relying on categorical interpretations, new studies are needed to understand how we perceive fine-grained differences and similarities between facial expressions of emotions.

In this project, I took an emotion profiling approach to investigate what information is perceived from facial expressions of emotions; to determine whether a complex emotional content, rather than a single emotion category, can better explain how we process facial expressions of emotions; and to uncover factors that may influence the way we perceive fine-grained differences between facial expressions of emotions. To do so, I used scenario-induced facial emotion stimuli and collected rich profiles of information about their emotional, semantic, contextual and physical aspects through behavioural studies and responses of computational models. I obtained a variety of measures of perceptual and physical similarity among the same set of stimuli and investigated how these could be potentially linked together. Understanding how people determine perceptual similarity between facial expressions of emotion would shed light on how facial emotions are perceived and represented.

In the following sections of this Chapter, I am going to review the prevailing existing theories on facial emotions, highlighting how recent research demonstrates the limitations of classical research methods and theories in the field, and how new research approaches may help establish a deeper understanding of facial emotion processing in real life.

1.2 Theories of facial emotion: the role of emotional Categories and Dimensions

1.2.1 Basic emotion theories and categorical approaches to understanding emotions

One of the most influential theories on facial emotion is The Basic Emotions Theory (BET). Built upon Darwin's pioneering theory on the evolutionary origins of facial expressions (Darwin, 1897), Ekman and colleagues proposed that certain facial expressions were reliably associated with specific emotions and that these expressions could be consistently recognized across cultures (Ekman & Friesen, 1971). Ekman termed these emotions as basic facial emotions. According to the BET emotions are biologically determined and influenced by our evolutionary heritage. Throughout evolution each emotion has been tailored to solve a specific adaptive problem and, because of this, it is characterized by unique components that differentiate it from other emotions, resulting in discrete and distinct emotion categories (Ekman, 1992; Ekman & Davidson, 1994; Tracy, 2014). Each emotion is believed to be characterized by four components (Keltner & Gross, 1999). These are the *Cognitive Appraisal*, emotions are thought to be influenced by our cognitive interpretation of a situation or event; the *Physiological Response*, emotions are also believed to be accompanied by physiological changes in our body (e.g., heart rate, breathing rate, sweating); the *Behavioural Response*, emotions are often expressed through our behaviour (e.g., facial expressions, body language, and other forms of nonverbal communication); and, the *Subjective Experience*, emotions are thought to be accompanied by a subjective experience or feeling. Facial expressions of emotions are a relevant part of our *Behavioural Response* and are thought to have developed due to their specific value in human survival and social communication. As a result, facial emotions are conveyed by prototypical configurations of facial movements that can be organized into a limited set of discrete basic categories which are similarly recognized and produced across different cultures (Ekman & Cordaro, 2011; Ekman & Davidson, 1994; Matsumoto et al., 2008).

Ekman and colleagues proposed that facial expressions could be represented as a pattern of facial movements, called Action Units (AUs) (Ekman & Friesen, 1978). Each of the 44 AUs, correspond to a specific set of facial muscles that provide a standardize way to identify and measure all facial expressions, creating a comprehensive coding system called the Facial Action Coding System (FACS) (Ekman & Friesen, 1976, 1978; Rosenberg & Ekman, 2020). Each basic emotion is represented by a standard pattern of AUs. Anger, for example, is typically characterized by a combination of specific facial muscle movements including a furrowed brow (AU4; AU7), which is created by the contraction of the corrugator muscle between the eyebrows, and a tightened jaw (AU5; AU23), which is created by the contraction of the masseter muscle in the jaw (see Ekman et al., 2002, Table 10-1). The BET recognizes that slightly variations of the prototypical expressions may occur, however, these are mainly considered a result of changes in the intensity and duration of the emotions being expressed. The FACS system can also be used to decode secondary emotions or blended emotions which, according to Ekman, arise from conflictual emotional experience or transition from one emotion to another. In this case, following an additive approach, blended emotions would result in predictable facial expressions composed by the action units of the different basic emotions involved (Ekman, 1972, 1973; Ekman & Friesen, 1975).

According to the BET, the expression and recognition of facial emotions may be influenced by culture, which play a top-down role in the process. Cultural norms and expectations may shape the way emotions are expressed and interpreted, interrupting the otherwise full coherence between emotional experience and emotional display and leading to some differences in the *intensity*, *duration*, and *frequency* of the emotional expression across cultures. However, it is important to notice that this within-category variation is considered to be the result of processes that are not emotion-related. The intensity and duration of an emotion are thought to be influenced by the cognitive appraisal of a situation (i.e., individual's subjective interpretation of a state based on own beliefs, values, and past experiences, see (Lazarus & Folkman, 1984), and cultural learned factors (Ekman, 1992; Ekman & Cordaro, 2011). Individuals from different cultures may learn different display rules from their social environment, norms and emotion-regulation strategies that dictate when and how emotions should be expressed in public, which in turn may affect how they express emotions by leading them to deintensify, intensify, neutralize, and mask an affect display (Ekman et al., 1969).

The core idea of the BET has been inherited by many different models of emotion perception, though these often diverge in some aspects. Specifically, the number of basic emotions they identify, the labels they attribute to those categories, the prevalence they recognize to the basic emotions in daily life, and the weight accorded to cultural influence in the process (Clore & Ortony, 2013; Tracy & Randles, 2011). For instance, while Ekman and Cordaro recognize seven basic emotions - Happiness, Sadness, Fear, Anger, Disgust, Contempt, and Surprise (Ekman & Cordaro, 2011), Izard's model replaces Surprise with Interest (Izard, 2011). Levenson, on the other hand, identifies five basic emotions - Enjoyment, Sadness, Fear, Anger, and Disgust (Levenson, 2011), while Panksepp and Watt propose Play, Panic/Grief, Fear, Rage, Seeking, Lust, and Care (Panksepp & Watt, 2011). Moreover, while these models recognize the critical role played by basic emotions in early development, they also argue that basic emotions tend to interact to each other, are influenced by higher order cognitive process, and often evolve into more complex emotional states (Tracy & Randles, 2011). For example, Izard (2011) argues that emotion schemas, and not basic emotions, characterize emotion experience in everyday life. Emotion schemas represent an individual's organized knowledge about emotions, which includes their personal experiences, cultural influences, and social learning. These are more complex cognitive structures that involve both basic emotions and additional cognitive and behavioural components. Finally, these models also differ in terms of the degree of influence that culture has on emotion perception. Ekman considers cultural differences in facial emotion expression and recognition as mostly impacting the intensity of the emotional expression or originating a rapid succession of qualitatively different emotions. However, some authors attribute a more prominent role to culture. Elfenbein, in his Dialect Theory, suggests that subtle variations may shape the very expressive elements of an emotion, characterizing cultures in different ways. According to Elfenbein’s theory, while Basic Emotions are still considered universal, different cultures may have their own “emotional dialects” (Elfenbein et al., 2007; Matsumoto, 2006). These cultural variations are mostly subtle and enable a successful cross-cultural communication yet may give rise to potential miscommunication (Elfenbein & Ambady, 2002, 2003).

Despite the various differences that characterize each theory, basic emotion theories adopt a categorical approach to understand emotions, agreeing on dividing emotions into discrete categories that may either be universal and basic or socially constructed. Due to the evolutionary ontology of basic emotions, each category evokes a unique and specific pattern of physiological responses, behavioural expression, and subjective experience. Thus, basic emotions should have identifiable antecedents, neural networks, physiological responses, and behavioural outputs. These assumptions often promoted the search for a one-to-one correspondence between prototypical expression - emotion category – and its underlying brain activity.

1.2.2 Emotional dimensions as building blocks of emotional experience

To comprehensively understand emotional experience, it is essential to define the fundamental elements of its ontology. The categorical approach tried to reduce the complexity of emotion by searching for the primary distinct categories that serve as the foundational building blocks for understanding the diverse range of human emotional experiences. Emotions such as Happiness, Fear, and Anger have been considered as irreducible entities that formed the basis of research models and theories in this field. Dimensional approaches have arisen as a viable alternative to categorical theories, highlighting the blended, multidimensional nature of emotions. Within this framework, emotions are defined as complex and dynamic entities that can be experienced simultaneously, without discrete borders that clearly differentiate them from one another (Russell & Fehr, 1994). Categories of emotions are explained as configurations of fundamental elements, (i.e., dimensions). Identifying these underlying dimensions, going beyond the prototypes of emotion categories, is essential to adequately describe human emotional experience (Russell, 2003).

Woodworth (1938) pioneered to investigate emotional experience through a dimensional perspective. He used participants' error distributions to conceptualize the relationship among different emotions. As a result, he ordered emotions categories along a one-dimensional space, with a scale ranging from 1 (happiness) to 6 (contempt), where closer categories represent those more commonly confused to each other (Woodworth, 1938). Schlosberg (1941) revised Woodworth's linear two-poles scale into a circumplex model where emotion categories were arranged along a circular structure and any emotion could be represented as a linear combination of different dimensions. He proposed two dimensions to define emotions: (1) "pleasant/unpleasant", which refers to the positive or negative nature of an emotion, and (2) "attention/rejection", which refers to subject's focus of attention when experiencing the emotion (Schlosberg, 1952). He later integrated this model with a third dimension: "sleepy/tense", which refers to the arousal level of the emotion (Schlosberg, 1954).

Several other dimensional models have been proposed, many of which closely resembling Schlosberg's original model, while others suggesting the need for additional dimensions. For instance, Osgood (1966) reported evidence for three major dimensions labelled as Pleasantness, Activation, and Control. In Frijda's and Philipszoon's study (1963) ratings of facial expression on 22 bipolar scales were intercorrelated and subjected to factorial analysis resulting in 4 factors: "Pleasantness-Unpleasantness", "Control of expression-Intensity of expression", "Attentional Activity-Disinterest", which align with Schlosberg's three dimensions, and "Naturalness and Submission-Artificiality and Condescension". Other works using multidimensional scaling based on similarity judgments suggested that Schlosberg's dimension "attention-rejection" seemed to be superfluous, supporting the view of a two-dimensional framework (Abelson & Sermat, 1962; Shepard, 1962). Despite the diverse terminologies adopted to describe its dimensions and the diverse emotional experience investigated, the two-dimensional structure has consistently emerged in numerous studies (Lang et al., 1998; Russell, 1980; Thayer, 1989; Watson et al., 1999), and it remains the most commonly used model in dimensional research on emotion perception.

Among various dimension models, Russell's (1980) two-dimensional Circumplex Model, has had a profound impact on the field. Through extensive studies adopting multidimensional scaling and factor analysis of self-reports of affective states (Feldman Barrett & Russell, 1998; Russell, 1980), emotion-denoting words (Russell, 1980) and similarity ratings of facial expressions (Russell & Bullock, 1985), Russell revealed two independent bipolar dimensions underling emotion experience: *valence* and *arousal*. The former varies from positive to negative emotions, whereas the latter ranges from low to high levels of physiological activation. The two dimensions are uncorrelated and emotions at opposite sides of the same dimension do not co-occur, indicating that blended emotions cannot be formed by feelings that drastically differ from each other (e.g., happy, sad). Within this framework, emotional experience is defined by two components: the Core Affect and the Prototypical Emotion Episodes. The Core affect refers to the basic, immediate, and non-specific experience of valence and arousal that underlies all emotions. It is the most fundamental and elemental aspect of emotions, and is believed to be biologically based, arising from the interaction between an individual's physiological state and their cognitive appraisal of the environment. To support the evolutionary origin of states of valence and arousal, Russell reported evidence of how these two dimensions were pan-cultural (Russell, 1991), presented in children (Russell & Bullock, 1985), and relied on linguistic studies showing how emotions are constantly described in terms of valence and arousal in different languages around the world (e.g., sentences like "I feel bad" and "I feel good" are presented in all languages) (Russell, 1991; Wierzbicka, 1999). On the other hand, the Prototypical Emotion Episodes refer to specific, discrete, and recognizable emotional experiences that are learned and shaped by social and cultural factors. These are the commonly recognized emotions, which are typically accompanied by specific behavioural, physiological, and cognitive patterns (Russell, 2003; Russell & Barrett, 1999).

In summary, dimensional approaches to the study of facial emotions aim to understand emotions based on their underlying dimensions, which are considered as the fundamental building blocks of our emotional experience. Researchers in this field seek to identify the smallest number of dimensions that can represent all emotions. Although conceptualized in different ways (Lang et al., 1998; Russell, 1980; Thayer, 1989; Watson et al., 1999), two main dimensions have been constantly revealed. In recent models, and mostly in line with

Russell works, these dimensions are often identified in terms of *arousal* and *valence*. Thus, discrete emotions such "fear" and "happy", emerge from the interaction of these underlying dimensions and the cognitive appraisals of the self and the environment. By adopting a dimensional approach, researchers aimed to move away from rigid and fixed categories of emotions and instead promote a more nuanced understanding of emotions as multidimensional and dynamic experiences that arise from the interaction of underlying dimensions and cognitive appraisals.

1.3 The rise of new approaches and findings challenge the study of facial emotions

1.3.1 Critiques to the standard methods used to investigate emotion perception

While the original assumptions of Ekman's BET are vastly influential (Ekman et al., 1969; Ekman & Friesen, 1971), the theory has also been the subject of extensive critique. Researchers have raised questions about the methodologies employed, and the ecological validity of the stimuli used, with some research denouncing theory-driven biases (Aviezer et al., 2012; Crivelli & Fridlund, 2019; Gendron et al., 2018; Jack et al., 2012; Nelson & Russell, 2013; Russell, 1994). Nonetheless, the approaches used in Ekman's early studies (Ekman et al., 1969; Ekman & Friesen, 1971) have become the "standard methods" (Nelson & Russell, 2013), for much subsequent research in this field. The same database of photographs adopted and created by Ekman is still a widely recognized set of stimuli in emotion research, extensively employed in studies investigating emotion recognition on children, individuals with autism, primates, and brain's activity (Parr et al., 2007; Tracy & Randles, 2011; Walle et al., 2017; Whalen et al., 2013).

Critics of the standard methods argue that these mainly rely on the few identified basic emotions, static stimuli, prototypical posed facial expressions, forced-choice response format, and within-subjects designs. Within this framework, above chance performance on recognizing facial emotions across cultures is often interpreted as evidence of the universality of basic emotions. However, constraints of the standard tasks may influence participants' performance and artificially inflate matching scores between emotion label and facial expression, leading to a false impression of universality (Nelson & Russell, 2013). For example, the forced-choice paradigm may induce participants to believe that any emotion category not listed is not a viable option, which can influence their interpretations and steer them towards the category specified by the experimenter. When presented with multiple facial expressions, participants often make their judgments by comparing the target expression to the ones previously observed (Russell, 1991), or rejecting the options already selected in prior trials (DiGirolamo & Russell, 2017; Russell, 1994). Moreover, using chance performance as a threshold may be too low to support the universality thesis, and not sensitive enough to detect specific patterns of differences in recognition performance across groups (Jack et al., 2016). Nelson and Russell (2013) reviewed studies that use standard tasks to support the universality thesis. They found that while matching scores for happiness were particularly high 89.6%, the percentage dropped drastically for negative emotions (e.g., sadness, anger, fear, and disgust), especially for non-Western participants who scored 57% when literate, and 39% when non-literate. Matching scores seemed lower than expected, as Haidt and Keltner (Haidt & Keltner, 1999) predicted that a 70% to 90% agreement on the emotion labels is needed to support the BET. According to Nelson & Russell, while we can conclude that humans do not answer randomly when asked to categorize a facial expression, evidence are not strong enough to support a thesis of universality. Alternative hypotheses cannot be fully ruled out, as other information, related to valence and arousal or social messages (Russell, 2003; Yik et al., 2013), may have been used by participants to deduct the emotion category. Regarding the ecological validity of the stimuli used in standard methodologies, it has been a concern that a limited set of static, posed, prototypical facial expressions may not capture the full range of human emotional experience, being mostly distant from the rich variations experienced in real life. Research has shown that there are significant morphological and dynamic differences between spontaneous and posed facial expressions (Buck & Arthur Vanlear, 2002; Namba et al., 2017 Delannoy & McDonald, 2009; Smith et al., 1986; Valstar et al., 2007; Park et al., 2020), and that participants' evaluation of morphological aspects of facial emotion can differ depending on whether facial expressions are spontaneous or posed (Johnston et al., 2010). Furthermore, facial expressions used as stimuli are frequently performed by trained actors who, aiming at conveying messages to a wide audience, tend to exaggerate their expressions by intentionally and strategically manipulating their facial movements into more artificial expressions (Buck & Arthur Vanlear, 2002; Carroll & Russell, 1997).

To overcome these limitations studies started to adopt less constrained designs, and discovery-based tasks (e.g., free labelling, cue-cue matching). These works often revealed both similarities and differences in the categorization of emotions across cultures (for a review, Gendron et al., 2018). For example, when members of a small-scale indigenous society in Paupua Guinea were shown with spontaneous facial expressions produced by members of another small-scale indigenous society, they rarely agree with Ekman's predicted labels. Completing a forced choice and free label tasks, accuracy scores ranged from 13% to 38%, for the forced choice task, and from 0% to 16% for the free labelling task. Nonetheless, participants scores given to dimensions of valence and arousal were largely in line with those obtained with Westerners' people (Crivelli et al., 2017). The emergence of conflicting results compared to previous research seemed to confirm that commonly used methods may fail in capturing the complexity of emotion processing. Cowen, Sauter et al. (2019), argued that traditional low-dimensional models, focusing on limited number of emotions and prototypical stimuli, are only able to account for approximately 30% of the systematic variability in emotional experience. Similarly, Snoek et al. (2023), showed how models based on AUs, are not able to decode all the relevant information used to infer emotion from faces, and cannot explain all the variance of emotion categorization behaviour.

Moreover, traditional discrete-category and dimensional frameworks both suggest that the processing of affective information, categories or dimensions, is relatively unaffected by the context. However, research has demonstrated that the same prototypical facial expressions can be labelled differently depending on the expressive behaviour simultaneously occurring in other modalities, such as posture or gesture (Aviezer et al., 2008, 2012), and that perception of facial expression can be influenced by the context (Greenaway et al., 2018; Wieser & Brosch, 2012) and the conceptual knowledge of the perceiver (Brooks & Freeman, 2018). Highlighting how to get a comprehensively understanding of human emotional experience, theories and models should also consider other information that may influence how we perceive facial emotions.

1.3.2 Moving toward complexity: the emergence of new approaches and evolving theories

To gain insights into the nature of emotions, as well as cultural and individual variabilities in processing facial expressions, recent studies employed data-driven and multi-disciplinary approaches by using a vast array of stimuli, collecting large-scale behavioural data, moving beyond a one-to-one mapping along a few emotion categories or dimensions, and aiming for a more sophisticated understanding of the complex process in emotion perception (Cowen & Keltner, 2018; Jack et al., 2018; Jack & Schyns, 2017; Keltner et al., 2023).

As a result, recent works conducted along the basic emotion framework moved largely beyond the prototypical facial expressions of the six basic emotions, revealing emotional experience to be *multimodal*, *richer*, and *nuanced* (Cowen & Keltner, 2017; Keltner, Tracy, et al., 2019). One interesting example of a theory taking such an evolving view on emotion perception is Cowen and Keltner's Semantic Space Theory (Cowen & Keltner, 2020; Keltner et al., 2023). According to this theory, latent dimensions of the emotional experience can be derived quantitatively from behavioural data by identifying patterned responses that occur systematically within an emotion modality. Thus, emotions can be organized into semantic spaces that can vary across individuals and cultures depending on their dimensionality, distribution and conceptualization. The dimensionality defines the number of distinct continuous kinds of emotions that organize a semantic space, the distribution determines the geometric arrangement of these emotions and their boundaries, and, finally, the conceptualization define whether emotions are better described by categories or broader appraisals. In particular, Cowen's and Keltner's works examining participants' judgments in terms of emotion categories, appraisals, free response, and ecological validity, on 1500 photographs of facial-bodily expressions, found how emotion perception is high-dimensional and much *richer* and *nuanced* than previously thought (Cowen & Keltner, 2020). It was possible to identify up to 28 emotion that were reliably characterized by distinct facial-bodily expressions, much more than the six basic emotions. Moreover, while emotional experience was better conceptualized in terms of categories more than appraisals (i.e., valence and arousal), contrary to the BET, boundaries between emotions categories seemed smooth and blended rather than discrete (Cowen & Keltner, 2020). Adopting the same methodology, similar results were also obtained analysing judgments of 2,050 emotional vocal bursts. In this case, 24 distinct emotions were identified and, again, boundaries between categories were bridged by smooth gradients of meaning (Cowen et al., 2019). While previous studies often focused on the role played by facial expressions, emotional information conveyed through other modalities such as vocalization, bodily movement, posture, tactile contact, gaze, and autonomic responses starts to receive more research attention (Cordaro et al., 2016; Dael et al., 2012; Graham & LaBar, 2007; Grewe et al., 2009; Hertenstein et al., 2009; Laukka et al., 2014; Lobmaier et al., 2008). Showing how the same facial expression can be interpreted differently depending on the posture or gesture simultaneously occurring (Aviezer et al., 2008), or how intense positive and negative emotions are mostly discriminated based on information communicated by the body rather than the face (Aviezer et al., 2012). As a result, facial expressions are increasingly treated as a part of a *multimodal* dynamic patterns of emotional behaviour.

By embracing the complexity of the emotional experience, new studies have started to more closely investigate the within-category variations in the expression and perception of emotions, especially across cultures, moving away from the traditional dichotomy that oversimplifies emotional experience as being solely shaped by evolutionary factors or constructed by culture. To explain the differences in the expression and recognition of posed facial expressions often found in cross-cultural investigation, Elfenbein et al. (2007) proposed the Dialect Theory of communicating emotion. The theory suggests that, due to social norms and experiences, different cultures develop unique "dialects" of facial expressions, just as they do with language. In line with Elfenbein's theory, Cordaro et al. (2018) conducted a cross-cultural study to examine the extent to which emotional expressions are universal or culturally specific. Participants from five different cultures were presented with 22 emotion concepts and associated story. Participants were asked to generate a felt experience of the emotion before expressing this nonverbally. Results showed that while about the 50% of the behavior produced in response to each emotion could be classified as culturally universal, variations unique to individual culture were also identified.

Finally, the introduction of innovative methods has also given rise to a novel common ground where traditional dimensional and categorical approaches could be effectively compared. Supporters of both theories started to acknowledge the value of new integrative models where both, dimensions and categories, play a significant role in explaining emotional experience, and where factors such as culture (Cordaro et al., 2018; Gendron et al., 2020), prior beliefs (Brooks et al., 2019; Brooks & Freeman, 2018) and contextual information (Greenaway et al., 2018; Wieser & Brosch, 2012) may contribute to variances across individuals (Snoek et al., 2023a). Liu et al. (2022) recently employed a data-driven, perception-based methodology, and a computer-graphics based generative model of human facial movements, to agnostically generate facial expressions (i.e., random combinations of individual AUs) and model the specific components that elicit participants' perception of categories (i.e., basic and complex emotions), and dimensions (i.e., valence and arousal). Their results showed that a latent set of shared facial movements could jointly determine the perception of emotion categories and dimensions, indicating a common signalling basis. Furthermore, while facial signals of dimensional information could predict specific emotion categories, the opposite was not true, suggesting that emotion perception is supported by underlying signals of broad dimensions. Following a similar procedure, using reverse correlation combined with a dynamic facial expression generator, Jack et al. (2016) modelled the facial expressions of over 60 emotions across 2 cultures. They then used multivariate data reduction techniques to identify the latent patterns that are common across cultures and those that are culture-specific. Their result identified both AU patterns that are common across participants, and culture-specific "accent". Interestingly, each AU, based on participants' ratings along dimensions of valence, arousal and dominance, represented a fundamental social message (e.g., negative, high arousal). They proposed a theoretical framework which stands between the traditional discrete categorical approaches (Ekman et al., 1969), and the continuous space of the dimensions of valence, arousal and dominance (Russell, 1980). According to this framework, four basic emotion categories, conveying fundamental meanings, are shared across cultures. The message transmitted by the base AU is than accentuated by additional AUs, cultural accents, that refine the message and are influenced by culture. Base AU would be clustered around specific values in a three-dimensional continuous space defined by the dimensions of valence, arousal and dominance, and the variance would be attributed to the modulation due to cultural accents.

In sum, new research approaches and findings offer novel insights into the ontology of facial expression communication. Emerging theories go beyond traditional dichotomies of dimensions/categories, universal/culture-specific by embracing a more complex understanding of the emotional behavior. Emotions are now investigated as complex multimodal entities influenced by a multitude of interacting factors such as biology, culture, social contexts, and individual experience.

1.4 The present project

1.4.1 Investigating facial emotion perception in contemporary research

As reviewed in previous sections, researchers have sought to characterize emotional experiences by identifying elementary emotion categories (e.g., BET) or dimensions (e.g., Affective valence and arousal). The identification of specific affective dimensions and emotion categories captures some relevant aspects of our emotional experience and establishes common hypotheses and methodologies for investigating facial emotion processing. Nonetheless, the emergence of new methods, large-scale data, and data-driven approaches has consistently highlighted their limited power in explaining the richness and complexity of real-world emotion experiences (Cowen, Sauter et al., 2019; Snoek et al., 2023). In daily life, human emotions are rarely experienced as pure single entities, convey much more information than a single emotion label, and are significantly different from the posed, stereotypical facial expressions often used in research. Moreover, face perception occurs in contexts where multiple processes interact and influence each other (Greenaway et al., 2018; Wieser & Brosch, 2012; Brooks & Freeman, 2018). To fully comprehend how facial emotion processing occurs in real life, it is essential to reintroduce complexity into the research paradigm, exploring naturalistic stimuli, modulating factors such as contextual cues and culture, and collecting rich behavioural responses. Facial emotion models formulated in research have wide-ranging applications in various aspects of our lives and their ability to precisely capture the intricacies of our emotional experience, as it unfolds in real-world situations, it is of crucial importance. Nowadays, machine learning techniques are designed with the intent to classify combinations of facial movements into a defined set of categories (Altameem & Altameem, 2020; Fei et al. 2020; Kaushik et al., 2022). At the same way, tools intended to improve and promote human ability to read faces, such as adults training programs, games and prompts created to accompany children development, are created in accordance with the most accredited theories (Holmes, 2011; Margoudi et al., 2016; Payton et al., 2000). By aiming a high-dimensional models, research can bridge the gap with real-world emotional experiences, offering more comprehensive understanding of our ability to extract rich social meanings from facial expressions of emotions.

Toward dynamic and more natural stimuli

Even though the sensitivity to the dynamic nature of facial emotions is developed from early infancy (Braddick & Atkinson, 2011), and previous research has demonstrated that perception of facial emotion is enhanced when incorporating face stimuli (e.g., point-light, line drawings, schematic and computer-animated faces) with biological motion (for a review see Krumhuber et al., 2013), the majority of studies on facial emotions have primarily relied on static facial expressions (Krumhuber et al., 2013). Newborns exhibit a limited ability to accurately perceive shape and texture, but they demonstrate a remarkable sensitivity to biological motion (Fox & McDaniel, 1982; Sifre et al., 2018; Simion et al., 2008). They also show an innate preference for faces, as evidenced by their inclination to focus on and explore facial stimuli (Frank et al., 2009; Johnson et al., 1991; Nelson, 2001), and facial motion serves as a key mechanism through which they learn to understand and interpret emotions, and social cues (Carnevali et al., 2022; Farroni et al., 2007; Xiao et al., 2015). Infants as old as 2-days can already successfully discriminate between a happy or disgusted dynamic faces (Addabbo et al., 2018). Also in adulthood, motion offers a distinct advantage in the processing of faces, particularly when static information is limited, inadequate or unavailable (for a review see Krumhuber et al., 2013), and this advantage seems not to be attributed to an increased amount of static information. Ambadar, Schooler, and Cohn (2005), showed that the identification of subtle expressions was significantly superior when presented in moving sequences compared to "multistatic" images. Dynamic sequences appeared to provide functionally distinct information that could not be solely attributed to additional static cues. Similarly, Jack et al., (2014), showed that the dynamics of facial expressions transmit an evolving hierarchy of information through time. Initially, the expressions convey biologically rooted cues that facilitate the categorization of fundamental categories (e.g., approach/avoidance); subsequently, more complex signals emerge aiding in the categorization of a broader array of socially specific categories (e.g., emotion categories). Mortillaro et al. (2011) also showed that facial expressions can convey different emotional meanings (i.e., interest, pride, pleasure, and joy) depending on the duration or frequency of specific action units.

Similarly, while previous research has been relying on posed facial emotions, these have been shown to be significantly different in their dynamics and morphology from natural real-world expressions (Delannoy & McDonald, 2009a; Namba et al., 2017; Valstar et al., 2007). Furthermore, as often performed by professional actors, posed expression tend to be even more exaggerated and artificial (Buck & Arthur Vanlear, 2002; Carroll & Russell, 1997). Namba et al. (2017) compared spontaneous and posed facial expressions associated with surprise, amusement, disgust, and sadness. Spontaneous expressions were captured as participants were watching emotion-inducing videos, while posed expressions were recorded by instructing participants to intentionally display each emotion. Analysing outputs through the Facial Action Coding System (FACS), results revealed distinct dynamic patterns between the two types of expressions. In line with these results, other studies found differences in the temporal patterns of facial movements between spontaneous and posed facial expressions (Delannoy & McDonald, 2009a; Valstar et al., 2007). Park et al. (2020) analysing differences between posed and spontaneous smiles using three-dimensional facial landmarks and advanced machine analysis, found that spontaneous smiles exhibited higher intensities in the upper face while posed smiles showed higher intensities in the lower part of the face. The analysis also revealed that the left eyebrow displayed stronger intensity during spontaneous smiles compared to the right eyebrow. Importantly, differences between spontaneous and posed smiles have also been shown to be perceptible to observers. Individuals displaying spontaneous smiles of enjoyment were evaluated more positively than those displaying posed smiles, suggesting that the authenticity and naturalness of facial expressions significantly influenced how people perceived and responded to them (Johnston et al., 2010).

Nonetheless, in most of the available databases used in emotion perception research, static posed expressions have been preferred over spontaneous dynamic facial emotions. Facial emotion databases have often been created based on categorical assumptions (for a review see Krumhuber et al., 2017), stimuli were divided into mutually exclusive categories, often the 6 basic emotions (Ekman, 1992), and a one-to-one correspondence between emotional experience and prototypical patterns of facial was assumed (Ekman & Friesen, 1978). While these stimuli offer excellent experimental control, they may fail in representing real-world spontaneous emotional experience.

Conversely, none-posed dynamic facial emotions, while allowing for more fine-grained and natural expressions, are blended, variable and challenging to control. Dobs et al. (2018) conducted a comprehensive review assessing the effectiveness of various stimulus types incorporating dynamic information in face perception studies. The choice of stimuli enables researchers to address varying levels of naturalness, control over facial form (i.e. facial features and their configuration) and/or motion, and different levels of technical demand. For instance, while natural videos of faces are the most ecological stimuli in terms of facial form and motion and require a low technical demand, they provide limited control on both form and motion. On the other side, point-light faces are highly unnatural, require some technical demand, but offer a good control on facial motion. Image-based morphing and Synthetic facial animations, represent intermediate options in terms of naturalness on form and motion while offering different levels of experimental control. Finally, recent advances in photorealistic face rendering provide a way to achieve control over facial motion while preserving the natural facial form. In conclusion, the facial perception system exhibits a heightened sensitivity to natural facial motion, suggesting the importance of adopting dynamic face stimuli. However, depending on the specific aspect under investigation, the choice of stimuli may differ considering that each type of stimulus possesses unique advantages and disadvantages. The selection of appropriate stimuli remains a significant challenge in facial emotion research, requiring a delicate balance between ecological facial expressions and experimental control.

Toward high-dimensional frameworks and collection of rich behavioural responses

Every-day emotions are fluid and evolving phenomena that do not neatly fit into one predefined box or label; they blend and transition, making it challenging to capture their entirety through a single emotion label or a limited number of dimensions (Keltner, Tracy, et al., 2019). Even though still little is known about the perception of complex blends of emotion (Moore & Martin, 2022), researchers using FACS analysis showed that 21 distinct compound emotion categories can be visually discriminated by computational model (Du et al., 2014). Similarly, analysing naturalistic facial-bodily expressions have been identified 28 distinct categories of emotion that are bridged by smooth gradients of meaning (Cowen & Keltner, 2020). English speakers, for instance, can clearly distinguish dozens of emotional states such as: contempt, shame, pain, sympathy, love, lust, gratitude, relief, triumph, awe, and amusement (Cowen et al., 2019). Whether we approach emotions by using single categories or few affect dimensions (i.e., valence, arousal), we inevitably end up simplifying the complexity and richness of human emotional experiences, loosing precious information. To understand natural facial emotion perception, we need to acknowledge its complexity adopting research paradigms that gather rich behavioural responses rather than relying solely on single, discrete emotion labels.

A holistic model of emotion processing should also account for the impact that none-emotional factors may have on the perception of facial emotion. Greenway et al. (2018) identify three factors that, operating at different levels, influence how emotion is experienced, expressed, perceived, and regulated. At the personal-level, people have internal constructs including demographics, personality, and stimulus appraisals. At the situational-level, people have the characteristics of the immediate environment and the relationship with it. Finally, at the cultural-level, people have the socio-cultural background of the expresser and perceiver. For instance, more individualistic cultures tend to have more lenient emotional display rules, allowing individuals to express a wider range of emotions. In contrast, collectivistic cultures often have stricter emotional display rules, emphasising the suppression of negative emotions (Matsumoto, Yoo, et al., 2008). As a result, even though perceivers from different cultures may agree on the primary emotion that is conveyed by a facial expression, they may differ in how they perceive concurrent emotions from the same facial expression (Fang et al., 2018). Previous findings suggest that East Asians tend to experience multiple different emotions concurrently, while North Americans and Europeans are likely to experience specific feelings (Grossmann et al., 2016; Miyamoto et al., 2010). To sum up, even though contextual information such as culture, environment and person characteristics, often plays a key role in shaping the way we express and perceive emotions (Barrett, 2012; Greenaway et al., 2018; Lindquist & Barrett, 2008; Matsumoto & Hwang, 2013; Wieser & Brosch, 2012), emotional experience has often been extracted from its context (i.e., cultural background, specific environment, personal characteristics). In order to develop holistic account of emotion perception, it is necessary to adopt sensitive methods that can capture the rich human emotional responses and the factors that may interact with it.

1.4.2 The aim of the project

The best approach to conceptualising and/or measuring emotion perception and understanding is a contentious question for contemporary researchers (Snoek et al., 2023). Theories so far have mostly explained the way we process and decode emotions from facial expressions by hypothesizing a categorization process (e.g., BET). Meaning is attributed by classifying the perceived facial expression into one of different possible basic discrete categories (Ekman, 1992; Tracy, 2014; Matsumoto et al., 2008 Panksepp & Watt, 2011), or it is the result of the perception of different basic dimensions (e.g., Affective valence and arousal) (Russell, 2003) which combined are interpreted as - again - a specific discrete category. Many now challenge the prevailing view that emotions are discrete constructs that can be meaningfully described with a single semantic label. Instead, they are seen to be multidimensional, dynamic, blended, and potentially varying along a gradient (Barrett & Satpute, 2019; Cowen & Keltner, 2020). More recent theories, with the aim to be able to explain the complex emotional experience that characterize our daily life, identify up to 28 distinct categories of basic emotions and consider emotions as being organized into semantic spaces, where the boundaries between categories are smooth and blended rather than discrete (Cowen & Keltner, 2020) revealing a more *rich and nuanced* processing. Critically, however, methods advances in the field have not kept pace with theory: we lack measures sensitive enough to capture these highly complex constructs. Most of the tasks commonly used to collect behavioural responses still reflect the assumptions of classical theories of emotions, potentially neglecting precious information, asking participants to sort, categorize or match facial emotion with one category.

To address the current knowledge gap in methodologies, my PhD embraced the challenge of exploring a novel nuanced approach which I validated with respect to established behavioural and computational benchmarks. In particular, extensive research has been conducted on the process of discerning and differentiating facial emotions, mostly relying on categorical interpretations and adopting methods that tend to minimize the distance between emotion of the same category and maximize the distance between emotion of different categories (e.g., prototypical stimuli, forced-choice response format, attributing one-label only). In this project, I developed tasks that are more sensitive to the rich and multidimensional emotional, semantic, and contextual information conveyed by faces. Going beyond a one-to-one correspondence between facial expression and single labels, and investigating how different sources of information (i.e., different measures of stimuli similarities) may explain the way we perceive similarities between facial expressions will help us understand how the same emotion can be expressed via a variety of within-category expressions (Sauter, 2010; Shiota et al., 2017; Szameitat et al., 2009), and how similar expressions can acquire different meanings depending on the expressive behaviour simultaneously occurring (Aviezer et al., 2008), the context in which they occur (Greenaway et al., 2018; Wieser & Brosch, 2012), as well as the cultural background of both the perceiver and the person exhibiting the expressions (Elfenbein, 2013; Elfenbein et al., 2007).

Differently from previous works, I asked participants to define stimuli along multiple dimensions, instead of a single emotion category, and I translated responses into detailed, multidimensional profiles for each stimulus and participant. Treating profiles as vectors, I obtained different measures of stimuli similarity and saw how participants perceived similarity might be explained by stimuli high-dimensional representations and other relevant factors such as their physical image-based similarity. Then, I considered the possibility that stimuli high-dimensional representations and perceived similarity may be influenced by participants culture, the context in which the stimulus is presented, and the presence of facial dynamic cues. In particular, while the influence of culture has been largely investigated in previous research, results may have been limited by the adoption of categorical approaches (Fang et al., 2018). The more sensitive tasks introduced in this project highlight how, even though perceivers from different cultures may agree on the primary emotion that is conveyed by a facial expression, they may differ in how they perceive concurrent emotions from the same face giving rise to a qualitatively different emotional experience. Finally, with the aim to explore the potential of contemporary computational models, based on categorical assumptions, in replicating the profiling performance of humans in facial emotion recognition, I tasked a machine learning to produce profiling responses by deriving posterior probabilities of how the classifier would respond to each of the eight learned emotional dimensions.

In conclusion, this approach allows a more fine-grained perspective on how rich information is extracted from faces, and how this information may differ, overlap, and dynamically change depending on factors such as context, culture, the presence of motion cues - going beyond the basic 8-emotions categorical approach.

Specifically, I conducted six behavioural and cross-cultural studies using high-dimensional profiling tasks and Representational Similarity Analysis with the main goal to (1) investigate whether complex, high-dimensional facial emotions are embedded in facial expressions; (2) determine whether high-dimensional representations of facial emotions outperform categorical models in predicting perceptual similarities between different facial expressions of emotion; (3) uncover what factors underlie perceptual similarity between facial expressions of emotions; (4) test the impact of culture, emotion intensity, facial motion, and context on perception of facial emotions; and (5) explore how human facial emotion processing may differ from machine learning approaches to emotion perception.

With the intent to move beyond databases commonly used in emotion perception research, I specifically selected stimuli from the MPI Facial Expression Database (Kaulard et al., 2012). Unlike the most common used databases of facial expressions, in this dataset fit for the purposes of our work very well. In particular, (1) facial expressions are available in both format, videos and pictures; (2) the database contains a large variety of natural emotional expression aside from the 6 basic ones; (3) within-category variations of emotional expression are available; (4) it balances between ecological facial expressions and experimental control by recording videos using an acting protocol, in particular, no-professional actors are instructed to mimic the expression elicited by a given every-day scenario (e.g., “You have reached a goal and you are happy”).

To assess participants' ability to extract rich complex information from facial expressions of emotions, I collected high-dimensional behavioural responses through the Profiling task. The Profiling task, asked participants to describe their perception along multiple dimensions, generating unique profiles representing each facial emotion under different conditions. To assess whether participants perceived facial expressions of emotions as a single, discrete emotion category (i.e., dominant response for the target dimension, accompanied by near-zero scores for other dimensions) or as a profile of mixed emotions, (i.e., multiple prominent dimensions), I planned to run t-tests comparing target and no-target dimensions of the same profile, expecting to find multiple no-target dimensions being no significantly different from the target dimension. As comparisons were planned, were deemed appropriate and justifiable without the need for adjustment for multiple comparisons. Then, I trained a Random Forest machine learning to classify the observed emotion profiles into the three emotions and applied a feature importance algorithm to determine the relative importance of the different dimensions (i.e., one relevant dimension or more?).

To investigate whether factors prevalent in our daily experience may influence the way we perceive facial expressions of emotion, response profiles were compared across different cultural backgrounds (Chinese vs. British participants), facial motion (Static vs. Dynamic), emotion intensity (High vs. Low), and emotional contexts (Physical vs. Social scenarios). To do so, profiles generated in response to the three facial emotions where contrasted by averaging responses across all factors except the one investigated. The obtained profiles for each facial emotion were then submitted to a 2 (British vs Chinese / Dynamic vs Static / High vs Low intensity) by 8 (emotion dimensions) ANOVA and t-tests were conducted.

To uncover the underlying mechanisms of perceptual similarity between facial emotions, I first performed a Representation Similarity Analysis (RSA) to test whether emotion profiles are similar across different conditions, then I constructed multiple regression models to test how different factors contribute to the perceptual similarity between facial emotions. In particular, the RSA quantifies similarities or dissimilarities between the responses to pairs of stimuli resulting in a Representational Dissimilarity Matrix (RDM). These RDMs can be then compared across different modalities, often through correlation analysis, providing crucial insights into how facial information is processed and represented (Cutzu & Edelman, 1998; Cutzu & Edelmant, 1996; Kiani et al., 2007). To investigate what determines perceptual similarity between facial emotions, from participants profiling responses I derived various types of stimuli similarity based on stimuli's emotional meaning (i.e., Emotion Similarity), emotional intensity (i.e., Intensity Similarity), and the social and physical contexts they may usually be associated with (i.e., Context similarity). I also obtained a physical image-based similarity between facial emotions (i.e., Physical Similarity), using the Gabor-jet model (Lades et al., 1993). It is relevant to note that while I originally planned to compare stimuli similarity between 4 conditions. Specifically: (1) *Same*, faces displaying the same emotion at the same level of intensity; (2) *Within-Within*, faces displaying the same emotion at different levels of intensity; (3) *Between-Within*, stimuli displaying different emotions at the same levels of intensity (4) *Between-Between*, faces displaying different emotions at different levels of intensity. Results from our first study guided me in adopting a more fine-grained approach that also consider the specific emotion category displayed by the facial expression, comparing responses between Happy, Fear, Pain facial expressions at two levels of intensity (High, low). A multilinear regression analysis was then conducted with the aim to assess how these various types of similarity indexes may contribute to participants' perceptual similarity. To determine whether high-dimensional representations of facial emotions outperform categorical models in predicting perceptual similarities between facial expressions of emotions, two different models of stimuli similarity based on their emotional meaning (i.e., Emotion Similarity) were generated. A Categorical model, computed from participants confusion scores obtained in a categorization task (i.e., matching an emotion label with a facial expression), and a Profiling model, computed from response profiles obtained in an emotion profiling task. These two models were compared to determine whether the information contained in profile responses could improve model performance in predicting perceptual similarity and, if that was the case, determine the extent of this improvement.

Finally, I also compared human performance with algorithm-based emotion perception based on a widely employed machine learning approach. I trained computational models to perform both the emotion categorization task and the emotion profiling task, and then tested whether humans and algorithms responded differently to facial expressions of emotions. The findings may be potentially helpful to bridge the gap between human and machine performance.

1.5 Overview of the thesis

In *Chapter 2* I present two behavioural studies investigating the role of Categorical, Physical and Intensity information in perception of dynamic and static facial emotions. In Study 1 participants were first asked to distinguish between facial emotions by categorizing static or dynamic stimuli into three different labels to measure the confusability between facial expressions based on their emotion categories. Then, they were asked to directly judge the perceptual similarity between pairs of stimuli. Results from both the categorization task and similarity rating task allowed me to investigate the influence of facial motion, categorical confusability, and physical similarity in perception of non-stereotypical, non-posed facial emotions. A multiple linear regression model was used to estimate the extent to which categorical and physical models of similarities contribute to explaining variations in participants' perceived similarities and whether the combination of physical and categorical information yielded a stronger explanation for participants' perceptual similarity compared to the contributions of individual models. In Study 2, participants were asked to rate emotional intensity of dynamic or static facial emotions before judging their similarity. Results allowed me to test how differences in emotion intensity, facial motion and physical similarity may contribute to perceptual similarity between facial expressions of emotions, and whether the combination of physical and intensity information yielded a better explanation for perceptual similarity compared to their individual contributions.

In *Chapter 3* I present two behavioural studies conducted on British and Chinese participants with the intent to capture the complex representation of facial emotions through an emotion profile task while investigating the influence of facial motion, emotion intensity and culture in the perception of facial emotions. In Study 3, participants first rated all dynamic or static facial emotions along 8 emotional dimensions, and then were asked to judge their similarity. I tested whether similarity based on Emotion Profiles can better explain perceptual similarity compared to that based on categorical responses generated from Studies 1 and 2. A multiple linear regression model was employed to estimate the extent to which Emotion Profiles models of similarity contribute to explaining variations in participants' perceived similarities compared to categorical models, and in combination with, physical and intensity models obtained from previous studies. Finally, a machine learning algorithm was trained to generate machine-based high-dimensional emotion profiles that have been compared to human performance. In Study 4, participants first rated all dynamic or static facial emotions along 6 semantic dimensions closely related to the target emotion category, and then were asked to judge their similarity. High-dimensional emotional profiles were first analysed with the intent to determine whether facial expressions of emotions are linked to more semantic concepts beyond the target emotion category, and then compared between different intensities, cultures, and stimulus type (e.g., dynamic, static).

In *Chapter 4,* I present two behavioural studies conducted on British and Chinese participants that investigate the role of contexts in the perception of facial expression of emotion. In Study 5, participants first rated all dynamic or static facial emotions along 6 different social or physical emotional scenarios by judging the likelihood of the stimuli being displayed in each scenario (i.e., context profiling task), then were asked to judge their similarity. The context rating task allowed me to establish a context profile for the facial emotion displayed, and the observed context profiles were used to test how emotion intensity, culture and facial motion may affect emotion perception. As for previous studies, participant ratings of perceptual similarity were used to investigate how context profiles correlate with perceptual similarity. In Study 6, participants first rated all dynamic or static facial emotions along 8 emotional dimensions (i.e., emotion profiling task), while facial emotions were presented in context, with some were congruent with the facial expressions and others were not. The context was adopted from Study 5. This design allowed me to investigate whether congruent context enhances emotion perception and how incongruent context affects emotion perception. I also tested whether emotions profiles obtained with context were influenced by facial motion, semantic intensity, and culture.

In *Chapter 5*, I summarize relevant findings, drawing conclusions on how a profiling approach may better capture the complex nature of facial expressions of emotion, I explain how we perceive differences and similarities between facial expressions, and explore the role played by facial motion, emotion intensity, culture and context in the process. Possible future directions are also considered.

Chapter 2

The role of Physical, Categorical, and Intensity information in the perceptual similarity of dynamic and static facial expression of emotions

2.1 Introduction

The perception of facial emotions is guided by various sources of information, which determine how we interpret of facial expressions as instances of emotion. Within similar facial expressions, we extract subtle differences that shape our understanding of the conveyed emotions. Conversely, quite different facial expressions can be interpreted as conveying the same underlying emotional meaning based on specific relevant cues. For example, even though facial expressions of happiness and pain convey very different emotional meanings, their physical expressions often involve similar configurations of facial muscles. This shared similarity makes it challenging to distinguish between intense facial displays of happiness or pain when these are presented without contextual information (Barrett et al., 2011). Previous research has primarily focused on unraveling the factors underlying our ability to categorize facial emotions, while little research has explored those that influence the way we perceive subtle differences and similarities between facial emotions. Nonetheless, by investigating the perceptual similarity between facial emotions, we could gain a fine-grained perspective on how we differentiate among facial expressions and identify the underlying information that shapes our representation of facial emotions.

Previous studies have shown that both physical cues and conceptual information influence performance on tasks requiring the perceptual matching and categorisation of facial expressions of emotion. In particular, perceivers' conceptual knowledge of emotional concepts dynamically interact with the processing of facial cues, and idiosyncratic differences in emotion concept knowledge can predict subtle differences in how emotions are perceived (Brooks & Freeman, 2018). Similarly, physical cues such as shape and textures of facial expressions influence the way we perceive facial emotions. In a study of Sormaz, Watson, et al. (2016), researchers employed Procrustes analysis to compute similarities between the shapes of pairs of emotional faces, while image-based correlations between pixel intensities were used to determine the similarities in their surface textures. These physically based properties of stimuli were found to predict subjective ratings of similarity between facial emotions. More recently, Murray et al. (Murray et al., 2021) investigated the roles of shape, surface, and conceptual information in the perception and categorisation of facial expressions. Their findings revealed that while conceptual similarity affects both the categorization and perception of facial emotions, the shape and texture of faces assume more crucial roles in emotion perception and categorization, respectively.

In this Chapter, I present two behavioural studies conducted with British participants, with the main objective of investigating how the similarity between emotion categories, emotion intensity and physical appearance of two facial emotions affect their perceived similarity. In particular, I focused on the role of Categorical, Physical and Intensity information in the perception of both static and dynamic facial expressions. Similar to emotion categories, the intensity of emotion (i.e., high vs low levels of physiological activations) has been widely investigated and considered as a basic dimension underling our emotional experience. Together with valence (i.e., positive vs. negative emotions), emotion intensity has been recognized as a fundamental, biologically based, aspect of emotions. In classic dimensional theories, such as Russell's (1980), emotion intensity is acknowledged as a biologically grounded dimension of emotional experience, thus likely to significantly influence how we perceive the similarities and differences between different facial emotions. Comparing results generated in response to static and dynamic stimuli we wanted to also test for any influence of facial motion on participants' performance. Previous research has highlighted how facial motion enhances performance in tasks related to emotion categorization and recognition. This advantage of motion becomes particularly apparent when the static information is limited, inadequate or unavailable (e.g., through degradation in geometry, shape, or texture) (for a review see Krumhuber et al., 2013). Exploring the perception of dynamic facial emotions also allows for the identification of potential limitations in prior research, which heavily relies on static facial emotions. Given that participants completed the study online, in their preferred time and environment, it was important to implement measures intended to mitigate possible distractions or disengagement. Limiting the study duration to one hour was one such measure. This constraint informed the decision to focus on no more than three distinct emotions for investigation. Specifically, I selected two well-investigated basic emotions: one characterized by a positive valence, 'Happy,' and one with a negative valence, 'Fear,' both of which are part of the seven basic emotions originally recognized by Ekman (Ekman, 1972). Additionally, I wanted to explore responses to a less investigated emotion, and I selected 'Pain,' which has recently been introduced into the list of 28 distinct basic emotion categories recognized by modern developments of the Basic Emotion Theory (BET) (Cowen & Keltner, 2020). The choice of investigating participants' responses to 'Happy' facial expression, was also guided by our interest in exploring whether, and how, the well-established "Happy advantage" would emerge in our studies. Due to their perceptual and categorical distinctiveness, Happy facial expressions are in fact recognized more quickly and accurately than other emotions (Calvo et al., 2012; Calvo & Beltrán, 2013; Leppänen & Hietanen, 2007). This phenomenon, termed the "Happy advantage", has consistently emerged in studies comparing recognition across different stimulus sets for the six basic emotion categories (Calvo & Lundqvist, 2008; Ekman & Friesen, 1976; Leppänen & Hietanen, 2004).

2.2 Study 1. How do physical and categorical similarities contribute to perceptual similarity of dynamic and static facial emotions?

Study 1 investigated the role of stimuli physical proprieties (Physical similarity) and emotion information (Categorical similarity) in the perception of similarities between dynamic and static facial emotions. Participants were first asked to categorize static or dynamic facial expressions of Happy, Fear and Pain using a forced choice task (i.e., a standard task used in BET). Then, they rated the perceptual similarity between pairs of facial emotions.

Adopting a representation similarity analysis, we computed Physical, Categorical and Perceptual models of similarity for static and dynamic stimuli in the form of dissimilarity matrices. Each cell of a matrix contained values representing the discrimination between the corresponding pair of emotions using a different source of information for each model. Specifically, categorization patterns of individual subjects have already been used in previous works to encode the similarity of emotions as reflected by the behavioural confusions between a facial expression and an emotion label (Skerry & Saxe, 2014). Similarly, we used confusability measures obtained from the categorization task to compute models of stimuli similarity based on their conveyed emotion category for each participant in response to both static and dynamic stimuli. Regarding physical similarity, in face perception research is particularly challenging to find quantitative metrics to precisely measure differences in perceived faces. One widely used approach is the implementation of the Gabor-jet model (Lades et al., 1993). This is a biologically inspired model that emulates the response of simple cells in the early visual cortex (V1), assuming V1 captures metric variation, and enabled us to compute a single value that represents the psychophysical similarity of two images. Prior works have shown how the Gabor similarity of facial stimuli correlate with perceptual similarity of facial identities (Yue et al., 2012), expressions (Xu & Biederman, 2010), and facial movements (Dobs et al., 2014) and it has been used to determine objective measures of facial expression similarity for both images (Lyons et al., 1998; Susskind et al., 2007) and videos (Dobs et al., 2014; Xu & Biederman, 2010). After nearly a quarter of a century since its development, the Gabor-jet model remains highly relevant and is widely used in modern neurocomputational models of vision (Margalit et al., 2016). We adopted Gabor dissimilarity measures to compute the physical dissimilarity matrices for both static and dynamic stimuli. Finally, to calculate perceptual similarity of pairs of facial expressions, previous works have often used subjective judgements on a seven-point scale (Said et al., 2010; Sormaz et al., 2016). We adopted the same procedure and then calculated two perceptual similarity matrices for each participant from the ratings given to pairs of static and dynamic stimuli. Finally, to obtain a measure of how strongly Physical and Categorical models relate to perceptual similarity for static and dynamic stimuli, these were correlated with their corresponding static or dynamic perceptual similarity for each participant. Then a multiple linear regression model was used to estimate the extent to which Categorical and Physical models of similarities contribute to explaining variations in participants' perceived similarities and whether the combination of physical and categorical information yielded a stronger explanation for participants' perceptual similarity compared to the contributions of individual models.

2.2.1 Methods

Participants

Given no strong precedent for calculating sample size, I defined my sample based on previous literature where a similar designed was implemented (Brooks & Freeman, 2018).
Eighty participants were recruited using the SONA System at the University of East Anglia (11 males, 69 females; age ranged between 18-33 yrs., M = 20.1, SD = 2.72). The participants sample included 59 British, 16 no-British European, 4 Asian, and 1 African. All participants were naïve to the purpose of the investigation, provided informed consent before taking part in the study, and were debriefed at the end, receiving course credits as compensation. The study's experimental procedure was approved by the Ethics Committee of the School of Psychology at the UEA.

Stimuli, Materials, and Tasks

Dynamic and static facial expression stimuli were created using the MPI Facial Expression Database (Kaulard et al., 2012). This database contains 55 distinct natural basic or subordinate expressions performed by non-professional actors who, following an acting protocol, were instructed to mimic the expression elicited by given scenarios. From the database I extracted the facial emotions of interest (i.e., 'Happy', 'Fear', 'Pain'), each including a high- and a low- intensity expression prompted by a total of 6 different scenarios (for details, see

Appendix A, Figure A.1). Facial emotions and associated scenarios were validated in Kaulard's study (2012).

I randomly picked 10 identities (5 females and 5 males) from the MPI database, obtaining a total of 60 videos (10 actors * 3 emotions * 2 intensities). Static stimuli were created by extracting the frame conveying the peak of the emotion from each video. It is important to note that while posed prototypical stimuli obtained through more traditional means (e.g. adopting morphing techniques, instructing participants on their muscle movements or on the specific emotion to display) often have a crescendo in the muscle contractions toward a peak, following the onset-apex-offset model, this may not be necessarily true for elicited non-stereotypical facial expressions. More natural expressions can in fact have multiple apexes, or may apex at a low intensity, following a more complex development (e.g., onset-apex-onset-apex-offset) (Delannoy & McDonald, 2009). Participants involved in the creation of the MPI database were free to express the facial expression elicited by a specific scenario without any specific constraint. Consequently, videos varied in length and exhibited a more complex, non-linear dynamic, making the identification of an emotional peak challenging (see Figure 1a for examples of frames extracted from raw videos). To address this issue, a pilot study was conducted to determine the peak of the expressions. Ten volunteers from the University of East Anglia were asked to categorize each video first, and then to select the frame they believed conveyed the emotional peak. Participants were allowed to re-watch and pause the video until a decision was reached. Results from the pilot study allowed us to cut the videos at the most commonly identified pinpointed emotional peak. The outcomes of the pilot study also prompted the exclusion of one identity, as participants encountered notable challenges in accurately identifying the intended emotions conveyed by videos featuring this specific actor. As a result, the finalized list of stimuli included a total of 54 static stimuli and 54 dynamic stimuli (9 actors * 3 emotions * 2 intensities). Before composing the final set of dynamic stimuli, Adobe Premier software was used to normalize videos. In the finalized videos, facial expressions were preceded by 5 frames of neutral expression, all faces were centered on the screen, the green markers on actors' head were covered with a black mask, and all videos were converted to grayscale (see Figure 1b).

Stimuli were converted to grayscale with the aim to specifically focus on the dynamic changes in face shape involved in different facial emotions. While dynamic changes are also indicated by shading cues, converting stimuli to grayscale results in loosing information regarding possible differences in skin pigmentation (e.g., blushing). Both shape and surface (i.e., pigmentations and shading) features of a face contribute to the perception of facial expression (Bruce & Young, 1998). However, changes in shape seem to play a relatively dominant role (Butler et al., 2008; Etcoff & Magee, 1992). Degrading surface information while leaving shape information intact has little impact on perceptual and neural responses to facial expression (Bruce & Young, 1998; Harris et al., 2012; Magnussen et al., 1994). In everyday life, facial color changes during emotions - faces flushes during anger or goes pale when experiencing fear - and recent studies have highlighted how facial color is an important signal in the perception of facial emotion, enhancing our ability to recognize expressions (Nakajima et al., 2017). However, the interplay between skin pigmentation and emotion recognition, while crucial for developing a comprehensive model of emotion recognition, remains less explored. Hence, my decision to focus specifically on shape information, which is well-established as a key factor in decoding facial emotions.

The videos durations consistently ranged from 1 to 2 seconds, ending once the identified peak of the emotion was reached). However, it is noteworthy that videos featuring the same actor exhibited uniform lengths, either approximately 1000ms or around 2000ms, without any additional manipulation. The final frame of each video was extracted to form the static stimuli (see Figure 1c for example of the finalized stimuli).

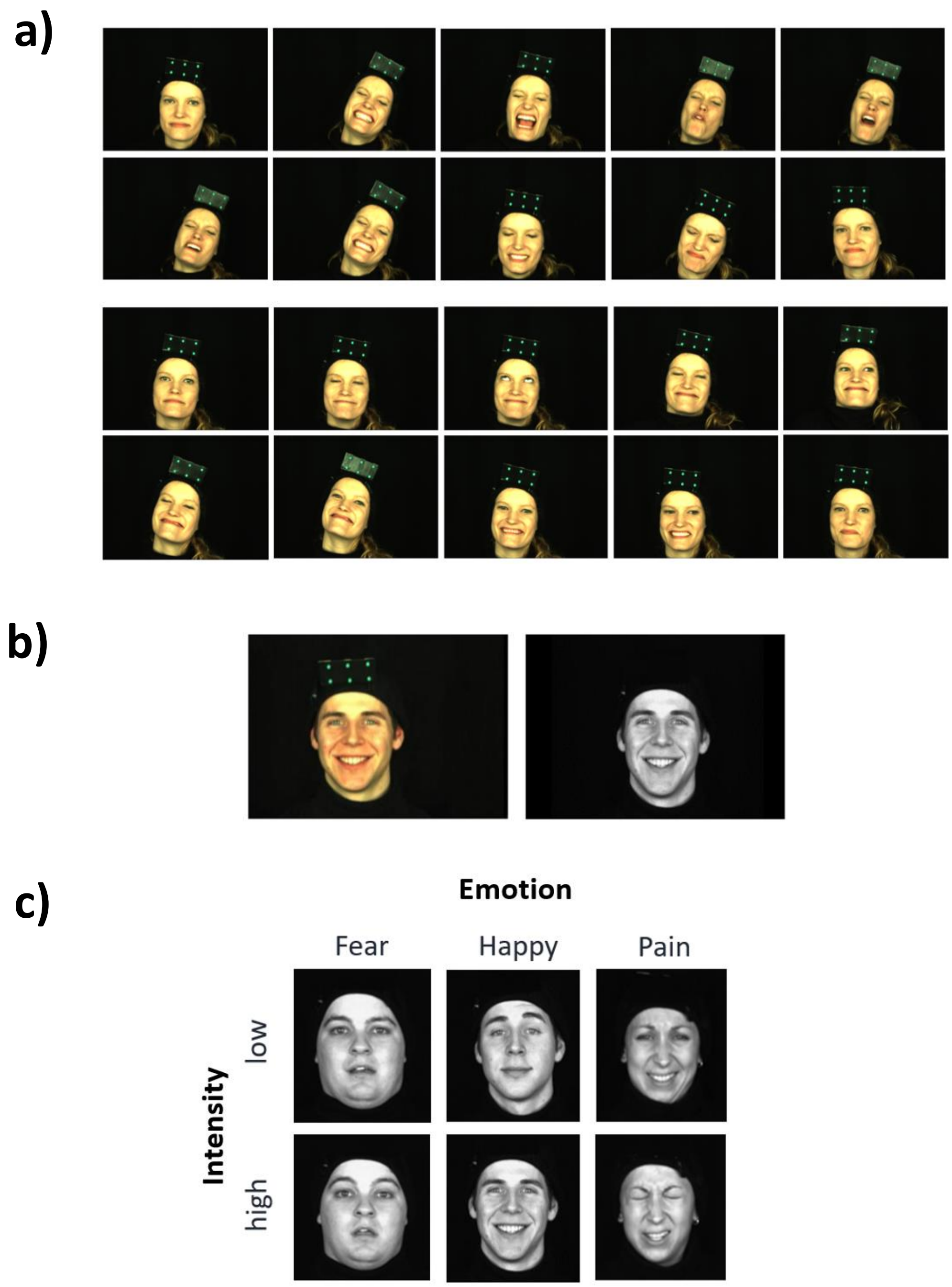

Figure 1. **Examples of stimuli**. *a) Frames, presented in temporal order, extracted from raw videos of facial expressions elicited by Happy (upper rows) and Pain (bottom rows) scenarios. b) Raw (left) and final (right) version of the facial emotion stimuli. c) Examples of high- and low-intensity expression of Fear, Happy and Pain.*

Participants performed an integrated emotion categorization and comparison task. For each trial, they saw pairs of facial emotions presented sequentially. In particular, the decision not to present faces simultaneously was driven by the study's aim to measure the similarity of stimuli based on participants' internal representation of the expressions seen. Essentially, we sought to prevent participants from specifically looking for physical, image-based, details of the images/videos. It is important to note that presenting the stimuli sequentially introduced a working memory component to the task. We might anticipate that altering the timing between comparison stimuli could affect similarity judgments. To mitigate this potential issue, the timing between stimuli was kept constant across all the studies of this work that integrated a similarity rating task.

Each trial started with a fixation cross (1000ms) followed by a text screen announcing the first stimulus ("Image 1" for the Static task or "Video 1" for the Dynamic Task, 500ms). Then the first stimulus appeared centered on the screen. In the dynamic version the stimulus lasted 2000ms or 1000ms (depending on the video length), while in the static version the image was shown for 2000ms. The stimulus was followed by a response screen where the participants could press one of three buttons to categorize each facial expression as Pain, Fear, or Happy. The order of the buttons was counterbalanced across participants. The participant responded by pressing one of three keys: "g" for the label on the left, "h" for the centre, and "j" for the right. Participants had 5000ms to respond, or otherwise, the task went ahead to the following presentation. The same structure was repeated for the second stimulus, a fixation cross (1000ms) was followed by a text screen announcing the second stimulus, and then the second stimulus appeared centered on the screen. The second stimulus showed a facial expression of the same actor lasting for the same amount of time. Finally, the last response screen asked participants to rate stimuli similarity on a 7-points Likert scale, ranging from 1 (Totally different) to 7 (Exactly the same). With the aim to obtain a measure of perceived similarity and not to guide participants responses, no specific criteria were provided on what information to use for judgment (e.g., physical similarity, emotion similarity). Participants were generally instructed to "Rate the extent of similarity or dissimilarity between the two facial expressions".

As the study took place online, measures were taken in the construction of the tasks to reduce and account for potential participants' distractions or disengagement. In particular, (1) access to the study was restricted to PCs or laptops (i.e., no mobile phone or tablet allowed); (3) participants were required to self-report the reliability/usefulness of their data (e.g., due to lacked attention) at the end of the experiment, and were assured that their response would not affect their compensation in any way; (4) a time-limit was imposed on each screen, participants could only pause the experiment by closing the browser window and reopening it through the provided link. To prevent this scenario, participants exceeding the maximum allotted time for completing the experiment were excluded from the study; (5) control trials were included in the study, where participants were asked to rate the similarity of identical images.

Procedure

The study was run online using the Gorilla platform (https://gorilla.sc/), and participants gained access through a URL link using their desktop computer or laptop (i.e., no tablet or phone access was allowed). First, information about the study was displayed to obtain informed consent. Once consent was given, participants were directed to a demographic questionnaire. Closed-ended questions asked for their hand dominance (right-handed/left-handed) and the gender they identify with (male/female/other); while open-ended questions asked for their age and nationality. Detailed instructions were then displayed, followed by a practice stage to familiarize the participant with the task.

Differently from the majority of previous works, the stimuli adopted in these studies represented no-prototypical, scenario-driven, facial expressions which may be more challenging to recognize and differentiate compared to the facial emotion adopted by more traditional datasets. Moreover, the integrated emotion categorization and comparison task introduced in this study has never been used before. Thus, with the aim to define the difficulty of the task and the processing of these stimuli, trials where participants were asked to categorize and rate the similarities of the same stimuli, no physical difference, were introduced. These trials were also adopted as a control to identify participants who completed the task attributing casual scores or who were particularly distracted, a valid risk to consider particularly since the study was conducted online.

In particular, each participant went through 160 experimental trials. With the aim to generate and control for variability in our stimuli sample, the 160 experimental trials included: 40 trials showing faces with the same emotion and same intensity (i.e., identical), 40 trials showing faces with the same emotion but with different intensity (e.g., high intensity happy vs low intensity happy), 40 trials showing faces with different emotions at the same level of intensity (e.g., low intensity happy vs low intensity fear), and 40 trials showing faces of different emotions at different level of intensity (e.g., high intensity happy vs low intensity fear). These 160 pairs of stimuli were randomly selected out of 216 possible stimuli combination, with the aim to maintain the 40/40/40/40 ratio, please see

Appendix A, Figure A.2. Trials were fully randomized across experiment trials.

During each trial, participants first categorized each of the two facial expressions and then rated the similarity between them using a Likert scale (Figure 2). Half of the participants were randomly allocated to the Dynamic task, and the other half to the Static task. They were instructed to focus on the task and to respond as quickly and accurately as possible. At the end of the study, participants were debriefed and completed a self-report questionnaire about the reliability of their data. Participants took about 35-40 minutes to complete the study. They were allowed a break of no more than 5 minutes, which occurred after 80 trials.

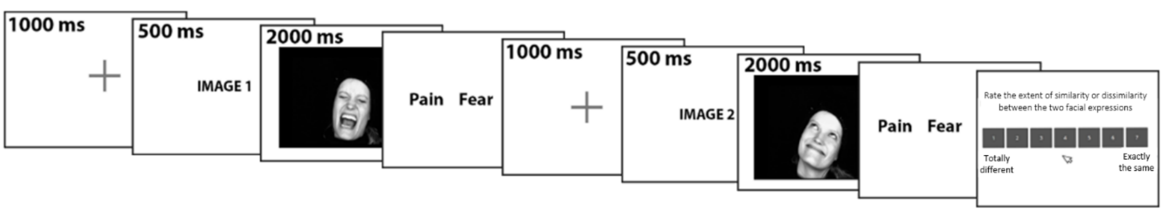

Figure 2. **Emotion categorization and similarity rating task**. *In each trial, participants categorized two facial expressions according to their emotion: Happy, Fear, or Pain. Subsequently, using a Likert scale ranging from 0 to 7, they rated the degree of similarity between the two facial expressions.*

Computing stimuli physical similarity between facial emotions

We calculated the physical similarity between two facial emotions using Gabor similarity (Lades et al., 1993). A Gabor jet was placed at each node of a uniform 10 x 10 grid covering the entire image. The square grid marked the positions in the image space from which filter convolution comparison values would be extracted. Each jet consisted of 80 filters (or kernels) at 8 equidistant orientations, 5 spatial scales, and 2 phases (sine and cosine). The two kernels produced at each location by the 90-deg phase shift allowed sensitivity to the direction of contrast in the images, which is particularly important for detecting complex stimuli such as faces. With 100 locations, nodes of the grid, we had a total of 8,000 kernels per image. The image was convolved with each jet, and the model stored the magnitude and phase values of the image. Consequently, a high-dimensional feature vector was obtained in response to all the Gabor jets by concatenating the two values from each of the 8,000 filters. The differences in activation between the corresponding filters of two stimuli were used to determine the similarity of the images and the measure of similarity was computed as the Euclidean distance between the two features' vectors derived from convolution with Gabor filters.

For the Static face stimuli, we computed the physical similarity of the two grayscale images (256 by 256 pixels) in each of the 160 trials and then averaged across all trials in each of the experimental conditions (3 emotions * 2 intensities). Although the Gabor similarity measure is commonly used for processing static images, this method of analysis has also been successfully applied to video animation (Dobs et al., 2014). For the Dynamic stimuli, we calculated the physical similarity between every frame of the two videos, converted in grayscale images (256 by 256 pixels), in each of the 160 trials. We then determined the mean Gabor similarity as a measure of video similarity and averaged across all trials, as did with static faces.

2.2.2 Results and Discussion

Data from 5 participants were excluded, resulting in a final sample size of 75 participants for the analysis (35 static, 40 dynamic). These five participants either failed to respond to more than 50% of the trials, rated over half of identical facial emotions as below 3 (on a 7-points Likert scale for similarity rating), or reported that their data were unreliable in the final questionnaire.

Dynamic cues improve emotion categorization

I first compared the overall error rates in the categorization of Static vs Dynamic facial expressions (Fear, Happy, and Pain) using an independent t-test. As shown in Figure 3a, there were significantly fewer errors overall in the dynamic face task (17.65%) than in the static face task (22.13%), $t(72) = 2.96$, $p = .004$. Next, I examined how different facial expressions were mistakenly categorized in dynamic and static faces (i.e., where the errors of emotion recognition originated). The confusion rates for the three facial emotions are illustrated in Figure 3b. Participants tended to confuse facial expressions of Fear more with Pain compared to Happy [for Dynamic faces, 79.5%, $t(38) = 6.68$, $p < .001$; for Static faces, 63% $t(34) = 2.79$, $p = .009$], facial expression of Happy more with Pain compared to Fear [for Dynamic faces, 78.1%, $t(38) = 4.69$, $p < .001$; for Static faces, 67.9%, $t(34) = 3.9$, $p < .001$], and Facial expressions of Pain were confused more with Fear compared to Happy [for Dynamic faces, 76% $t(38) = 8.69$, $p < .001$; for Static faces, 67.8%, $t(34) = 5.15$, $p < .001$]. Finally, overall, participants made significantly more errors distinguishing facial expressions of Pain (for Static, M = 37; for Dynamic, M = 34), followed by Fear (for Static, M = 21; for Dynamic, M = 14) and Happy (for Static, M = 11; for Dynamic, M = 6) for both, Static (all $ts(34) > 5.27$, $p < .001$) and Dynamic (all $ts(34) > 6.15$, $p < .001$).

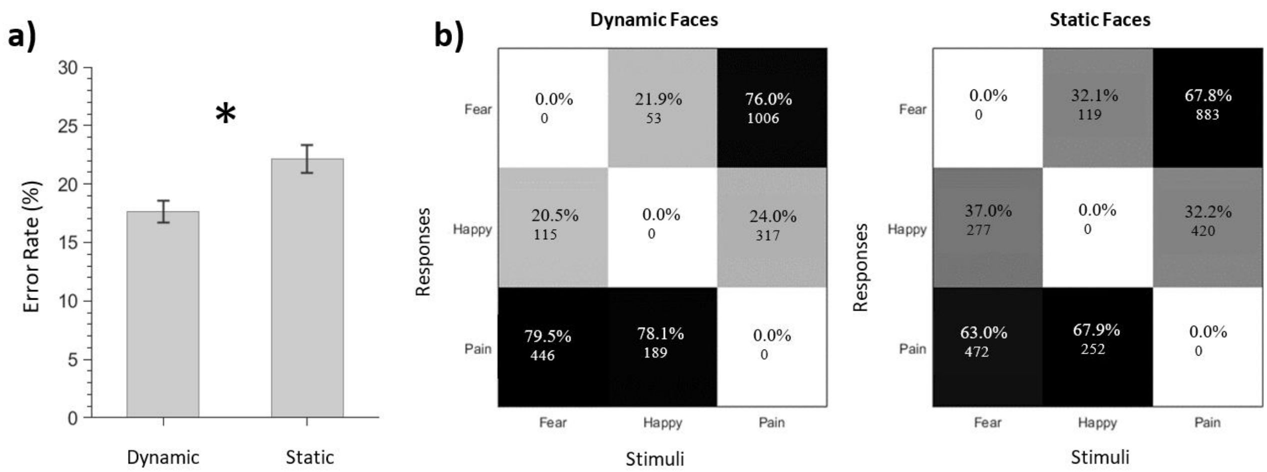

Figure 3. **Participants performance to the Categorization task**. *Results of the emotion categorization task. a) Overall error rates for categorizing static and dynamic facial emotions. b) distribution of errors during emotion categorization. X axis represents the stimuli shown, and Y axis represents participants responses. Percentage in each cell were calculated based on the overall wrong answer given to the stimuli of a specific emotion category (i.e., each column of the matrix is equal to 100%). The value reported under percentage is the total number of the answers given to that specific label at that specific stimulus across all participants.*

Physical similarity correlates with perceptual similarity for both static and dynamic facial emotions

Stimuli-based similarity was calculated using Gabor similarity between two face images or videos for each of the 160 trials (Dobs et al., 2014). A correlation analysis was performed to test whether physical similarity predicts perceptual similarity between two facial expressions shown in a trial. As shown in Figure 4, there were significant correlations between image- or video-based similarity and participants' ratings about their similarity in terms of facial emotion, for Static faces, $r = .87$, $p < .001$; for Dynamic faces, $r = .88$, $p < .001$.

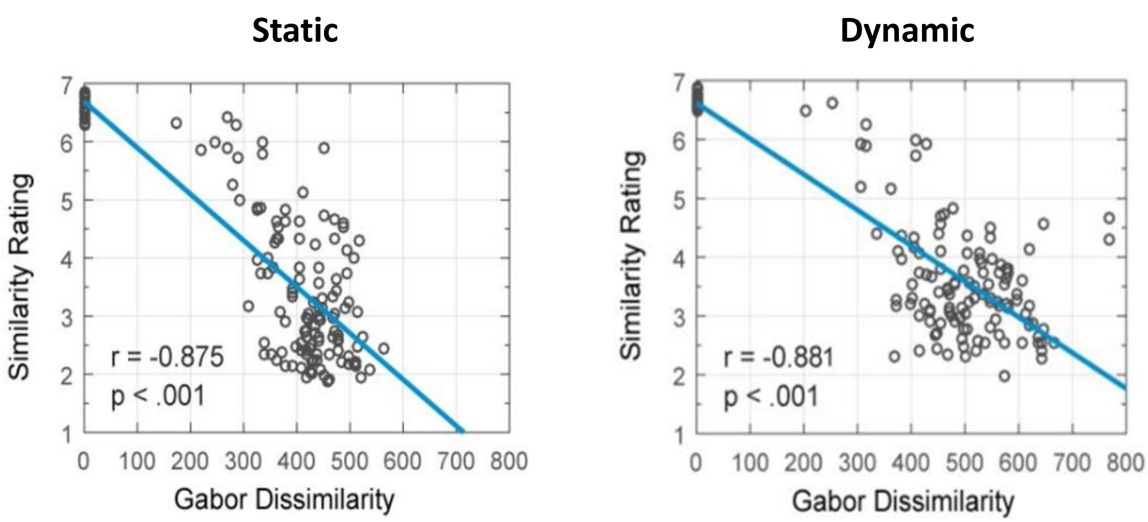

Figure 4. **Correlation between physical similarity measures and participants similarity ratings.** *Each dot represents one of the 160 pairs of videos or images shown in the Similarity Rating task.*

Categorical similarity correlates with perceptual similarity for both static and dynamic stimuli

Next, I tested whether similarities inferred from the emotion categorization task (i.e., Categorical similarity) are related to perceptual similarity obtained directly from the similarity eating task. To derive categorical similarity, I used confusion error rates from the categorization task as a measure of distance (or dissimilarity) between facial expressions of emotions.

Using confusions errors, I first computed a similarity matrix across the three facial emotions for each participant and for each facial motion condition (dynamic or static). These matrices were then averaged across participants, resulting in 2 matrices representing an indirect measure of perceptual similarity (Figure 5a). For perceptual similarity, I averaged participants' responses to the similarity rating task across each combination of facial emotion (e.g., Fear vs Happy) to produce a perceptual similarity matrix for each participant, which were then averaged across participants, resulting in 2 matrices representing a direct measure of perceptual similarity (Figure 5b). Correlation analysis between the two types of similarity matrices, using Spearman rho coefficient, showed a strong correlation between categorical similarity and its corresponding perceptual similarity for both dynamic and static facial emotions (for Static, $r = .99$, $p < .001$; for Dynamic, $r = .97$, $p < .001$). This result suggests that, on a group-level, categorical emotion information may play a role in perceptual similarity. To assess the consistency of these findings across participants, I analysed participant-level data by calculating the Spearman's rho coefficient between each participants' emotion categorical matrix and their corresponding perceptual similarity matrix. As can be seen in Figure 6. For the Dynamic condition, all participants exhibited a significant correlation between their categorical and perceptual similarity matrices (all rs $\geq .75$, all ps $\leq .01$). Similarly, for the Static condition, all participants showed significant correlations (all rs $\geq .74$, all ps $\leq .02$) except for one ($r = .28$, all $p = .45$). Therefore, how confused two facial expressions are in a categorization task, is related to how people rate their similarity directly.

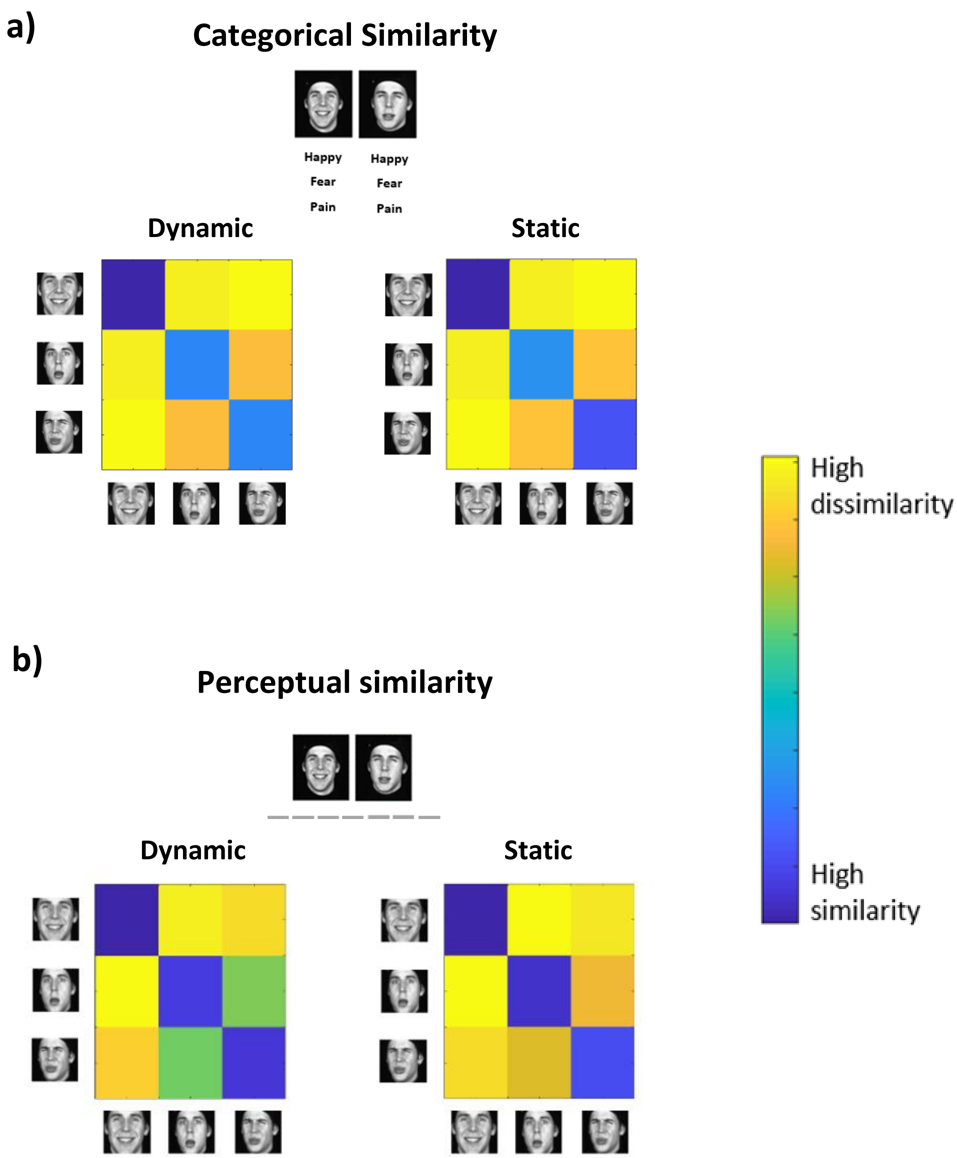

Figure 5. **Categorical and Perceptual similarity matrices across facial emotons.** *a) Categrocial similarity derived from the emotion categorization task. b) Perceptual similarity obtained from the similarity rating task. Along the diagonal are values obtained comparing response to the same facial emotion. Facial emotions from top to bottom (Y axis) and from left to right (X axis) are Happy, Fear, and Pain respectively.*

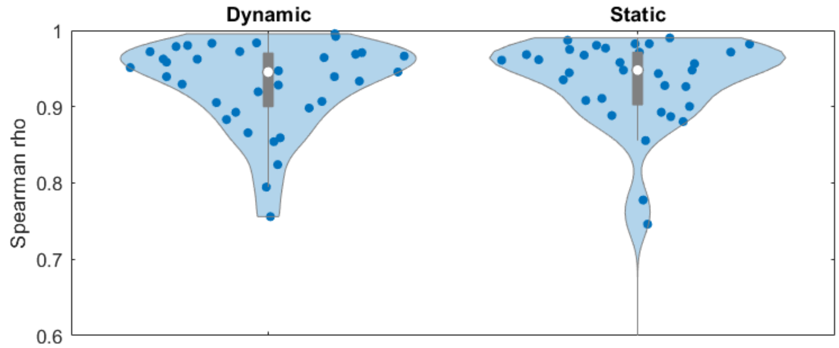

Figure 6. **Correlation coefficients between perceptual and categorical similarity matrices for individual participants**. *All but one participant who took part in the Static task showed significant correlations.*

Physical and Categorical information combined outperform independent predictors and can better explain the way participants perceive differences and similarities between facial expressions of emotions

As both physical and categorical emotion information are correlated with perceptual similarity between facial expressions of emotions, my next objective was to assess the relative contributions of these two factors, and to test whether their combination provides a better account for perceptual similarity between facial emotions. I performed a multiple linear regression model with physical and categorical similarity as predictors for perceptual similarity for the 160 trials. For static facial emotions, Categorical similarity accounted for a significant proportion of the variance in the dependent variable, R^2 adjusted = .542, $F(1, 158) = 189$, $p < .001$. Physical similarity alone gave higher predicting power, accounting for more variances, R^2 adjusted = .825, $F(1, 158) = 750$, $p < .001$). Interestingly, combining both Categorical similarity and Physical similarity yielded the best fit to the data, R^2 adjusted = .859, $F(2, 157) = 484$, $p < .001$ (see Figure 7). Comparisons between the three models (i.e., Categorical similarity; Physical similarity; combined) were conducted to evaluate the incremental contribution of the predictors. When comparing Categorical model to the combined model, the addition of Physical similarity significantly improved the model fit ($\Delta R^2 = .315$, $F(1, 157) = 355$, $p < .001$). Similarly, the inclusion of Categorical similarity alongside Physical similarity resulted in a significant enhancement in model fit compared to Physical model alone ($\Delta R^2 = .034$, $F(1, 157) = 38.8$, $p < .001$). Comparable results were obtained for the dynamic facial emotions (Figure 8). Categorical similarity alone accounted for a significant proportion of the variance in perceptual similarity, R^2 adjusted= .660, $F(1, 158) = 310$, $p < .001$. Physical similarity alone showed higher contribution to perceptual similarity, R^2 adjusted = .776, $F(1, 158) = 552$, $p < .001$. Again, the combined model provided the best fit to the data, R^2 adjusted = .868, $F(2, 157) = 524$, $p < .001$ (see Figure 8**Error! Reference source not found.**). When c omparing the contribution of the Categorical model to the combined model, the addition of Physical similarity significantly improved the model fit ($\Delta R^2 = .092$, $F(1, 157) = 111$, $p < .001$). Similarly, the combined model showed a significant enhancement in model fit compared to Physical model alone ($\Delta R^2 = .133$, $F(1, 157) = 208$, $p < .001$). These findings indicate that both Categorical similarity and Physical similarity contribute to perceptual similarity, though Physical similarity seemed to be a more dominate predictor.

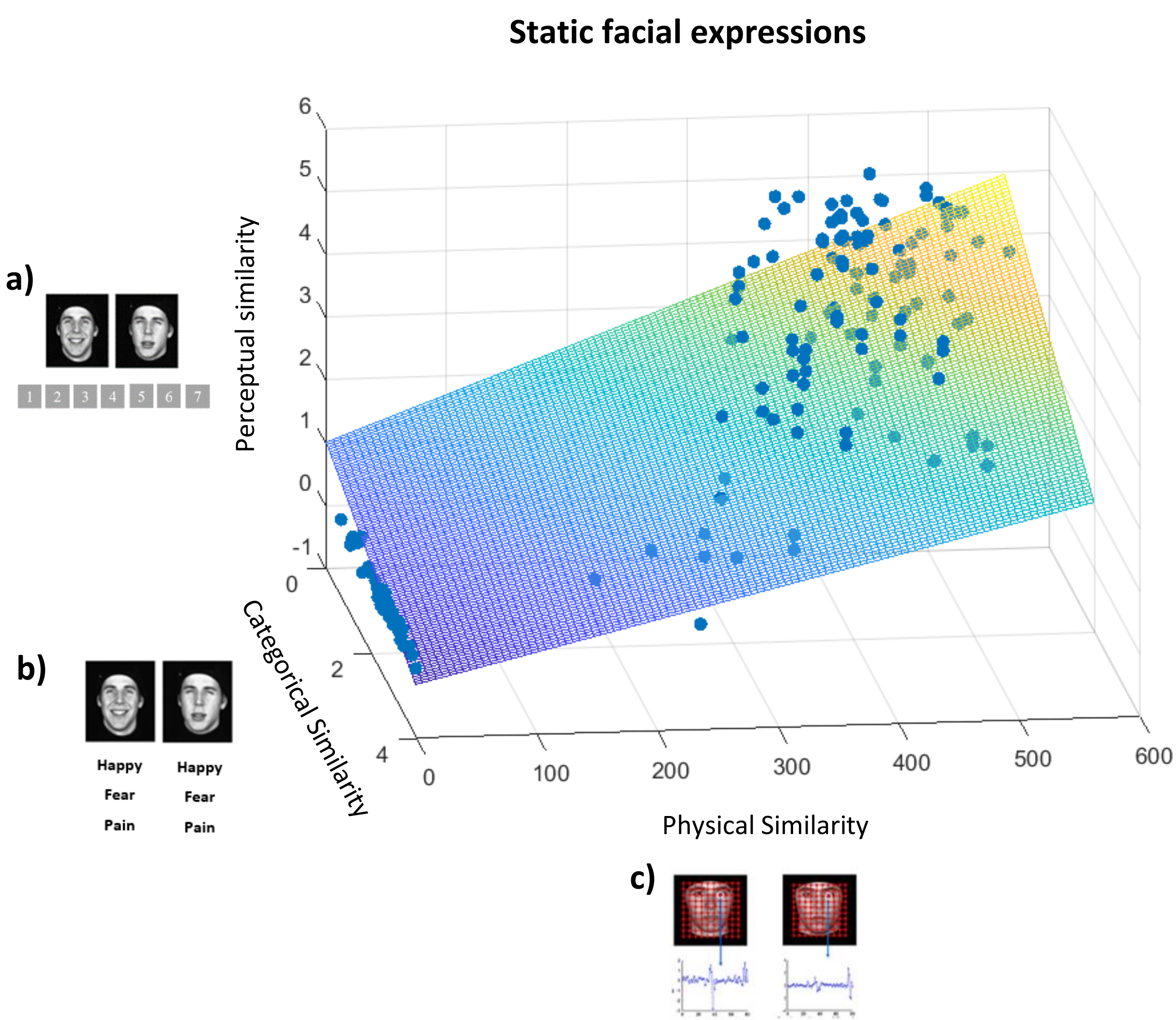

Figure 7. **Multilinear regression model integrating Physical and Categorical similarity to predict participants' perceptual similarity of static facial expressions.** *Each dot represents the similarity measure for one of the 160 pairs of static facial emotions. a) Perceptual similarity obtained from participants responses to a similarity rating task. b) Physical similarity computed using Gabor similarity. c) Categorical similarity computed from confusions errors in the emotion categorization task.*

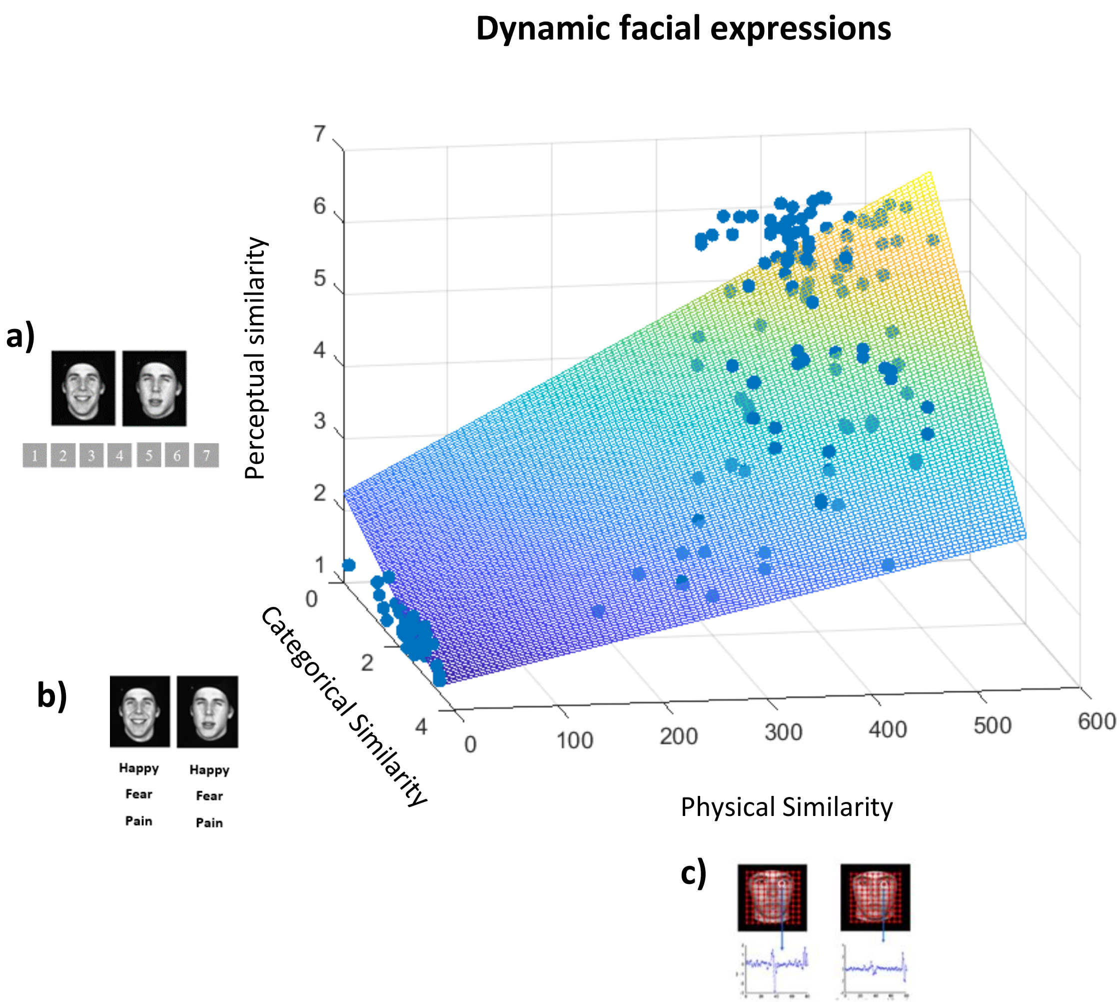

Figure 8. **Multilinear regression model integrating Physical and Categorical similarity to predict participants' perceptual similarity of dynamic facial expressions.** *Each dot represents similarity the measure for one of the 160 pairs of dynamic facial emotions. a) Perceptual similarity obtained from participants responses to a similarity rating task. b) Physical similarity computed using Gabor similarity. c) Categorical similarity computed from confusions errors in the emotion categorization task.*

2.3 Study 2. How does emotion intensity contribute to perceptual similarity of dynamic and static facial emotions?

Study 1 showed that both Physical and Categorical information of facial emotions contribute to how we perceive their similarities. In Study 2, I investigated how emotion intensity information affects perceptual similarity between dynamic and static facial emotions. To this end, participants were first asked to rate the intensity of dynamic or static facial emotions before judging their similarity. It is worth noting that we defined emotion intensity differently from previous studies on facial emotions. Different levels of emotion intensity are usually obtained by directly controlling the physical muscles involved in that specific emotion, such as exaggerating facial movements, cutting the video before the peak of the emotion is reached, or using a morphing between two facial expressions. A few assumptions are at the basis of these methods which are mostly in line with categorical and low-dimensional theories. One is that emotional intensities differ primarily in a quantitative rather than qualitative manner. In other words, experiencing intense happiness is not fundamentally distinct from experiencing mild happiness in terms of their intrinsic nature. Instead, they share the same underlying experience that is heightened in intense emotions. As a result, different levels of intensity are expressed through the same patterns of facial muscles. A second assumption is that, as for the underlying emotional experience, the expression of the emotion intensity differs quantitatively more than qualitatively. This means that the more facial muscles are contracted or displayed, the more intense the expressed emotion is. Different from previous studies, stimuli adopted in this project have been selected to convey a more natural and semantic definition of emotion intensity. In particular, different levels of emotional intensities have been obtain by asking participants to mimic the facial expression evoked by specific emotional scenarios which have been validated to elicit a high- or low- intensity emotion (e.g., high-intensity: ‘You are lying on your couch after a delicious dinner’; low-intensity: ‘You have reached a goal and you are happy to have accomplished it’).

We also slightly varied the way participant rate the similarity between facial emotions to test whether the results obtained in Study 1 generalize to new conditions tested in Study 2. Finally, by integrating results from both studies, I was able to examine what factors may produce the best explanatory power for perceptual similarity between facial expressions of emotions.

2.3.1 Methods

Participants

Based on results from our previous study, a priori power analysis was conducted using G*Power (Faul et al., 2007). The analysis was conducted for a bivariate normal model correlation. I adopted a conservative approach and assumed a H1 = .80, and a no correlation at all for H0. With a significance criterion of $\alpha = .05$ and power = .95, the minimum sample size needed for condition (Static or Dynamic) is 13. Considering the potential issues generated by conducting the study online (e.g., lack of attention), I collected 20 participants per condition. Forty participants were recruited from the University of East Anglia using the SONA System (11 males, 29 females; age ranged between 19-30 yrs., M = 20.8, SD = 2.97). The participants sample included 29 British, 8 no-British European, 2 Asian, and 1 American. None of the participants had taken part in Study 1. All participants were naïve to the purpose of the investigation, provided informed consent before taking part in the study and were debriefed at the end, receiving course credits as compensation.

Stimuli, Materials, and Tasks

The stimuli and materials were the same as in Study 1, including 160 pairs of facial images or videos showing 3 different emotions (Happy, Pain, Fear) at 2 intensities (High, Low). The tasks used were similar to Study 1 with the exception that participants rated the emotion intensity of each facial emotion (rather than defining the emotion category) before judging their similarity. Specifically, participants were asked to rate the intensity of each facial expression on a 7-points Likert. They had 10000ms to respond, otherwise, the task went ahead to the following facial expression. The same procedure was repeated for the second stimulus. As in Study 1, the last response screen asked them to “Rate the extent of similarity or dissimilarity between the two facial expressions”. However, different from Study 1, where a 7-point Likert scale was used, participants in Study 2 made their judgments by moving the handle of a 0-100 slider where 0 represented “Totally different” and 100 “Exactly the same” (Figure 9). The handle was initially placed on 50. This manipulation allowed me to check whether the specific way used to collect participants responses may influence similarity ratings.

The same measures taken in the first study were implemented with the aim to reduce, and account for, potential participants’ distractions or disengagement. In particular, (1) access to the study was restricted to PCs or laptops; (3) participants were required to self-report the reliability/usefulness of their data (e.g., due to lacked attention) at the end of the experiment, (4) a time-limit was imposed on each screen, and participants exceeding the maximum allotted time for completing the experiment were excluded from the study; (5) control trials were included in the study, where participants were asked to rate the similarity of identical images

Procedure

The procedure was the same as in Study 1, and, similarly, the study was conducted online using the Gorilla platform. During the experimental trials, participants first rated the intensity of each of the two facial expressions of the same actor and then judged their similarity (Figure 9). Half of the participants were randomly allocated to the Dynamic task, and the other half to the Static task. Before the experimental task, participants had 3 practice trials to familiarize themselves with the procedure. At the end of the experiment, participants were asked to judge the reliability of their data, considering possible distractions or lack of attention that occurred during the completion of the experiment. Participants took about 30-40 minutes to complete the study. They were allowed a break of no more than 5 minutes, which occurred after 80 trials.

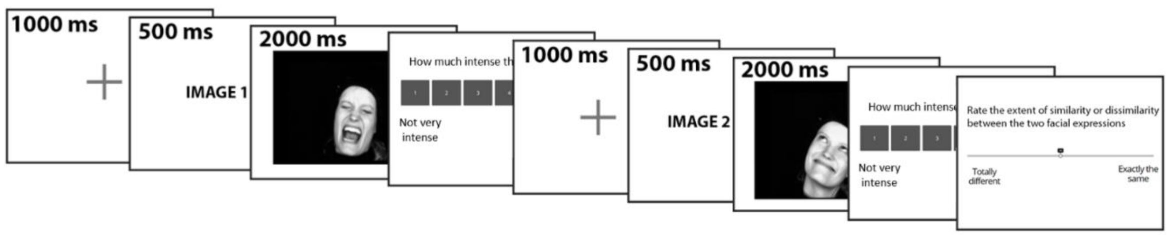

Figure 9. **Emotion intensity and similarity rating task**. *In each trial, participants first rated the intensity of two facial emotions. Then, using a slider ranging from 0 to 100, they rated the similarity between the two facial emotions.*

2.3.2 Results and Discussion

Following the same exclusion criteria set in Study 1, no data was excluded from the analysis, resulting in a final sample size of 40 participants for the analysis (20 static, 20 dynamic).

Static facial emotions are perceived as more intense than dynamic facial emotions

We conducted a 2 by 2 mixed model ANOVA to investigate whether perceived emotion intensity is influenced by facial motion and intensity of facial expressions. As shown in Figure 10, there was a significant main effect of stimuli's intensity, $F(1,38) = 115.99$, $p < .001$, with facial expressions elicited by high intensity emotional scenarios being perceived as more intense than those elicited by low intensity emotional scenarios. The main effect of facial motion was also significant, $F(1,38) = 4.03$, $p < .05$, with static emotions perceived as more intense than dynamic stimuli. There was no significant interaction between intensity and facial motion, $F(1,38) = .041$, $p = .84$. Further contrasts showed significant differences between emotion intensity for both dynamic and static stimuli, for Dynamic, $t(19) = 6.84$, $p < .001$; for Static, $t(19) = 8.86$, $p < .001$ (see Figure 10). Similarly, a difference between static and dynamic facial emotions was observed for facial expressions elicited by both high and low emotional scenarios, for High intensity, $t(38) = 2.03$, $p < .05$; for Low intensity, $t(38)= 1.97$, $p < .05$ (see Figure 10).

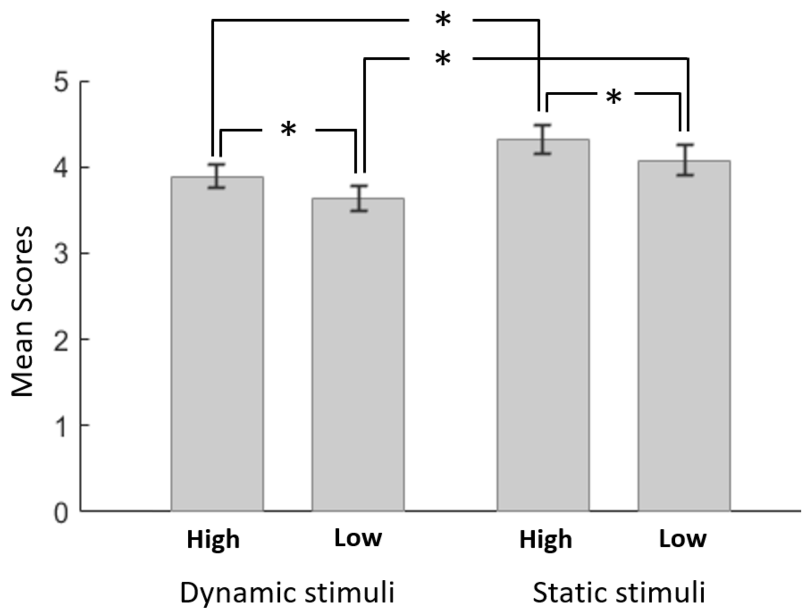

Figure 10. **Perceived emotional intensity from dynamic and static facial expression elicited by high and low intensity emotional scenarios.**

Completing either a categorization task or an intensity rating task before the similarity rating task does not seem to influence participants' performance

Next, I tested whether procedural differences between Studies 1 and 2 affected perceptual similarity scores. In Study 1, participants performed an emotion categorization task before providing similarity judgments, whereas in Study 2, they rated emotional intensity first. The two studies also used different rating scales for the similarity rating task. Following the same procedure as Study 1, I computed similarity matrices from participants' rating responses for both dynamic and static stimuli. I then correlated the perceptual similarity matrices for dynamic and static stimuli across the two studies (see Figure 11). Results showed that similarity matrices were highly correlated for both Static, $r = .95$, $p < .001$, and Dynamic facial emotions, $r = .98$, $p < .001$. Similarly, at a trial-by-trial level, perceptual similarity observed from the two studies were also highly correlated for both Static, $r = .96$, $p < .001$, and Dynamic facial emotions, $r = .97$, $p < .001$ (see Figure 12).

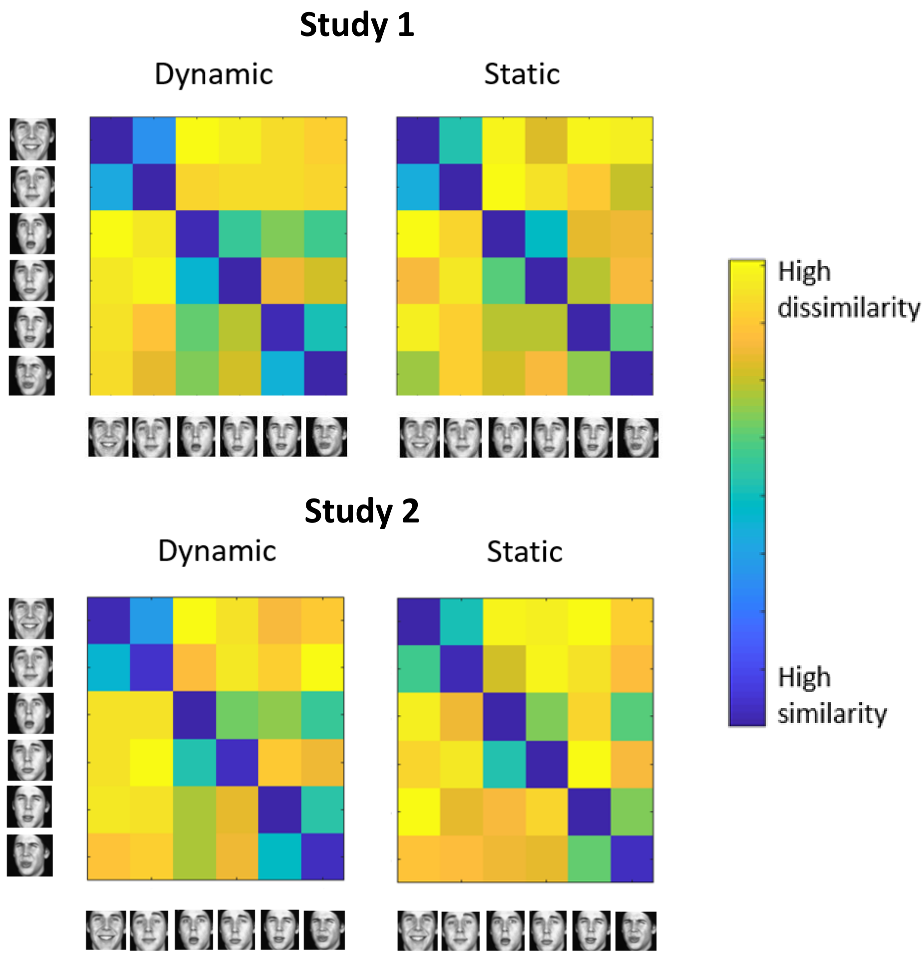

Figure 11. **Perceptual similarity matrices observed in Study 1 and Study 2.** *Facial expressions from top to bottom (Y axis) and from left to right (X axis) are high intensity Happy, low intensity Happy, high intensity Fear, low intensity Fear, high intensity Pain, low intensity Pain.*

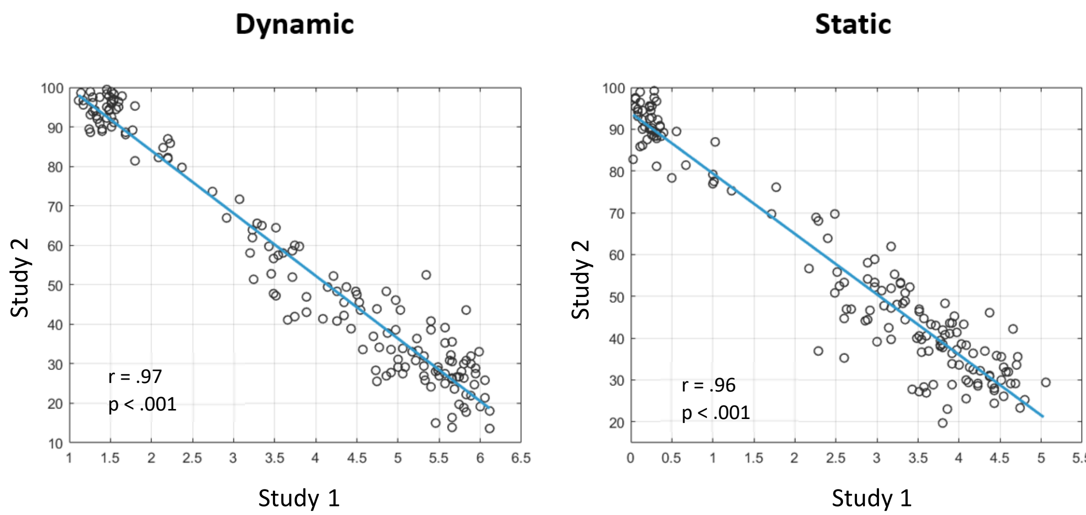

Figure 12. **Correlation between participants Similarity ratings to the first and second study.** *Each dot represents mean perceptual similarity across participants for one of the 160 pairs of videos or images shown in the Similarity Rating task.*

Physical, stimulus-based similarity correlates with participants' perceptual similarity for both static and dynamic stimuli

In attempt to replicate results from Study 1, I tested whether stimulus based physical similarity correlates with perceptual similarity. As shown in Figure 13, consistently with Study 1, there were significant correlations between physical and perceptual similarity for Static faces, $r = .90$, $p < .001$, and for Dynamic faces, $r = .88$, $p < .001$.

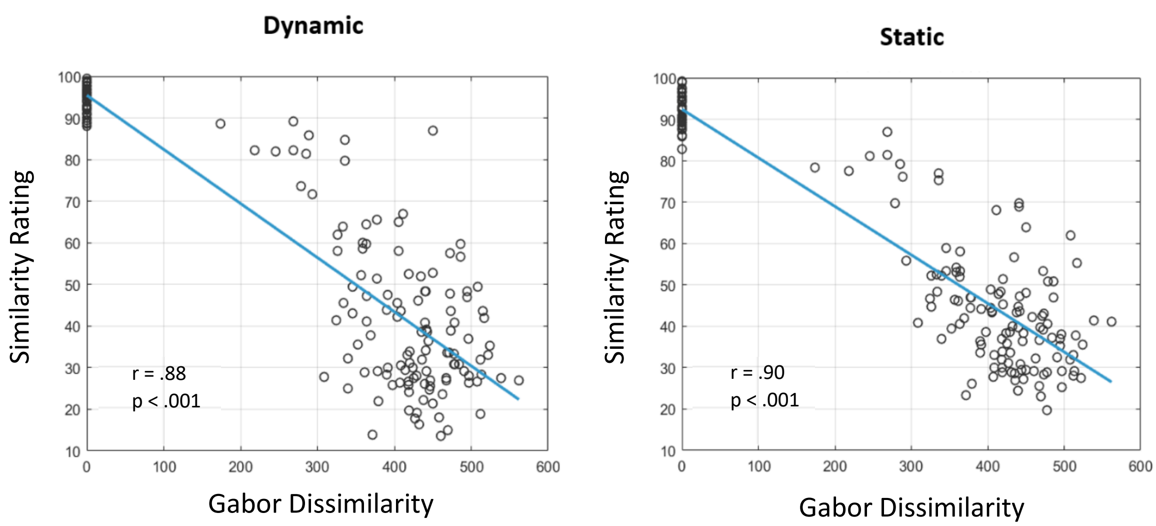

Figure 13. **Correlation between stimuli-based physical similarity and perceptual similarity.** *Each dot represents the similarity for one of the 160 pairs of videos or images shown in the Similarity Rating task.*

Intensity similarity correlates with perceptual similarity on a trial basis but not on a group- and participant- level

For each pair of facial emotions in each trial, intensity similarity (or distance) was defined as the absolute difference between their rated intensity scores. The intensity similarity obtained for the 160 trials was then used to create a similarity matrix across facial emotions for each participant, which was then averaged across participants , resulting in 2 matrices, one for each face motion condition (see Figure 14a). As in study 1, participants' responses to the similarity rating task across each combination of facial emotion (i.e., Fear, Happy, Pain) were used to produce perceptual similarity matrices for dynamic and static facial emotions (see Figure 14b). For Intensity similarity, values in the diagonal of the matrix, where stimuli pairs were identical, were all 0. To prevent diagonal values from inflating our results, I extracted and computed the Spearman rho coefficient using the upper triangle of the matrices. Since the perceptual similarity matrix lacks symmetry along the diagonal line, the mirrored cells in the upper and lower triangles were averaged. Results showed no significant correlation between intensity similarity matrices and its corresponding perceptual similarity matrices for both dynamic and static stimuli (Static, r = .37, p = .16; Dynamic, r = .34, p = .20). This suggests that, on a group-level, intensity information cannot be used to predict perceptual similarity. Similarly, analysing data on a participant-level by calculating the Spearman's rho coefficient between each participants' emotion intensity matrix and their corresponding perceptual similarity matrix, resulted in no significant correlations between matrices for all participants (for the Static, all rs ≤ .47 ; all ps ≥ .07 ; for the Dynamic, all rs ≤ .49 ; all ps ≥ .06) except for 2 participants taking part to the Static version of the study (both rs ≥ .50; both ps ≤ .05) and 2 taking part to the Dynamic version of the study (both rs ≥ .69; both ps ≤ .004). However, when conducting a correlation analysis on a trial-level, there was a significant correlation for both Static and Dynamic facial emotions (see Figure 15), for Static, r = 63, p < .001; for Dynamic r = 67, p < .001.

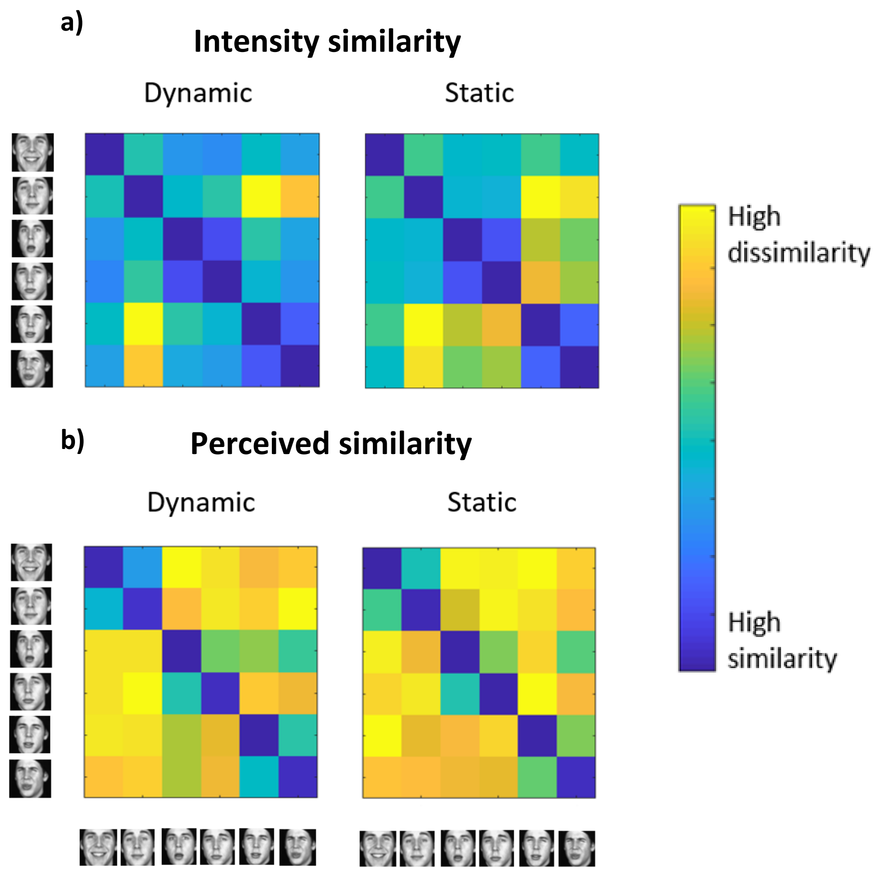

Figure 14. **Intensity and Perceptual similarity matrices obtained for Static and Dynamic facial emotions.** *Facial expressions from top to bottom (Y axis) and from left to right (X axis) are high intensity Happy, low intensity Happy, high intensity Fear, low intensity Fear, high intensity Pain, low intensity Pain.*

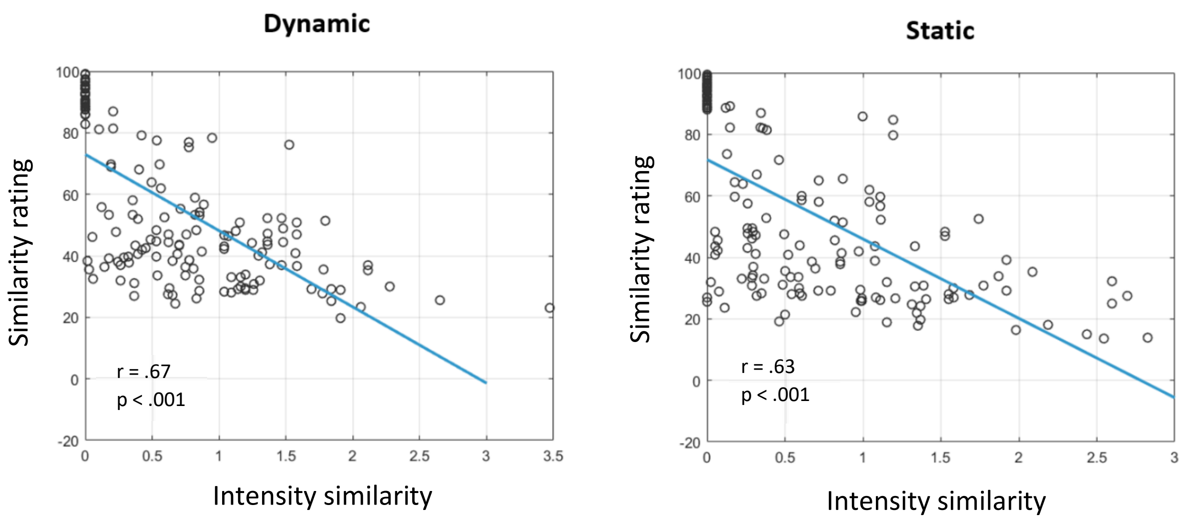

Figure 15. **Correlation between Intensity similarity and Perceptual Similarity observed for individual trials.** *Each dot represents one of the 160 pairs of facial emotions shown in the Similarity Rating task.*

Adding Intensity similarity slightly enhances the prediction of perceptual similarity compared to Physical similarity alone

I further examined how intensity similarity, physical similarity and their combination performs in predicting perceptual similarity across the 160 trials with a multiple linear regression analysis. For static facial emotions, physical similarity alone accounted for a significant proportion of the variance in perceptual similarity, R^2 adjusted = .825, $F(1, 158) = 750$, $p < .001$. Intensity similarity alone can account for significant but a smaller proportion of variances, R^2 adjusted = .442, $F(1, 158) = 127$, $p < .001$. The combined model incorporating both Physical and Intensity similarity showed the highest performance, R^2 adjusted = .838, $F(2, 157) = 412$, $p < .001$ (see Figure 16). Models comparisons showed a small but significant increment in performance when intensity similarity is added to physical similarity, $\Delta R^2 = .013$, $F(1, 157) = 13.7$ $p < .001$, and a substantial increment in performance when physical similarity is added to intensity similarity, $\Delta R^2 = .394$, $F(1, 157) = 387$ $p < .001$. Similar results were obtained with dynamic facial emotions. Intensity similarity accounted for a smaller but significant proportion of the variance in perceptual similarity, R^2 adjusted = .503, $F(1, 158) = 162$, $p < .001$, compared to the proportion of variance explained by Physical similarity alone, R^2 adjusted = .776, $F(1, 158) = 552$, $p < .001$. Again, the combined model yielded the best fit to the data, R^2 adjusted = .826, $F(2, 157) = 379$, $p < .001$ (see Figure 17). When comparing the contribution of the Intensity similarity with the combined model, addition of Physical Similarity substantially improved the model fit, $\Delta R^2 = .323$, $F(1, 157) = 295$, $p < .001$. Similarly, the inclusion of Intensity similarity alongside Physical similarity resulted in a significant enhancement in model fit compared to Physical model alone, $\Delta R^2 = .0509$, $F(1, 157) = 46.6$, $p < .001$. These findings indicate that both intensity and physical similarity contribute to perceptual similarity between facial emotions, with physical similarity playing a prevailing role.

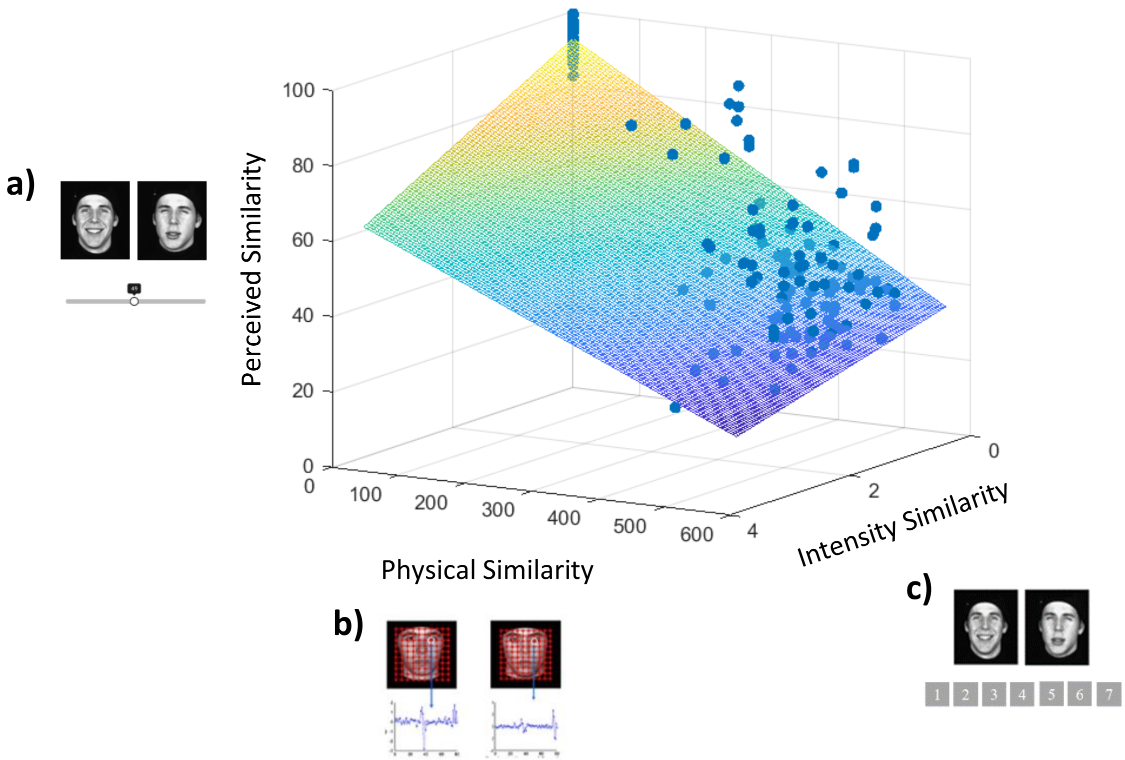

Figure 16. **Multilinear regression model integrating Physical and Intensity similarity to predict participants' perceptual similarity of static facial expressions.** *Each dot represents the similarity measure for one of the 160 pairs of static facial emotions. a) Perceptual similarity obtained from participants responses to a similarity rating task. b) Physical similarity computed using Gabor similarity. c) Intensity similarity was assessed as the absolute difference between rated intensity for the two facial emotions.*

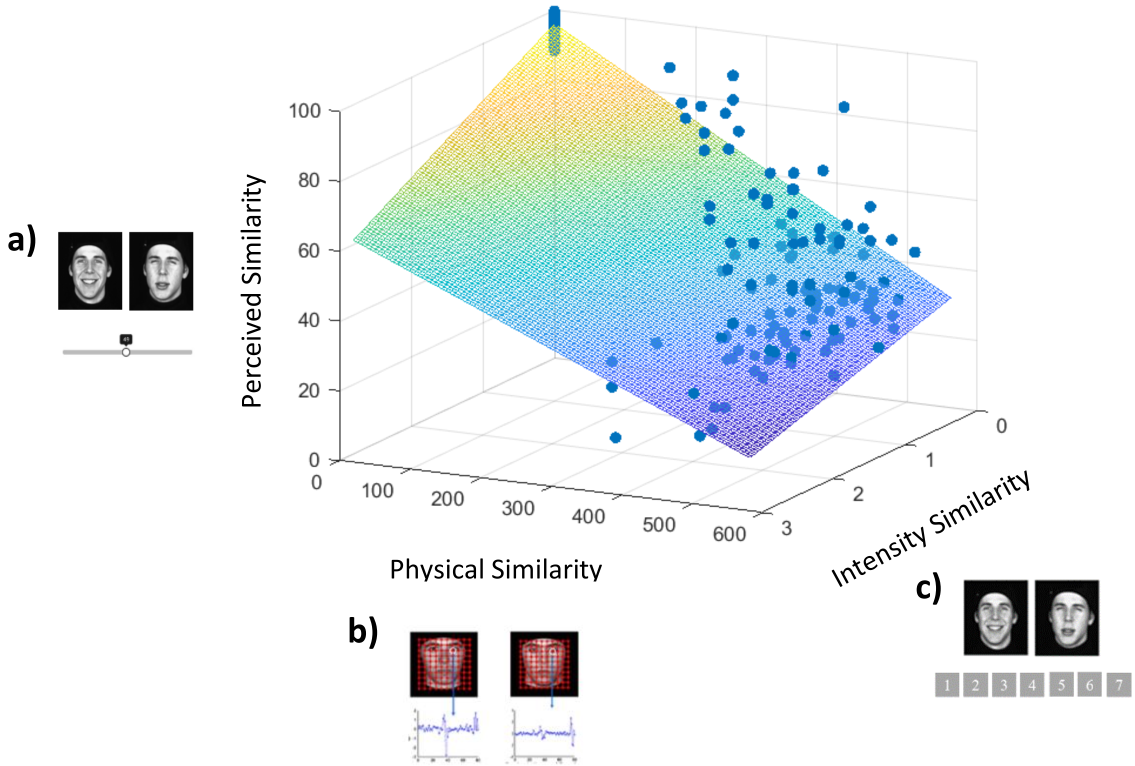

Figure 17. **Multilinear regression model integrating Physical and Intensity similarity to predict participants' perceptual similarity of dynamic facial expressions.** *Each dot represents the similarity measure for one of the 160 pairs of static facial emotions. a) Perceptual similarity obtained from participants responses to a similarity rating task. b) Physical similarity computed using Gabor similarity. c) Intensity similarity was assessed as the absolute difference between rated intensity for the two facial emotions.*

Integrating Physical, Categorical and Intensity similarity provides best prediction to perceptual similarity of Static facial emotions, however, for Dynamic facial emotions Intensity information became redundant

The above results showed that physical information alone well accounted for the variability in perceptual similarity. Nonetheless, its performance significantly improved when either categorical or intensity similarity is added to the model. To further investigate which combination of predictors yield the optimal explanatory power for perceptual similarity, I tested the performance of a model integrating Categorical and Intensity similarity.

For Static facial emotions, the model integrating Categorical and Intensity information accounted for a significant proportion of the variance in perceptual similarity, R^2 adjusted = .699, $F(2, 157) = 154$, $p < .001$. Models' comparisons showed a significant increment in Categorical model's performance when intensity information is added, $\Delta R^2 = .158$, $F(1, 157) = 83.4$, $p < .001$. However, this model still performs lower compared to the physical model alone, R^2 adjusted= .825, $F(1, 158) = 750$, $p < .001$. The best performance is reached when Physical, Categorical and Intensity information are combined, R^2 adjusted= .872, $F(3, 156) = 363$, $p < .00$. Models' comparisons showed a significant increment in Physical model performance when categorical, $\Delta R^2 = .0345$, $F(1, 157) = 38.8$ $p < .001$, and then intensity information is added, $\Delta R^2 = .0142$, $F(1, 156) = 17.6$ $p < .001$. For Dynamic facial emotion, the results were slightly different. The model integrating Categorical and Intensity information accounted for a significant proportion of the variance in perceptual similarity, R^2 adjusted = .658, $F(2, 157) = 154$, $p < .001$. However, in this model the effect of Intensity is not significant, t = .02; p = .98, meaning that adding intensity information to the Categorical model alone does not improve model performance. As a result, the highest explanatory power for Perceptual similarity of Dynamic stimuli is reached by combining categorical and physical information, R^2 adjusted = .868, $F(2, 157) = 524$, $p < .001$.

2.4 General discussion

Facial emotions convey a wealth of subtle cues that guide the way we perceive and make sense of them. In these two studies, I investigated what key factors determining perceptual similarity between static or dynamic facial emotions. I started by examining participants' categorical and dimensional perception of spontaneous dynamic and static facial emotions. Not surprisingly, even though emotions in the current study did not result from posed stereotypical facial expressions, and the different levels of intensity were not defined by manipulating the movement of facial muscles, the results align with previous findings from classical standard tasks (i.e., emotion categorization or intensity rating). Facial expressions could effectively be categorized into three discrete emotion categories in a classic forced choice paradigm and were perceived as conveying high or low levels of intensity in expressions.

Interestingly, a dynamic advantage in emotion recognition was observed, with fewer errors/confusion in the categorization of dynamic than static facial emotions The beneficial effect of dynamic cues has often been detected for suboptimal situations, such as with point-light or blurred stimuli, where information is somehow limited (Krumhuber et al., 2013). However, this advantage seemed to disappear when information was fully available or when emotions displayed high intensity (Bould et al., 2008; Fiorentini & Viviani, 2011). The current results, on the other hand, revealed a dynamic advantage even when facial information was fully available and for both high- and low- intensity emotions. This is further supported by the pattern of confusion observed in the categorization task. While error is more equally distributed between the two wrongly selected categories for static facial emotions, responses to dynamic facial emotions seem to show much less confusion with a more polarized, even if wrong, responses. These results suggest that dynamic cues can influence the processing of non-stereotypical, non-posed elicited facial emotions.

Additionally, an effect of facial motion was observed in the perception of emotion intensity. When participants rated the intensity of facial expressions, static facia emotions were perceived as more intense than dynamic ones. Interestingly, these results deviate from previous findings where dynamic expressions were often reported to elicit higher emotion judgments in terms of intensity, arousal and authenticity (Krumhuber et al., 2013). In a study of Biele and Grabowska (2006), animations of angry and happy faces received higher intensity ratings compared to photographs. This intensity effect is referred to as “representational momentum” and seems to be rooted in the idea that dynamic changes imply a forward shift in the direction of the observed motion (Yoshikawa & Sato, 2008). In Yoshikawa’s and Sato’s work (2008), participants observed short dynamic sequences and were subsequently asked to select the last image they perceived. The final images of dynamic stimuli that participants perceived displayed facial configuration with stronger emotional intensity than the actual presented image. Furthermore, the size of this effect increased with the velocity of the animation. Following on this theory, the discrepancy between our results and previous research might be attributed to differences in the temporal dynamics of posed and spontaneous facial expressions. While posed prototypical expressions often follow the onset-apex-offset model, characterized by a crescendo in the muscle contractions toward the peak, this pattern may not necessarily apply to spontaneous facial expressions. Temporal analyses have indicated that more natural expressions could feature multiple apexes or reach a peak at low intensity, following a more complex development (e.g., onset-apex-onset-apex-offset) (Delannoy & McDonald, 2009). Similarly, it has been shown that the order of appearance and velocity of facial Action Units often differ in spontaneous compared to posed facial expressions (Namba et al., 2017). These findings suggest that commonly assumed properties of posted facial emotions may not necessarily apply to spontaneously elicited facial emotions. Our findings support the view that relying on static, stereotypical posed expressions may limit our understanding of how facial emotions are processed in real-life.

We then investigated the role played by physical, categorical and intensity information in perceptual similarity. For both static and dynamic facial emotions, the stimulus-based physical similarity strongly correlated with their perceived similarity. It is worth noting that the perceptual similarity obtained from both studies was highly consistent, thus ruling out the possibility that exposure to the emotion labels during the categorization task could have influenced participants’ responses to the similarity rating task. Furthermore, given that presenting stimuli sequentially introduced a working memory component to the task, the similarity judgments obtained in these studies should not be generalized and may vary depending on the timing between comparison stimuli.

Consistent with our results, previous research has demonstrated that Gabor similarity in facial stimuli correlates with perceptual similarity of expressions (Xu & Biederman, 2010) and facial movements (Dobs et al., 2014). However, in Xu and Biederman (2010) participants did not directly assess facial expression similarity. Researchers inferred perceptual similarity indirectly by measuring the impact of introducing differences in facial expressions on participants' performance in an identity matching task. Similarly, also Dobs et al. (2014) assessed perceptual similarity indirectly. Participants judged which of two possible approximations, differing in the amount of information about natural facial motion they contained, was perceptually closer to an original animation of a facial emotion. In the present study I adopted a direct measure of perceptual similarity by asking participants to rate the similarity of facial emotions and examined how this measure could be predicted by their physical similarity.

For both static and dynamic stimuli, the categorical similarity correlated with their perceptual similarity. In a study by Brooks et al. (2018), they explored how a perceiver's specific knowledge of emotional concepts dynamically interacts with their perception of facial emotions. Their participants initially rated the similarity of various emotional concepts. Subsequently, participants engaged in a two-choice categorization task where they had to select the right emotion category for a facial expression. Perceptual similarity was computed from the mouse trajectory-deviation toward the unselected category. The researchers found that when a participant's conceptual knowledge about an emotion category overlapped with their knowledge about an alternative category, participant perception was biased, probably due to a co-activation of the two categories, and the hand trajectory deviated toward the alternative category response. Similarly, in the present study, the more a facial emotion was confused, mistakenly categorized, with an alternative emotion category, the closer these emotions were perceived to be.

Lastly, intensity similarity shows correlations with participants' similarity ratings on a trial- level, but not on a group- level or participant-level. However, this trial-level correlation can likely be attributed to the substantial variability in conveying intensity information across different identities (i.e., actors displaying emotions). During the behavioural task, each trial presented two facial expressions from the same identity. These were first rated for their intensity and then for their similarity, involving a total of 9 different identities throughout the task. Facial expressions were gathered by instructing non-professional actors to freely express the emotion they would experience in various every-day scenarios. It appears that intensity similarity scores exhibit a correlation with perceptual similarity when the identity of the actor conveying the emotion remains consistent.

It is important to consider that among the 164 trials presented during the integrated categorization and similarity rating task, there were 40 trials where participants were asked to categorize first and then rate the intensity of identical stimuli. In each trial, the two stimuli were presented sequentially, and each stimulus was categorized first. Despite being recognized as highly similar, identical stimuli were often not rated as identical. Similarly, around 20% of the stimuli were wrongly categorized in the categorization task. This performance highlights how identical non-prototypical facial expressions, when not presented simultaneously, may be challenging to process. Participants' scores to identical stimuli were not excluded from our analyses. While this decision may have slightly inflated the scores resulting from our correlation analysis, we also conducted analyses excluding trials containing identical faces, and the results were not significantly different.

To sum up, the results suggest that perceptual similarity between facial emotions is jointly supported by their categorical, intensity and physical similarity. However, we also found that their contributions are not equivalent. When considering static facial emotions, the integration of Physical, Categorical and Intensity similarity yields the optimal model for explaining variances in perceptual similarity, with Physical similarity making the largest contribution to the model, followed by Categorical similarity and, finally, Intensity similarity. Conversely, for dynamic facial emotions, Intensity similarity appears to be redundant and does not enhance the model's performance when combed with Physical and Categorical similarity.

These findings indicate that the perception of facial emotions is not solely dependent on the emotion concepts attributed to them. Physical and intensity information also play a role in shaping how we perceive differences and similarities between facial emotions. Murray et al. (2021) also found that both stimulus-based cues (i.e., shape and surface) and conceptual information play roles in the perception and categorisation of facial expressions. In their emotion perception task, participants were asked to select, among three faces displaying one of two possible expressions, which one displayed a different expression from the other two (odd one out). They computed the similarity of emotion concepts based on ratings given to 200 short stories in which the characters experienced one of the six basic emotions. They found that emotional concepts were related to behavioural responses in both categorical and perceptual tasks. Physical cues, such as face shape measured using Procrustes analysis, were more closely linked to the perception task than the categorical task, whereas physical cues based on surface textures, computed from the Fisher's Z-transformed Pearson's correlation coefficient between the pixel intensities, were more strongly associated with the categorical task than the perception task. In line with Murray's findings, our participants' scores in a similarity rating task could be explained by both stimuli's physical similarity (i.e., Gabor similarity), which accounted for the majority of the variance in similarity scores, and stimuli's categorical similarity that specifically examined how participants represented the stimuli based on their attributed emotion categories, enhancing the model's performance when integrated with physical information. Moreover, our results also showed that while intensity similarity contributes to improving model performance in explaining participants' responses to static stimuli, the same effect was not observed with dynamic stimuli. This result suggests that the contribution of physical and conceptual information to emotion perception is modulated by facial motion.

Chapter 3

Facial expressions of emotions are perceived as high-dimensional emotional and semantic profiles

3.1 Introduction

Facial expressions have traditionally been assumed to be relatively unambiguous, serving as direct reflections of our internal emotional states and evolving to be easily understood by our conspecifics. According to the BET, in the absence of any conflicting contextual information or extreme stimulus ambiguity, facial expressions of emotion are thought to be directly associated with the perception of discrete emotion categories. However, our daily emotional experiences often prove to be more intricate. Individuals frequently struggle in accurately discerning and articulating their own emotions. When asked to report their mood, they typically do not rely on a single emotion category (Moore & Martin, 2022; Saarni, 1999; Trampe et al., 2015). Emotions are not experienced as isolated and distinct entities; instead, they manifest as complex and interconnected feelings that frequently exhibit ambiguity and overlap.

Unlike classic theoretical models that propose a few affective dimensions or emotion categories to effectively capture how people experience and process emotions, recent research argues that facial expressions typically do not convey a pure, single affective state (Cowen & Keltner, 2020; Moore & Martin, 2022), and emotions are represented within a semantic space that include plentiful of terms that refer to a rich variety of emotional states (Barrett, 2009; Sabini & Silver, 2005; Shaver et al., 1987). Studies that transcend low-dimensional approaches have consistently highlighted the limited explanatory power of traditional models in capturing the richness of real-world experiences (Cowen et al., 2019; Snoek et al., 2023). They have demonstrated that emotional behavior, although still conceptualized in terms of basic emotion categories, is high-dimensional and far more intricate and nuanced than previously believed (Cowen & Keltner, 2020). For instance, Cowen and Keltner (2020) identified up to 28 categories of emotion based on daily facial and bodily expressions, which often finding overlaps with neighboring emotion categories rather than clear distinctions between them.

Facial expressions of emotions convey a wealth of subtle cues that can influence the way we perceive and interpret them, and detecting these subtle cues is crucial for interpretating daily life facial expressions. For instance, positive emotional states - such as interest, pride, pleasure, and joy - subtly diverge in their facial expressions, showing differences in the frequency and duration of several action units (Mortillaro et al., 2011). Even within the same facial expressions, we can identify nuanced differences that shape our understanding of the conveyed emotions. Similarly, quite different facial expressions can be interpreted as conveying the same underlying meaning, depending on specific relevant cues that we perceive. A single, discrete emotion label may not be powerful and sensitive enough to detect these subtle differences, potentially resulting in a loss of valuable social and emotional information in the process of emotion perception.

Furthermore, examining emotional experiences in a complex and nuanced manner can enhance our ability to detect how non-affective factors, such as culture, may shape the perception of facial emotions (Greenaway et al., 2018). For instance, it has been proposed that our emotional experience is rooted in culturally learned emotion concepts (Barrett, 2017; Barrett et al., 2007), and social rules (Matsumoto, Keltner, et al., 2008), which dictate how individuals perceive and communicate emotions. For instance, in some cultures, it may be considered inappropriate to publicly display negative emotions such as anger or sadness (Matsumoto & Hwang, 2013). Cultures that emphasize emotional expressiveness may have more frequent and intense facial expressions of emotion compared to cultures that value emotional restraint, which may have more subdued facial expressions (Matsumoto, Keltner, et al., 2008). Furthermore, decoding rules, learned skills and strategies for recognizing and interpreting emotions based on their expressions may also vary across cultures. Some cultures may rely more on contextual cues, such as situational factors or body language to interpret emotions, while others may place more emphasis on facial expressions alone. These differences can affect the accuracy and reliability of emotional recognition across cultures (Matsumoto, 1993; Matsumoto & Hwang, 2013). As a result, while we might agree on the primary emotion conveyed by a facial expression, differences can arise in how we perceive concurrent emotions from the same facial expression (Fang et al., 2018) or in the way we extract semantic information concurrently activated by emotional content (Jackson et al., 2019; Liu et al., 2022).

Establishing holistic and more comprehensive models of facial emotion perception would not only deepen our understanding of how emotions are experienced and processed in real life but also have practical implications for the technologies based on theoretical understanding of emotion processing. For instance, machine learning technique are widely applied into our daily lives with the intent to extract relevant information from faces (Altameem & Altameem, 2020; Kaushik et al., 2022). Most of these automated systems rely on the prototypicality of facial emotions, often trained on datasets displaying posed and stereotypical facial expressions. These systems generally employ a categorical approach where combinations of facial movements are classified into a defined set of emotion categories (Pantic & Stewart, 2007). As a result, while algorithm-based automatic emotion classification systems (e.g., FaceReader, CERT, FACET) often outperform human performance for prototypical facial expressions (Del Líbano et al., 2018; Lewinski et al., 2014), their performance significantly decrease for non-stereotypical facial expressions (Yitzhak et al., 2017).

In this Chapter, I present two studies that adopt a new research paradigm that emphasizes the complex and blended nature of emotion perception. Specifically, I introduce an emotion profiling task and a semantic profiling task to measure multiple dimensions of emotions and the diverse concepts and meanings associated with the emotional content. These tasks collect rich behavioral responses with the aim to detect emotional blends and to be sensitive enough to uncover influences of factors playing a role in the process such as culture. Differently from previous works, I asked participants to define stimuli along multiple dimensions, instead of a single category, and I translated responses into detailed, multidimensional profiles for each stimulus and participant. A few previous studies have asked participants to rate facial expressions along multiple emotion dimensions, however they usually focus on a few dimensions (e.g., asking participants to indicate the emotions they perceived on the face along three emotion scales) and with the specific aim to test the prediction that a given expression would be rated higher on morphologically similar emotion dimensions than dissimilar ones (Fang et al., 2018), or, when high dimensional responses are collected, these are treated as independent categories with the specific aim to determine how many emotion categories are reliably distinguished in the recognition facial-bodily expression (A. S. Cowen & Keltner, 2020). In my studies, I test whether high dimensional responses may be treated as a whole single vector of information that define the stimulus (i.e., profile), and how high dimensional representations of facial expressions may better explain the way we perceive and process facial emotions.

As for previous studies (Chapter 1), participants completed the tasks online, in their preferred time and environment. Also in this case, it was important to mitigate possible distractions or disengagement by limiting the study duration to one hour. This constraint led me to focus on no more than three distinct emotions for investigation. To allow continuity with my previous studies, I investigate participants responses to 'Happy' and 'Fear', two well-investigated basic emotions (Ekman, 1972), and to 'Pain,' a less investigated emotion which has recently been introduced into the list of 28 distinct basic emotion categories recognized by modern developments of the Basic Emotion Theory (BET) (Cowen & Keltner, 2020).

With these newly designed tasks, I conducted two cross-cultural studies with the following objectives: (1) assess whether an emotion profile, as opposed to a single emotion category, provides a better explanation of how we perceive spontaneous facial expressions of emotions; (2) examine whether an emotion profile can be used to predict perceived differences and similarities between spontaneous facial emotions; (3) investigate whether perception of facial emotions is linked to a broader range of semantic concepts beyond its target emotion category; (4) analyse the roles of facial motion, emotion intensity, and culture in shaping how we profile facial emotions and judge their similarity; and (5) evaluate whether computational algorithms commonly employed for emotion categorization can generate emotion profiles comparable to those formed by human participants.

3.2 Study 3. Emotion profiling: do we perceive more than one target emotion in facial expression of basic emotions?

The primary aim of Study 3 was to investigate whether people perceive complex and rich emotion information from non-stereotypical, spontaneous facial expressions of basic emotions and, if so, whether perceived profiles of rich emotion content could be used to predict the perceptual similarity between facial emotions. Additionally, I explored the role of facial motion, emotion intensity, and participants' culture background in shaping how individuals profile facial emotions and judge their similarity. Finally, I tested whether a machine learning-based algorithm trained to categorize facial emotions could produce emotion profiles comparable to those observed in human responses.

In both studies I used the same stimuli as in Chapter 2, which were specifically selected to convey more natural facial expressions of emotions. To gain a rich and fine-grained understanding of how facial expressions of emotions are perceived, I employed a Profiling task. In this task, participants were asked to rate each facial expression along eight different emotion dimensions. The scores provided for each of these emotion dimensions constituted participants' unique response profiles to each facial emotion (i.e., Emotion profile).

I used the eight basic emotions as the building blocks of the Emotion Profile: Happy, Sad, Fear, Surprise, Anger, Disgust, Neutral and Pain. Recent studies have indeed demonstrated that specific emotion categories provide a more robust representation of emotional experiences, expressions, and neural processing than the traditional valence and arousal dimensions (Cowen & Keltner, 2020). According to Cowen and Keltner (2020), the mechanisms underlying our high-dimensional and often blended emotional experience in daily life are grounded in unique emotion categories. They draw an insightful analogy between emotion and color perception. Just as colors can be broken down into primary color channels—red, green, and blue—our emotional experiences can similarly be understood through fundamental emotion categories. However, different from the BET, they assert that the boundaries between these fundamental emotion categories are not rigidly discrete; instead, they fluidly blend, similarly to how primary colors can be mixed to create a myriad of shades, allowing us to perceive more nuanced emotions. Additionally, much like we can discern various attributes from colors (e.g., whether they are warm or cold), we can infer diverse emotional characteristics from emotion categories, such as valence and arousal. These emotional characteristics are pivotal components of emotional experiences, yet they are less basic and more susceptible to cultural influences(Cowen et al., 2019; Cowen et al., 2020).

Consistent with Cowen and Keltner's view, Du et al (2014), utilizing a Facial Action Coding System analysis, demonstrated that the production of 21 compound emotion categories differs from one another but remains consistent within the subordinate categories in terms of facial actions. Moreover, the automatic categorization of these basic and compound emotions revealed that configural second-order features can serve as superior discriminant measurements of facial expressions of emotions. This suggests that facial emotions are better described using a rich set of basic and compound categories rather than a limited set of basic elements (Du et al., 2014). Similarly, employing a combination of perceptual expectation modeling, information theory, and Bayesian classifiers, Jack et al., (2016) found that dynamic facial expressions of emotion convey an evolving hierarchy of information over time. Initially, dynamic facial expressions transmit four basic emotion categories (i.e., happy, sad, fear/surprise, and disgust/anger), followed by the conveyance of more complex signals that support the categorization of a larger number of categories (i.e., the six basic emotions). Collectively, these findings suggest that we can use basic emotion categories and their blending to better elucidate the way we perceive and understand the complex content extracted from perception of facial emotions.

To investigate whether perceived similarity between facial expressions is driven by perceived emotion profiles I employed a representation similarity analysis, as in the first two studies reported in Chapter 2. I computed Physical, Profiling and Perceptual models of similarity in the form of similarity matrices for static and dynamic facial emotions, and Chinese and British participants. The Physical and Perceptual dissimilarity matrices were obtained following the same procedure as employed in previous studies. For the Profiling similarity between facial emotions, I calculated the cosine distance between the vectorised Emotion

Profiles for two facial emotions. Subsequently I conducted correlation analysis to assess the strength of the relationship between Physical/Profiling similarity and perceptual similarity, for static and dynamic faces, and Chinese and British participants. To evaluate the relative contributions of physical and profile similarity to perceptual similarity, I performed a multiple linear regression analysis. This analysis aimed to estimate the extent to which the Physical and Profiling models of similarities explain variations in participants' perceived similarities and whether their combination yields a better prediction compared to the contributions of individual models.

Finally, to assess whether machine learning-based algorithms for emotion categorization can generate emotion profiles similar to human responses, I trained a computational model for emotion categorization and computed model-based Emotion Profiles. Specifically, I used a Support Vector Machine (SVM), a widely employed prediction method applied to classifications and regressions problems, particularly well-suited for dealing with small training datasets. In fact, while Convolutional Neural Networks (CNN) are state-of-the-art tools for analysing visual imagery, these networks often require a large amount of training samples (Shao & Qian, 2019). In a study by Wang et al. (2021) comparing image classification algorithms based on traditional machine learning and deep learning, it was found that SVM has a better solution effect on small sample data sets, whereas CNN has higher recognition accuracy on large sample data sets. To enhance the performance of my SVM, I applied a feature extraction approach prior to classification, leveraging the representational power of a pre-trained deep neural network (Fei, Yang, Li, Butler, Ijomah, Li, & Huiyu, 2020; Ko, 2018). In particular, I extracted features from both the learning and test images using AlexNet, which is a CNN pretrained on over a million images from the ImageNet database (Russakovsky et al., 2015) and is capable of classifying images into 1000 categories (Krizhevsky et al., 2017). The obtained model-based emotion profiles were then compared to human performance to evaluate whether algorithms commonly employed for facial emotion categorization tasks could generate emotion profiles comparable to those generated by humans.

3.2.1 Methods

Participants

The size of the participants' sample was based on the results from our previous studies, suggesting that a minimum of 13 participants was required. Again, considering the fact that the study was conducted online we collected 20 participants for condition (dynamic/static).
Forty British participants were recruited from the University of East Anglia via the SONA System (5 males, 35 females; age ranged between 18-52 yrs., M = 21.6, SD = 6.88). Participants who did not identify themselves as British in the demographic questionnaire were excluded from the study. Thirty-eight Chinese participants were recruited from the Sun Yat-sen University in China (11 males, 27 females; age ranged between 18-26 yrs., M = 21.1, SD = 2.37). None of these participants had taken part in any previous studies of this project. All participants provided informed consent before taking part in the study and were debriefed at the end, receiving course credits or payment as compensation. The study's experimental procedure was approved by the Ethics Committee of the School of Psychology at UEA.

Stimuli, Materials, and Tasks

The stimuli were the same as used in Studies 1 and 2, including facial images or videos of 9 actors displaying 3 different emotions (Happy, Pain, Fear) at 2 intensities (High, Low) taken from the MPI Facial Expression Database (Kaulard et al., 2012). For the Emotion Profiling task, participants viewed images or videos, depending on the task condition (i.e., Dynamic vs Static), presented individually in a random order for a total of 108 trials (= 3 emotions * 2 intensity levels * 9 actors * 2 repetitions). For the Similarity Rating task, the same 160 pairs of facial emotions used in Study 1 and 2 were presented in a random order.

For the Emotion Profiling Task, each trial started with a fixation cross (1000ms), followed by the stimulus displayed on the left half of the screen (i.e., image for the Static condition, video for the Dynamic condition) and 8 different sliders on the right half of the screen. Participants were asked to rate each facial expression along these dimensions, indicating how much the facial expression displayed each of 8 possible emotions (Happy, Surprise, Sad, Disgust, Neutral, Anger, Fear, and Pain). Participants responded by moving the handle of the sliders (see Figure 18a). Each slider was independent from the others and ranged from 0 (not at all) to 100 (very much). The handles of all sliders were initially placed on 0. To ensure participants had sufficient time to provide their answers without overthinking, preventing the interference of more high-level processes, the response screen had a time-limit of 20 seconds. If the time limit was reached the experiment moved on to the next trial. Dynamic stimuli were presented on a loop until a decision was made or the time limit was reached.

Similarity judgments were collected in a similar way as Study 2. In the Similarity rating task, each trial started with a fixation cross (1000ms) followed by the first stimulus centred on the screen. In the static condition the image was shown for 2000ms, while in the dynamic condition the stimulus lasted either 2000ms or 1000ms, depending on the video length. After the first stimulus was displayed, a second fixation cross (1000ms) was followed by the second stimulus, once again centred on the screen and displayed for the same amount of time. Finally, as in Study 1, the last response screen asked them to “Rate the extent of similarity or dissimilarity between the two facial expressions”. Participants responded by moving the handle of a slider where 0 represented “Totally different” and 100 represented “Exactly the same”. The handle was initially placed on 50 (see Figure 18b).

To adapt the study to our Chinese-speaking participants group, all materials and instructions have been carefully translated by three Chinese native-speaker researchers, who also possessed proficiency in English. Researchers were affiliated with the Sun Yat-sen University in China, and they collaborated on the translation process to ensure comprehensive scrutiny. Their work was supported by the use of dictionaries and reference materials to guarantee accurate translation of key concepts (i.e., emotion categories) and preserve consistency across languages. Despite the careful approach taken in the translation process, it's essential to acknowledge the potential for semantic differences between languages, which can lead to subtle variations in interpretation.

The same measures taken in previous studies were implemented with the aim to reduce, and account for, potential participants’ distractions or disengagement. In particular, (1) access to the study was restricted to PCs or laptops; (3) participants were required to self-report the reliability/usefulness of their data (e.g., due to lacked attention) at the end of the experiment, (4) a time-limit was imposed on each screen, and participants exceeding the maximum allotted time for completing the experiment were excluded from the study; (5) control trials were included in the study, where participants were asked to rate the similarity of identical images.

Procedure

Participants performed the Emotion Profiling and Similarity Rating Task online through the Gorilla platform (i.e., https://gorilla.sc/). They gained access through a URL link using their desktop computer or laptop (i.e., no tablet or phone access was allowed). Once consent was given, participants were directed to a demographic questionnaire. Closed-ended questions asked for their hand dominance (right-handed/left-handed) and the gender they identify with (male/female/other); while open-ended questions asked for their age and nationality. Half of the participants performed the Dynamic task, with movies as stimuli, while the other half performed the Static task, with images as stimuli. Participants completed the Emotion Profiling task first, then, after a 5-minute break, they went through the Similarity Rating Task (see Figure 18). Detailed instructions and three practice trials were given before starting each experimental task. They had a second break of 5-minute halfway through the Similarity Rating Task and the whole study took about 1 hour to complete. At the end of the study, they filled out a self-report questionnaire regarding the reliability of their data and were then debriefed.

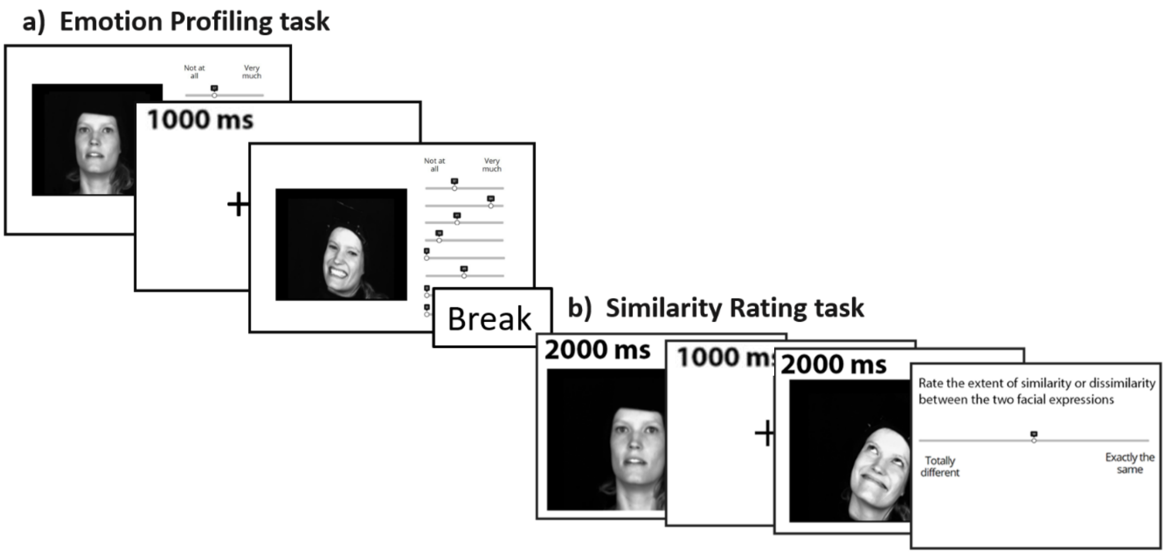

Figure 18. **Example of the Emotion profiling and similarity rating tasks**. *a) In the Emotion profiling task participants rated each facial emotion along 8 emotion dimensions. b) In the similarity rating task, participants saw two facial emotions in a row. Then, using a slider ranging from 0 to 100, they rated the degree of similarity between them.*

Algorithm-based emotion profiling responses

To obtain model-based profiling scores to the facial emotions used in the present study, I employed a Support Vector Machine (SVM) and applied feature extraction before classification to improve classification results (Fei, Yang, Li, Butler, Ijomah, Li, & Huiyu, 2020; Ko, 2018). Specifically, I extracted learned image features from the convolutional neural network AlexNet, which is pretrained on more than a million images of the ImageNet database and can classify images into 1000 categories (Krizhevsky et al., 2017). AlexNet consists of 8 layers, five of these are convolutional layers, some followed by max-pooling layers, and the last 3 are fully conected layers. Through the different layers the network creates a hierarchical representation of the input images, with deeper layers containing higher-level features. I extracted the feature representations of the training and test images using activations on the fully connected layer “fc7” of the net. Then, the feature vector and the feature matrix with training labels were formed. Features extracted from the training images were used as predictor variables and fitted to a SVM using a MATLAB Machine Learning toolbox built-in classifier, fitcecoc (Statistics and Machine Learning Toolbox). I classified the test images using the trained SVM model and the features extracted from the test images. To obtain posterior probabilities of how the SVM classifier responds to each of the eight learned emotional dimensions, in addition to a single label categorization, I used the MATLAB function fitPosterior to fits a score-to-posterior-probability transformation function to the scores. The transformation function computes the posterior probability that an image can be classified into each emotion category. To fit the requirement of the AlexNet (i.e., size: 227*227 pixels, number of color channels: 3) I pre-process our images creating an augmented image datastores, I specified the image size, and replicated the channel three times to convert our single channel grayscale pictures to three channels.

To ensure a balanced distribution of training images across the 8 emotion categories, I constructed the training set by combining stimuli from three datasets: the Karolinska Directed Emotional Faces (KDEF) (Goeleven et al., 2008; Lundqvist et al., 1998), the Delaware Pain dataset (Mende-Siedlecki et al., 2020), and the MPI Facial Expression Database (Kaulard et al., 2012). The resulting training set comprised a total of 560 images, with 70 images for each emotion category. As the testing images were sourced from the MPI Database, to avoid the same actor appearing simultaneously in the testing and training set, which might bias the results, I employed the leaving-one-out method during testing. Specifically, I used only one actor as a test subject at a time, while the remaining identities were incorporated into training. By running nine tests, one for each actor, I ensured that no image or identity appeared simultaneously in the testing and training sets.

3.2.2 Results and Discussion

Similar to Study 1 and 2, participants were not included in the data analysis if they failed to respond to more than 50% of the trials in the emotion profiling task, rated over half of identical facial emotions as below 50 (on a 100-point scale for similarity rating) or reported data unreliability in the final questionnaire. Following the application of these criteria, one British participant was excluded, resulting in a final sample of 39 British participants and 38 Chinese participants for subsequent data analysis.

Facial emotions are perceived as profiles of mixed basic emotions

To assess whether participants perceived facial expressions of emotions as a single, discrete emotion category or as a profile of mixed emotions, I first examined whether the performance on the Emotion Profiling task exhibited a dominant response for the target emotion, accompanied by near-zero scores for the other dimensions, or a more diverse and mixed profile with multiple prominent emotional dimensions. To gain an overview for each of the three facial emotions, participants' responses were averaged across culture (i.e., participants group), facial motion (Static/Dynamic), and emotion intensity levels (Low/High) (for details of emotion profiles before averaging, see Appendix B, Figure B.1). Results revealed that participants seemed to employ multiple emotion dimensions to interpret facial expressions of Happy, Fear, and Pain (see Figure 19a). Facial expression of Pain was characterized by a higher score on the target dimension (33%, SD = 13.8, paired t-test comparing target to other dimensions, all ts(76) ≥ 2.86, all ps ≤ .005), and co-occurring high scores in other dimensions such as Disgust (28%, SD = 13), Fear (18%, SD = 12.6), and Surprise (17%, SD = 12.1). For facial expression of Fear, we found a higher score for Surprise (46%, SD = 14.1) than the target emotion of Fear (24%, SD = 13.8). Paired t-test showed that response to Fear was higher than all other dimensions, all ts(76) ≥ 3.79, all ps <.001, except for Surprise (t(76) = 12.37, p < .001). Also in this case, there were co-occurring high scores to other dimensions such as Disgust (18%, SD = 10.2), and Neutral (15%, SD = 10.3). Finally, facial expressions of Happy showed a slightly different profile compared to Pain and Fear. While a higher score to the target emotion Happy (52%, SD = 14.2) was observed, all ts(76) ≥ 18.9, all ps < .01, responses to other dimensions were relatively low (e.g., 15% Neutral, SD = 9.48; 9% Surprise, SD = 9.26; 8% Disgust, SD = 8.94).

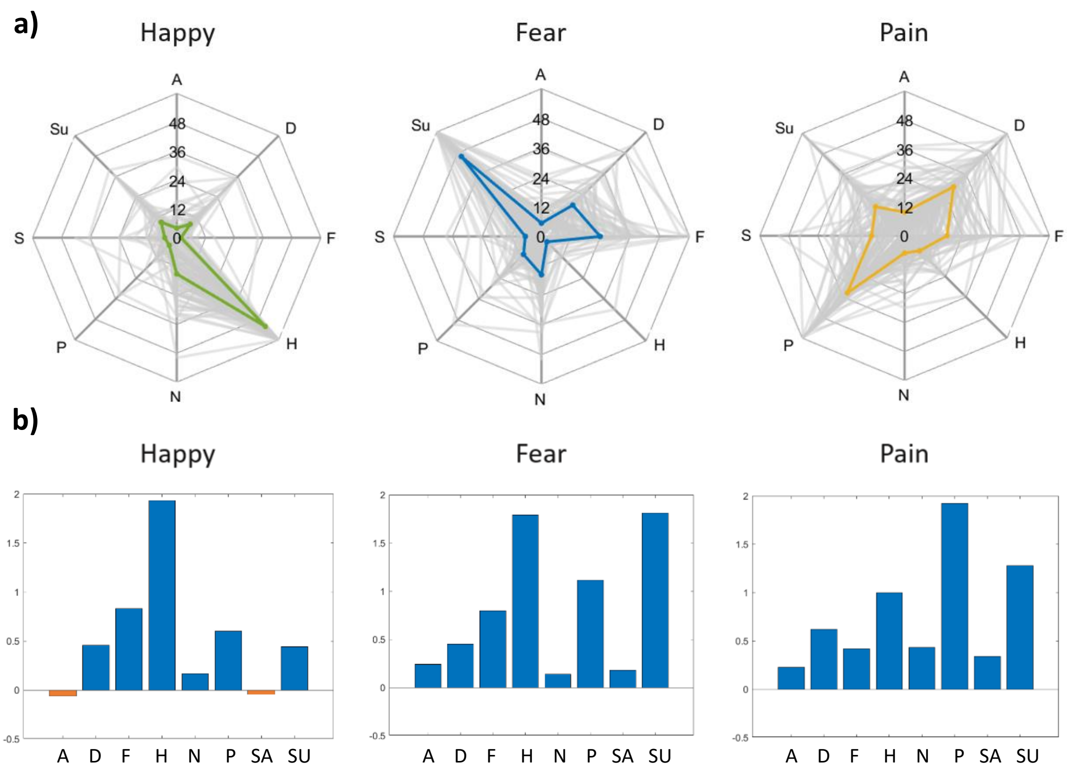

Figure 19. **Participants' Emotion Profiles of Happy, Fear and Pain and the relative importance of individual emotion dimensions in differentiating these emotion profiles.** *a) Emotion profiles. Colored lines are averaged profiling responses in percentage along 8 dimensions: Angry (A), Disgust (D), Fear (F), Happy (H), Neutral (N), Pain (P), Sad (SA), and Surprise (SU). Gray lines represent emotional profiles of individual participants, where scores higher than 60 are truncated for ease of visualization. b) Feature importance of individual emotion dimensions, which represents the relative contribution of each emotion dimension with arbitrary units of measurement. Negative values (shown in orange) indicates that the random permutation worked better than the original feature, suggesting that the specific feature does not have a role in the model's prediction.*

To further assess how individual emotion dimensions contribute to emotion perception, I trained a Random Forest machine learning model to classify the observed emotion profiles into the three facial emotions, and I applied a feature importance algorithm (i.e., Out-of-Bag feature importance) to highlight relative importance of the dimensions, or features, that contribute the most to the model's predictions. The Out-of-Bag (OOB) feature importance measure is estimated based on the decrease in prediction accuracy when the values of a particular feature are randomly permuted while keeping other features unchanged. In particular, for each feature, it measures the average increase in mean squared error across all trees in the ensemble and divides it by the standard deviation taken over the trees. The value obtained represents how much the prediction error (MSE) increases on average when the values of a particular variable are randomly permuted. Thus, a higher value indicates that the specific feature is much more crucial for accurate predictions, as permuting it leads to a larger decrease in the model's performance. To provide a more robust and representative estimation of the importance of each dimension I ran the feature importance analysis 10 times and averaged the results, this process helps to mitigate the potential impact of random variations in the model training process. The rankings across different iterations remained relatively stable, suggesting the consistency and reliability of the ranking on the importance of individual emotion dimensions (all SDs ≤ 0.19).

As shown in Figure 19b, analysis of feature importance showed that multiple emotion dimensions contributed to the model's ability to differentiate between emotion profiles, emphasizing the significance played by diverse dimensions in shaping participants' responses. For facial expressions of Pain, while Pain is in fact detected as the most important dimension to differentiate its profiles from the others (1.9), other co-occurring dimensions such as Surprise (1.3), Happy (1), and Disgust (0.6), also show high contribution to the model’s performance to differentiate it from other emotion profiles. The dimension of Anger has, instead, a close to zero score (0.2), suggesting its weak role in the model’s predictions. For facial expressions of Fear, Surprise is detected as the most important dimension (1.8), followed by Happy (1.7), Pain (1.1), and Fear (0.8), while dimensions of Anger, Sad and Neutral have all close to zero values (for Anger and Sad 0.2, for Neutral 0.1), denoting their weaker role to the model’s predictions. Finally, for facial expressions of Happy, higher score for the Happy dimension (1.9), are followed by Fear (0.8), and Pain (0.6), while, again, dimensions of Anger (-0.04) and Sad (-0.03) are both negative, suggesting their irrelevance to the model’s predictions. Note that higher importance of emotion dimensions is not necessarily characterized by higher scores in participants' responses. For instance, while both Happy (1.2) and Fear (0.9) were important dimensions for predicting emotion profiles of Fear, only 1.9% Happy responses were observed in the profiles of Fear. This is because the specific emotion profile that characterizes participants' responses to facial expression of Fear has constantly lower scores in Happy dimension compared to the emotion profiles in response to Happy and Pain.

Emotion Profiles are sensitive to Facial Motion, Emotion Intensity and Culture

Next, I tested whether emotion profiles were affected by facial motion (i.e., dynamic vs static stimuli), emotion intensity (i.e., high vs low) and participants' cultural background (i.e., Chinese vs British participants). To do so, I contrasted emotion profiles between corresponding conditions. For example, to obtain contrast profiles for Culture, I averaged the profiling responses to the three facial emotions across emotion intensity and facial motion conditions see Figure 20a (for details of emotion profiles before averaging, see Appendix B, Figure B.1). The obtained profiles for each facial emotion (i.e., Happy/Fear/Pain) were then submitted to a 2 (British vs Chinese / Dynamic vs Static / High vs Low intensity) by 8 (emotion dimensions) ANOVA. As shown in Figure 20, despite overall similarity across dynamic/static facial emotions, high/low emotion intensity, and British/Chinese participants, emotion profiles were also sensitive to these factors and shown some notable significant differences.

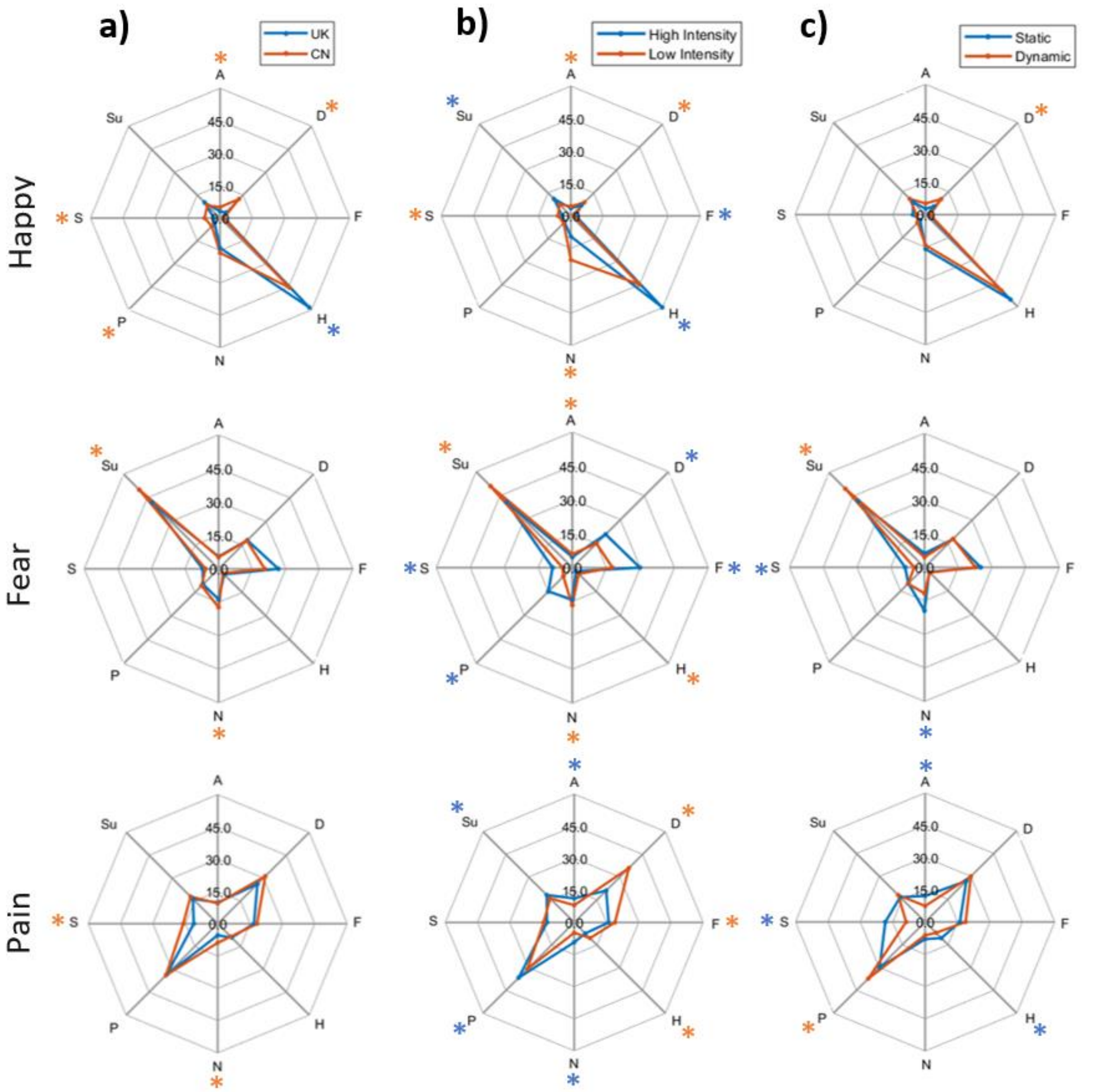

Figure 20. **Emotion-profiles for Happy, Fear and Pain as a function of Culture background, Facial Motion and Emotion Intensity.** *a) Responses of British, in blue, and Chinese, in orange, participants. b) Participants' responses to High and Low intensity facial expressions of emotions. c) Participants' responses to dynamic and static facial expressions of emotions. Asterisks indicate significant differences.*

British vs Chinese participants' responses. For facial expressions of Happy (see Figure 20a, top panel), the main effect of emotion dimensions was significant, $F(7,525) = 372.7$, $p < .001$, $\eta p^2 = .832$. This indicates that participants provided significantly different scores for the eight emotion dimensions. There was no significant difference in overall responses (across emotion dimensions) between British vs Chinese participants, $F(1,75) = 2.02$, $p = .159$, $\eta p^2 = .026$, while, more importantly, there was a significant interaction between Culture and emotion dimensions, $F(7,525) = 14.4$, $p < .001$, $\eta p^2 = .162$. Post-doc independent t-tests for each emotion dimension revealed significant differences between Chinese vs British participants' responses to the dimensions of Happy ($t(75) = 4.39$, $p < .001$), Anger ($t(75) = 2.34$, $p = .022$), Disgust ($t(75) = 5.53$, $p < .001$), Pain ($t(75) = 2.25$, $p = .027$), and Sad ($t(75) = 3.51$, $p < .001$). Their responses to other dimensions did not differ significantly (all ts$(75) \leq .84$, all ps $\geq .180$). Compared to British participants' responses, Chinese participants perceived less extent of Happy, and higher level of Pain, Disgust, Sad and Anger, providing a more diverse interpretation of happy expressions.

Facial expressions of Fear (see Figure 20a, middle panel) and Pain (see Figure 20a, bottom panel) showed a significant main effect of emotion dimension [for Fear, $F(7,525) = 243.23$, $p < .001$, $\eta p^2 = .764$; for Pain, $F(7,525) = 89.90$, $p < .001$, $\eta p^2 = .545$,], and both showed no significant difference between the two groups (across emotion dimensions) [for Fear, $F(1,75) = .38$, $p = .537$, $\eta p^2 = .005$; for Pain, $F(1,75) = 1.97$, $p = .165$, $\eta p^2 = .029$]. However, there was a significative interaction between Culture and emotion dimension for Fear, $F(7,525) = 4.43$, $p < .001$, $\eta p^2 = .056$, but not for Pain, $F(7,525) = 1.34$, p = .227, $\eta p^2 = .018$. In facial expressions of Fear, compared to English participants, Chinese participants tended to perceive less extent of Fear ($t(75) = 1.89$, $p = .06$) while detecting higher levels of Neutral ($t(75) = 1.92$, $p = .05$) and Surprise ($t(75) = 2.33$, $p = .023$). No significant difference was observed for other dimensions (all ts$(75) \leq 1.90$, all ps $\geq .062$). Also in facial expressions of Pain, Chinese participants perceived higher levels of non-target emotions like Neutral

(t(75) = 3.27, p = .002) and Sad (t(75) = 2.44, p = .017) and showed similar responses to British participants regarding other dimensions (all ts(75) ≤ 1.61, all ps ≥ .110).

High VS Low intensity facial emotions. As shown in Figure 20b, For all three emotions, I found a significant main effect of emotion dimension [for Happy, $F(7, 532) = 318.12$, $p < .001$, $\eta p^2 = .807$; for Fear, $F(7, 532) = 232.4$, $p < .001$, $\eta p^2 = .754$; for Pain, $F(7, 532) = 89.51$, $p < .001$, $\eta p^2 = .541$], a significant main effect of emotion intensity [for Happy, $F(1,76) = 6.52$, $p = .013$, $\eta p^2 = .079$; for Fear, $F(1,76) = 67.1$, $p < .001$, $\eta p^2 = .046$; for Pain, $F(1,76) = 6.67$, $p = .012$, $\eta p^2 = .081$], and more importantly, a significant interaction between Intensity and emotion dimension [for Happy, $F(1,532) = 107.79$, $p < .001$, $\eta p^2 = .058$; for Fear, $F(1,532) = 63.5$, $p < .001$, $\eta p^2 = .045$; for Pain, $F(1,532) = 44.50$, $p < .001$, $\eta p^2 = .369$]. Paired t-tests showed that for facial expressions of Fear, responses differed across all eight emotion dimensions (all ts(75) ≥ 2.30, all ps ≤ .024). For facial expressions of Happy, responses differed significantly for all the emotion dimensions except for Pain (all ts(75) ≥ 1.96, all ps ≤ .05, except for Pain, $t = 1.57$, $p = .121$). For facial expressions of Pain, all the emotion dimensions differed significantly except for Sad (all ts(75) ≥ 2.44, all ps ≤ .01,except for Sad, $t(75) = 1.60$, $p = .114$). In Figure 20b are reported the directions of the differences for each emotion displayed and labels. In particular, when a blue asterisk is reported on top of the specific label, scores were higher for the high intensity emotions. When a orange asterisk is reported on top of the specific label, scores were higher for the low intensity emotions. Responses to the target emotion dimensions (e.g., Fear dimension in facial expression of Fear) were consistently higher for high intensity emotions than low intensity emotions.

Dynamic VS Static facial expressions. As shown in Figure 20c, for all the three emotions, there was a significant main effect of emotion dimension [for Happy, $F(7, 525) = 326.71$, $p < .001$, $\eta p^2 = .813$; for Fear, $F(7, 525) = 251.6$, $p < .001$, $\eta p^2 = .770$; for Pain $F(7, 525) = 98.78$, $p < .001$, $\eta p^2 = .563$]. The main effect of facial motion was not significant [for Happy, $F(1,75) = 0.032$, $p = .858$, $\eta p^2 = 0$; for Fear, $F(1,75) = 0.68$, $p = .411$, $\eta p^2 = .009$; for Pain, $F(1,75) = 0.17$, $p = .677$, $\eta p^2 = .002$]. The interaction between facial motion and emotion dimension was significant [for Happy, $F(7, 525) = 2.77$, $p = .008$, $\eta p^2 = .036$; for Fear, $F(7, 525) = 8.18$, $p < .001$, $\eta p^2 = .098$; for Pain, $F(7, 525) = 8.81$, $p < .001$, $\eta p^2 = .105$].

Independent t-tests showed that for facial expressions of Happy, Static vs Dynamic faces differed for the dimension of Disgust ($t(75) = 2.09$, $p = .040$), but not other dimensions (all ts(75) $\leq$ 0.43, all ps $\geq$ .082). For facial expressions of Fear, Static vs Dynamic faces differed for the dimensions of Neutral ($t(75) = 4.28$, $p < .001$), Sad ($t(75) = 3.26$, $p = .002$), and Surprise ($t(75) = 2.86$, $p = .005$), but not other dimensions (all ts(75) $\leq$ 1.70, all ps $\geq$ .094). Finally, for facial expressions of Pain, significant difference was found for the dimensions of Anger ($t(75) = 2.77$, $p = .007$), Happy ($t(75) = 3.24$, $p = .002$), Pain ($t(75) = 2.65$, $p = .010$), and Sad ($t(75) = 4.83$, $p < .001$) but no other dimensions (all ts(75) $\leq$ 1.80, all ps $\geq$ .076). In Figure 20b are reported the directions of the differences for each emotion displayed and labels. In particular, when a blue asterisk is reported on top of the specific label, scores were higher for the static emotions. When an orange asterisk is reported on top of the specific label, scores were higher for the dynamic emotions. Therefore, while these results demonstrate that emotions profiles are sensitive to facial motion, I did not observe consistent dynamic advantage for perceiving the target emotion categories in facial expressions of emotions.

Emotion profiles predict perceptual similarity between facial expressions of emotions

The above results revealed that facial expressions of emotions are perceived as emotion profiles of mixed emotion categories. Next, I tested whether the similarity between emotion profiles could predict participants' perceptual similarity obtained from the similarity rating task. To do so, I first computed profile similarity for each of the 160 pair of stimuli presented in the similarity rating task. Profile similarity was measured as the Cosine distance between two vectorized response profiles. The Cosine distance is a well-established measure used for quantifying similarity between vectors in a multidimensional space. It computes the cosine of the angle between two vectors, providing a value ranging from -1 to 1, where 1 indicates that the vectors are identical and -1 indicates that they are opposite to each other. Profile similarity for all 160 trials were then averaged across each combination of facial emotions (i.e., High and Low intensity Happy, Fear and Pain), resulting in a similarity matrix. This procedure was repeated for each participant, resulting in 77 profile similarity matrices. These were then averaged across cultures (i.e., British vs Chinese participants) and facial motion (i.e., Dynamic vs Static) resulting in 4 matrices of profile similarity between facial emotions (Figure 21b). Participants' responses to the similarity rating task were also averaged across each combination of facial emotion (i.e., High intensity and Low intensity facial expressions of Happy, Fear and Pain) to produce a perceptual similarity matrix for each participant. Again, these were then averaged across cultures (i.e., British vs Chinese participants) and stimuli type (i.e., Dynamic vs Static stimuli) resulting in 4 matrices of perceptual similarity. My objective was to determine the correlation between profile similarity and perceptual similarity (Figure 21a). Since the direct perceptual similarity matrix lacks symmetry along the diagonal line, which has similar but not identical values, the mirrored cells in the upper and lower triangles were averaged. Finally, I computed the correlation coefficient using the upper triangle of the matrices.

Results showed a strong correlation between profile similarity (Figure 21b) and the corresponding Perceptual similarity (Figure 21a), all rs > .88, all ps < .001 (For British, Dynamic, $r = .91$, $p < .001$, Static, $r = .98$, $p < .001$; for Chinese, Dynamic, $r = .96$, $p < .001$, Static: $r = .88$, $p < .001$). This result suggests that, on a group-level, emotion profile can predict perceptual similarity between facial emotions. To assess the consistency of these findings across participants, I analysed participant-level data by calculating the correlation coefficient between each participants' profile similarity matrix and their corresponding perceptual similarity matrix. The results are presented in Figure 22. All British participants who performed the Dynamic task, exhibited a significant correlation between profile and perceptual similarity (all rs ≥ .60, all ps ≤ .01). Similar results were observed for Chinese participants who performed the Dynamic task (all rs ≥ .55, all ps ≤ .03). Similarly, all British and Chinese participants who completed the Static task, showed significant correlations, except for one British participant and two Chinese participants (for British, all rs ≥ .60, all ps ≤ .01, except for one, $r = .48$, $p = .066$; for Chinese, all rs ≥ .52, all ps ≤ .04, except for two, $r = .47$, $p = .073$ and $r = .22$, $p = .421$).

Interestingly, correlation coefficients seemed to be higher in the Dynamic conditions compared to the Static condition for both cultures. To explore this further, a 2 (British/Chinese) by 2 (Static/Dynamic) ANOVA analysis was conducted on the correlation coefficients obtained. The analysis revealed a significant main effect of facial motion (i.e., Dynamic vs. Static), $F(1, 73) = 11.86$, $p < .001$, with higher correlation coefficients for Dynamic than Static conditions. There was a marginally significant effect of culture, $F(1, 73) = 3.94$, $p = .05$, with higher correlation coefficients for British than Chinese participants. The interaction between Culture and facial motion was not significant, $F(1, 73) = 0.836$, $p = .364$. These results suggest that the association between profile and perceptual similarity is stronger for dynamic than for static facial emotions, and for both British than Chinese participants.

Finally, I performed a correlation analysis on all the 160 pairs of facial emotions. Results can be found in the Appendix B, Figure B.2. In line with previous analysis, profile and perceptual similarity were highly correlated at trial level, for both cultures and both dynamic and Static tasks (all rs $\geq .87$, all ps $< .001$).

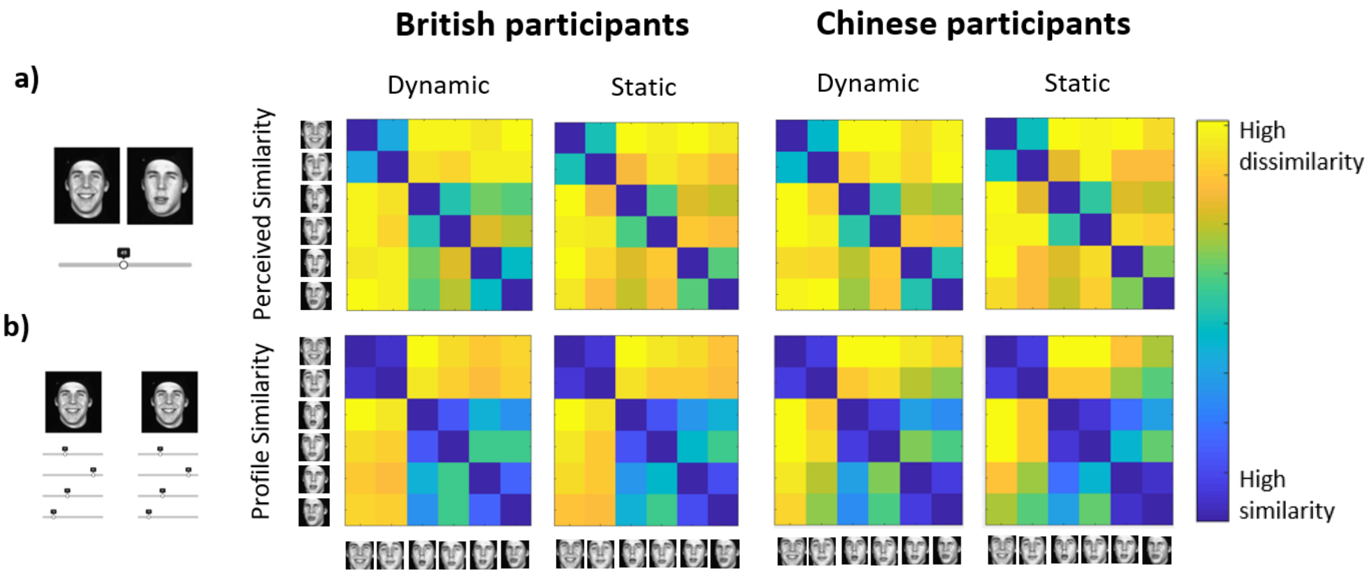

Figure 21. **Matrices of Perceptual similarity and Profile similarity as a function of facial motion and participants' cultural background.** *British and Chinese participants' perceptual similarity matrices highly correlate with their respective Profile Similarity matrices, for both dynamic and static facial emotions. a) Perceptual similarity obtained from participants' responses to the Similarity rating task. b) Profile similarity computed as the cosine distance between participants' vectorized responses to the emotion profiling task. Along the diagonal are values obtained comparing profiles in response to the same facial emotions (i.e., same emotion and same emotion intensity). Facial expressions from top to bottom and from left to right are high intensity Happy, low intensity Happy, high intensity Fear, low intensity Fear, high intensity Pain, and low intensity Pain.*

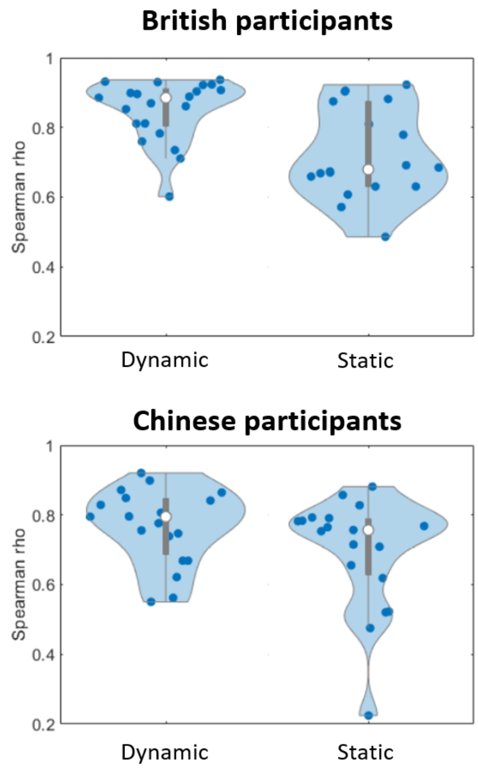

Figure 22. **Correlation coefficients between Perceptual and Profile similarity matrices for individual participants.**

These results provide compelling evidence that emotion profiles of individual facial emotions can predict their perceptual similarity across cultural backgrounds and stimulus types. These findings are supported on a group-, participant-, and trial-levels, indicating that participants may utilize emotion profile information to make judgments regarding the similarities between facial expression of emotions. Interestingly, a participant-level analysis revealed that facial motion seems to enhance the use of emotion profile information in similarity judgments, and that British participants tend to integrate this profile information more in their similarity judgments than Chinese participants. Finally, as shown in Figure 21, for perceptual similarity, pattern of responses seemed to be more different across facial motion than culture. Responses of British and Chinese participant are highly similar, indicating that perceptual similarity judgments are not heavily influenced by culture. However, profile similarity matrices showed an opposite tendency. The pattern of responses seemed to be more different across culture than facial motion. Thus, cultural differences seem stronger in the profile similarity than in perceptual similarity. Further analysis is needed to validate these observations.

Human and algorithm-based responses produce different emotion profiles

To test whether state-of-the-art algorithms commonly employed for facial emotion categorization can produce emotion profiles similar to human responses, I trained a Support Vector Machine to categorize emotions using a pretrained convolutional neural network (i.e., Alexnet, for details see Methods section). The trained model achieved a 67% accuracy in categorizing the facial emotions used in the present study. Subsequently, it was employed to compute the posterior probability of classifying a facial expression into each of the eight emotion categories, thereby generating an emotion profile.

I then explored the differences and similarities between human vs model-based emotion profiles (see Figure 23). For Happy, both human and model responses showed higher scores to the target dimension Happy (participants, 66% for high and 44% for low; model, 50% for high and 50% for low). However, participants attribute higher scores to the dimensions of Surprise (11% for high, and 7% for low) compared to the model (0.1%, for high, and 0.3% for low) and again to the dimension of Neutral (participants, 9% for high, and 20% for low; model, 4% for high, and 4% for low). The model attribute higher scores to the dimensions of Fear (12% for high, and 19% for low) compared to humans (2% for high, and 1% for low) and again to the dimension of Pain (model, 24% for high and 18% for low, participants, 4% for high and 4% for low).

For Fear, both human and model responses showed high scores to the target dimension Fear (participants, 30% for high and 17% for low, model, 42% for high and 45% for low,). However, participants gave the highest scores to Surprise (41% for high, and 51% for low), while for the model Fear was the dimension with higher scores, and no considerable scores were given to Surprise (2% for high, and 2% for low). Furthermore, compared to participants the model attribute higher scores to Happy and Pain (for Happy, model, 8% for high and 22% for low, participants, 2% for high and 3% for low; for Pain, model, 28% for high and 16% for low, participants, 15% for high and 5% for low), while participants attribute higher scores to Anger and Disgust (for Anger, participants, 4% for high and 6% for low, model, 0.1% for high and 0% for low; for Disgust, participants, 21% for high and 15% for low, model, 2% for high and 1% for low).

Finally, also for Pain, while higher level of the target dimension Pain was perceived by both the participants (40% for high, and 57% for low) and the model (37% for high, and 31% for low), participant seemed to perceive higher levels of all the other emotion dimensions compared to the model (for Anger, participants, 11% and 8%, model, 0.4% and 0.6%; for Disgust, participants, 21% and 36%, model, 7% and 5%; for Sad, participants, 12% and 14%; model, 5% high, 2% low; for Surprise, participants, 18% and 16%, model, 0.5% and 0.2%, respectively for high and low intensity), except for Happy where the model attributed higher score compare to the participants (model, 19% for high, and 19% for low, participants,7% for high, and 10% for low).

Finally, to assess the overall similarity between human and model-based emotion profiles, I performed a correlation analysis between each human vs model response vectors. Results showed no correlation between human vs model emotion profiles for Fear (for high intensity, $r = .23$, $p = .57$; for low intensity, $r = .13$, $p = .75$) and Pain (for high intensity, $r = .63$, $p = .090$; for low intensity, $r = .42$, $p = .29$). However, there was a significant correlation for Happy (for high intensity, $r = .84$, $p = .009$; for low intensity, $r = .75$, $p = .03$). Together, these results suggest that while human and model-based responses similarly detect the target emotion dimensions, their perceived emotions profiles differ significantly, especially for Pain and Fear.

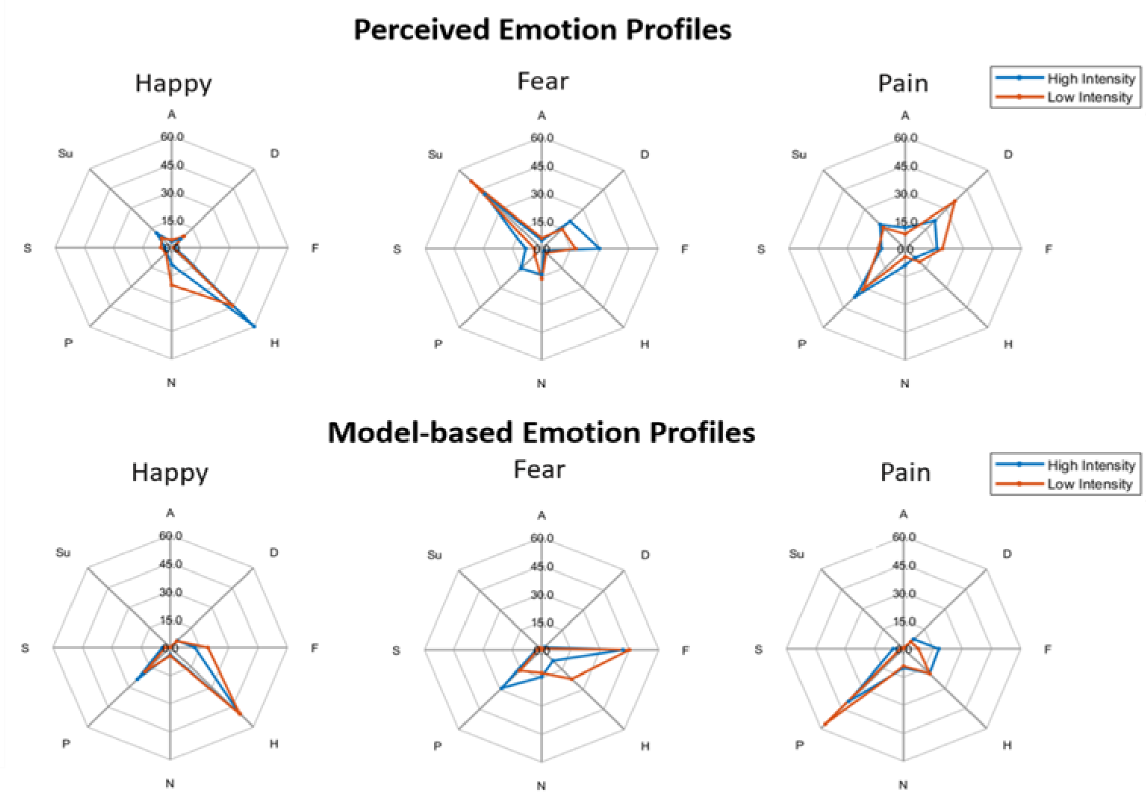

Figure 23. **Participants' and Model-based Emotion Profiles of Happy, Fear and Pain.** *On the top row, are participants' averaged profiling responses in percentage to high and low intensity facial expressions of Happy, Fear and Pain. On the bottom row, emotion profiles generated by a Support Vector Machine for classification. The 8 emotion dimensions are: Angry (A), Disgust (D), Fear (F), Happy (H), Neutral (N), Pain (P), Sad (SA), and Surprise (SU).*

Following the same procedure used to obtain participants' profile similarity (see Figure 24a), model-based emotion profiles were used to compute the profile similarity (i.e., Cosine similarity) between pairs of facial emotions for individual trials. As a result, I computed a model-based profile similarity matrix (see Figure 24c). I then calculated the correlation between model-based similarity matrix and participants' perceived and profile similarity. As for previous analysis, I extracted the upper triangle of the matrices and averaged the mirrored cells in the upper and lower triangles of the perceptual similarity matrix. Results showed no correlation between Model-based profile similarity and perceptual or profile similarity (for British perceptual similarity, r = .30, p = .28, for British profile similarity, r = .34, p = .21; for Chinese perceptual similarity, r = .41, p = .13, for Chinese profile similarity, r = .44, p = .10). Same results were found at a participants-level. For British participants, no participant showed significant correlation between their perceptual/profile similarity and corresponding model-based profile similarity (for perceptual similarity, all rs ≤ .48, all ps ≥ .07; for profile similarity, all rs ≤ .49, all ps ≥ .07). For Chinese participants, the results were similar with only a few exceptions (for perceptual similarity, all rs ≤ .46, all ps ≥ .09, except for five participants, all rs ≥.51, all ps ≤ .05; for profile similarity, all rs ≤ .49, all ps ≥ .07, except for two participants, both rs ≥ .55, ps ≤ .03).

Physical similarity correlates with perceptual similarity and emotion profile similarity

To test whether Physical, image-based, similarity is able to explain participants' perceived and emotion profile similarity, I performed a correlation analysis between the physical similarity matrix (obtained in Study 2 based on Gabor similarity) and participants' perceived and profile similarity (see Figure 24b). Following the same procedure of previous analysis, I averaged the mirrored cells in the upper and lower triangles of the perceptual similarity matrix and the Physical similarity matrices, then I extracted the upper triangle of the matrices and compute the correlations. For British participants, results showed a significant correlation between Physical similarity and participants' Perceptual similarity (r = .61, p = .016) and Profile similarity (r = .68, p = .005). Same results were found for Chinese participants, showing a significant correlation between Physical similarity and both Perceptual similarity (r = .68, p = .005) and the Profile similarity (r = .74, p = .002).

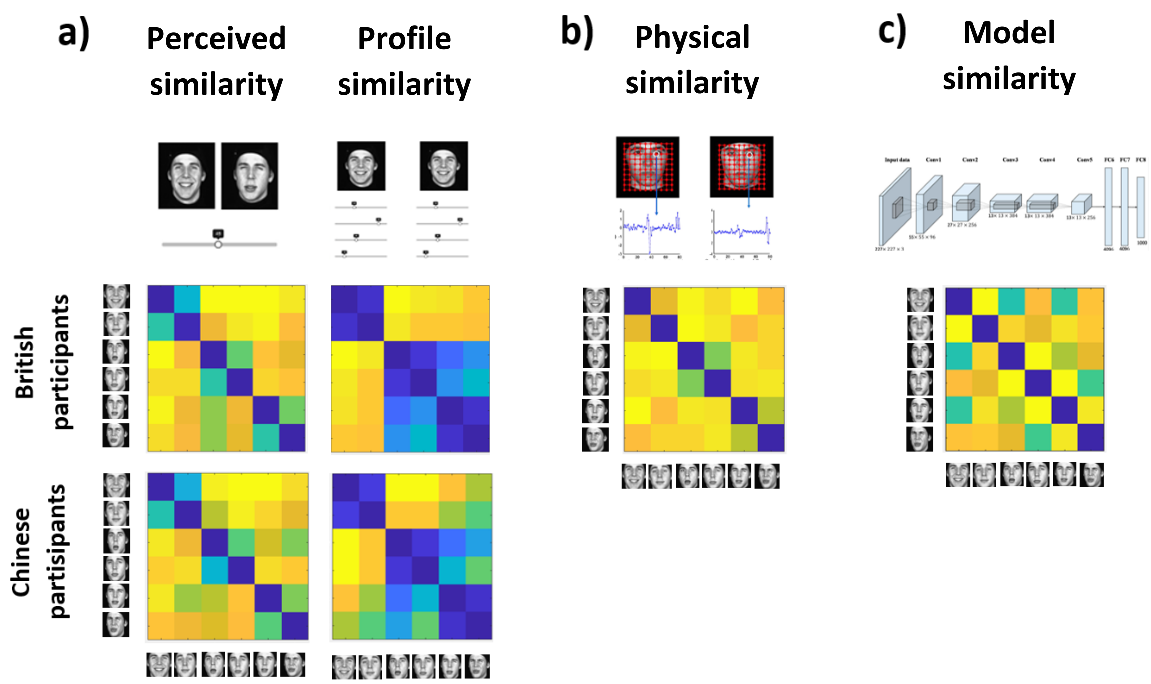

Figure 24. **Perceptual, Profile and Physical similarities between facial expression of emotions.** *Starting from the top row of each matrix are scores in response to, respectively, stimuli depicting high intensity Happy, low intensity Happy, high intensity Fear, low intensity Fear, high intensity Pain, low intensity Pain. a) Perceptual and profile similarity matrices for British and Chinese participants, generated from their response to the similarity rating task (left) and the emotion profiling task (right). b) Physical similarity matrix computed based on the Gabor similarity measure between each pair of facial emotions. c) Model-based profile similarity computed from emotion profiles generated by a Support Vector Machine classifier trained for emotion categorization.*

I also analysed data on a participant-level to assess the consistency of these findings. Results are presented in Figure 25. Perceptual similarity correlates with Physical similarity for all but five British participants (all *r*s ≥ .50, all *p*s ≤ .05; except five participants, all *r*s ≤ .49, all *p*s ≥ .06). Profile similarity correlates with Physical similarity for all but two British participants (all *r*s ≥ .51, all *p*s ≤ .05; except for two, both *r*s ≤ .45, both *p*s ≥ .09). Similarly, for Chinese participants, all but two participants show significant correlation between Perceptual similarity and Physical similarity (all *r*s ≥ .51, all *p*s ≤ .05; except for two participants, both *r*s ≤ .50, both *p*s ≥ .06), and between Profile similarity and Physical similarity (all *r*s ≥ .51, all *p*s ≤ .05, except for two, both *r*s ≤ .48, both *p*s ≥ .07).

Finally, I conducted a 2 (Culture) by 2 (Similarity type) mixed model ANOVA on the correlation coefficients to test whether Profile and Perceptual similarity differently associate with Physical similarity across cultures. The analysis revealed a significant effect of Similarity type (i.e., Profile similarity vs. Perceptual similarity), $F(1, 38) = 8.40$, $p = .006$, $\eta p^2 = 0.181$, a significant effect of Culture, $F(1,38) = 4.55$, $p = .04$, $\eta p^2 = .107$, and no interaction between Culture and Similarity type, $F(1,38) = .24$, $p = .63$, $\eta p^2 = .006$. These results indicate that Profile similarity is more strongly associated with Physical similarity compared to Perceptual similarity and that Chinese participants' behavioural performance is more strongly associated with stimuli property (i.e., Physical similarity) compared to British participants.

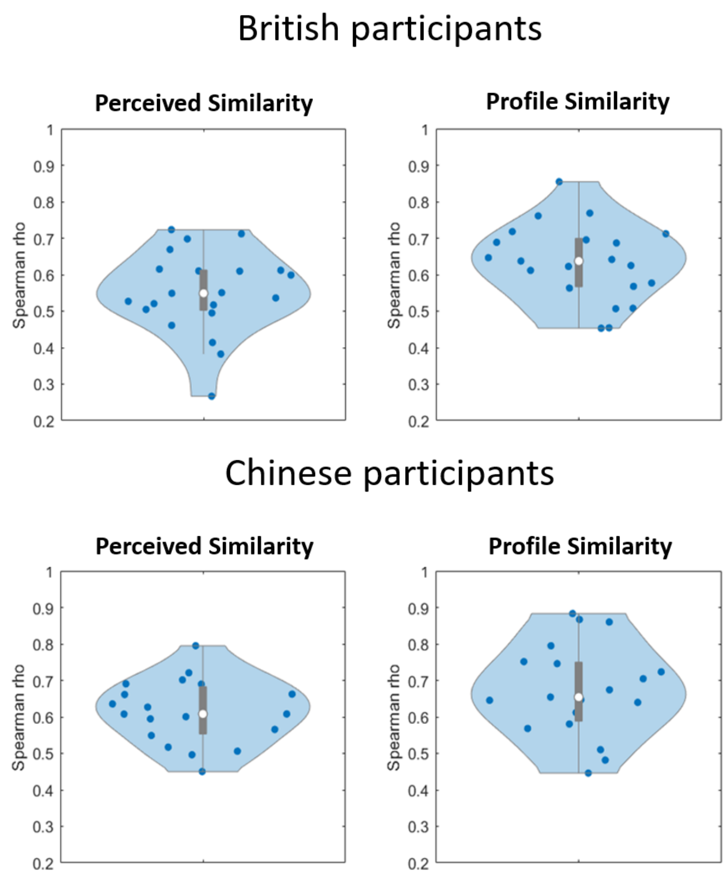

Figure 25. **Correlation between Physical similarity and Perceptual and Profile similarity for individual participants.**

Profile similarity can better explain Perceptual similarity than Categorical similarity, especially for dynamic facial emotions

To investigate how Profile similarity contributes to Perceptual similarity between facial emotions compared to other factors, I performed a multiple linear regression analysis including Categorical similarity, Physical similarity and Intensity similarity obtained from previous studies.

For static facial emotions, Physical similarity alone accounted for the higher variability in perceptual similarity, R^2 adjusted = .825, $F(1, 158) = 750$, $p < .001$, followed by Profiling Similarity, R^2 adjusted = .683, $F(1, 158) = 344$, $p < .001$, Categorical Similarity, R^2 adjusted = .542, $F(1, 158) = 189$, $p < .001$, and finally Intensity Similarity, R^2 adjusted = .442, $F(1, 158) = 127$, $p < .001$.

A model combining Physical, Categorical and Intensity information outperformed individual predictors, R^2 adjusted = .872, $F(3, 156) = 363$, $p < .001$. All three factors significantly contribute to the model, all ts ≥ 4.20, all ps<.001. A model comparison analysis showed that the introduction of Categorical similarity first and then Intensity information kept improving model performance. Adding Categorical information to physical similarity, $\Delta R^2 = .034$, $F(1, 157) = 38.8$, $p < .001$, further addition of Intensity similarity, $\Delta R^2 = .014$, $F(1, 156) = 17.6$, $p < .001$.

A model combining Physical, Profiling and Intensity information produced the best fit to the data, R^2 adjusted = .901, $F(3, 156) = 485$, $p < .001$. All three factors significantly contribute to the model, all ts ≥ 3.82, all ps $< .001$. A model comparison analysis showed that the introduction of Profiling and then Intensity information improves model performance. Introducing Profiling information in addition to Physical similarity, $\Delta R^2 = .068$, $F(1, 157) = 100.9$, $p < .001$, further introducing Intensity information to the model, $\Delta R^2 = .000$, $F(1, 156) = 14.6$, $p < .001$.

Models integrating Profiling information showed better performance compared to models integrating Categorical information. While adding Profiling similarity to Categorical similarity significantly improve the model performance (Categorical alone, R^2 adjusted = .542, $F(1, 158) = 189$, $p < .001$, introducing Profiling similarity, R^2 adjusted = .697, $F(2, 157)$

= 184, p < .001, Model comparison, ΔR^2 = .155, F(1, 157) = 81.5, p < .001), adding Categorical similarity to Profiling similarity significantly but only slightly improve the model performance (Profiling alone R^2 adjusted = .683, F(1, 158) = 344, p < .001, introducing Categorical similarity, R^2 adjusted = .697, F(2, 157) = 184, p < .001, Model comparison, ΔR^2 = .015 F(1, 157) = 7.86, p = .006). In particular, when Physical similarity is included in the model, Categorical information no more contributes to the model, t = .998, p = .320, becoming redundant (model comparison, ΔR^2 = 0, F(1, 156) = .99, p = .320).

Similar results were found for dynamic facial emotions. Physical similarity alone accounted for the higher variability in perceptual similarity, R^2 adjusted = .776, F(1, 158) = 552, p < .001, followed by Profiling Similarity, R^2 adjusted = .754, F(1, 158) = 488, p < .001, Categorical Similarity, R^2 adjusted = .660, F(1, 158) = 310, p < .001, and finally Intensity Similarity, R^2 adjusted = .503, F(1, 158) = 162, p < .001.

However, differently from static facial emotions, adding intensity similarity did not further improve model performance for (1) a model combining Physical and Categorical information (existing model performance, R^2 adjusted= .868, F(2, 157) = 524, p < .001, model comparison, ΔR^2 = 0, F(1, 156) = .714, p =.40, with Intensity factor not significantly contributing to the model, t = .84, p = .40), and (2) a model combining Physical and Profiling information (existing model performance, Adjusted R^2 = .916, F(3, 156) = 576, p < .001, model comparison, ΔR^2 = 0, F(1, 156) = 1.10, p = .29, with Intensity factor not significantly contributing to the model, t = 1.05, p = .296).

Again, models integrating Profiling information showed better performance compared to models integrating categorical information. Adding Profiling similarity to Categorical similarity significantly improve model performance (Categorical alone, Adjusted R^2 = .660, F(1, 158) = 310, p < .001, introducing Profiling similarity, Adjusted R^2 = .773, F(2, 157) = 271, p < .001, Model comparison, ΔR^2 = .113, F(1, 157) = 79.2, p < .001), whereas integrating Categorical similarity to Profiling similarity significantly but only slightly improve model performance (Profiling alone, Adjusted R^2 = .754, F(1, 158) = 488, p < .001, introducing Categorical similarity, Adjusted R^2 = .773, F(2, 157) = 271, p < .001, Model comparison, $\Delta R^2 = .020$, $F(1, 157) = 14.3$, $p <.001$). When Physical similarity is integrated, Categorical information only slightly contribute to the model, $t = 2.04$, $p = .043$, being mostly redundant, model comparison $\Delta R^2 = .002$, $F(1, 156) = 4.17$, $p = .043$.

Together, these results indicate that Physical similarity is the best individual predictor for participants' judgments of stimuli similarity. However, the model performance significantly improves with the integration of perceiver-based information, particularly Profiling or Categorical similarity. Regarding Intensity similarity, while it slightly enhances the model performance for static stimuli, it does not improve the performance for dynamic stimuli, suggesting that the availability of dynamic cues make explicit intensity information redundant. Importantly, Profiling similarity performs better than Categorical similarity, which mostly became redundant once Profiling information is combined with physical information. Overall, when dynamic cues are available behavioural models (i.e., Profiling, Categorical and Intensity similarity) can better explain perceptual similarity compared to their performance in response to static stimuli.

3.3 Study 4. Semantic profiling: do we perceive more than happy from facial expressions of happy?

Study 3 showed that our perception of facial expressions of emotions goes beyond a singular target emotion category. An emotion profiling approach proves more effective than a categorical one in explaining how we perceive the differences and similarities between facial expressions. In a similar way, in our daily life we represent emotional experience within a semantic space that include a great variety of dimensions, and when asked to report our mood we typically do not rely on one single discrete emotion category (Sabini & Silver, 2005; Shaver et al., 1987). The interpretation and processing of emotional content conveyed by facial expressions often carry additional connotations and meanings, resulting in a blend of coexisting concepts and semantics closely tied to the target emotion term (Jackson et al., 2019; Liu et al., 2022).

In Study 4 I explored the rich semantic information potentially conveyed by spontaneously elicited facial emotions. Specifically, I investigated whether participants employ different related semantic terms to define their perception of facial emotions, beyond the primary target one, thereby generating unique Semantic profiles for each facial emotion. In other words, I aimed to understand why we label a facial expression of happy as “happy” rather than “joy” or “excitement”. If “happy” is the only semantic concept we can attribute to the facial expression of “happy”, we would consistently use that term and not another. Additionally, I examined the role of facial motion, emotion intensity, and culture in the perception of semantic information from facial expressions. To this aim, I adopted the same natural facial expressions of emotions as in Studies 1 to 3, and introduced a Semantic profiling task to measure how each facial emotion is comprehended along different semantic dimensions. The selection of semantics related to "Happy," "Fear," and "Pain" was based on the work of Jackson et al (2019). Here, semantic information associated with emotion concepts across different languages has been investigated through colexification. Colexification occurs when multiple concepts are expressed by the same word across different languages, indicating their perceived similarity. Jackson et al. (2019) compiled colexifications into a database featuring 2474 languages and 2439 distinct concepts, including 24 emotion concepts (Rzymski et al., 2020). They estimated and compared networks of emotion concept colexification across different languages, identifiying a universal colexification network, shared by all language families, and networks specific to the 20 language families. In these networks, concepts are represented as distinct nodes, and the colexifications between concepts are represented as weighted edges, indicating the number of languages sharing the colexification. These findings suggest that the semantic meaning of an emotion concepts, such as “fear”, can vary across cultures. For example, while the concept of “fear” strongly links to the semantic concept of “anxiety” in the Indo-European language family, it is associated with the semantic concept of “surprised” in the Austroasiatic family. This suggests that individuals from these two language families may undergo slightly different emotion experience when processing facial expression of fear. Based Jackson et al.’s (2019) findings, I extracted 5 additional concepts semantically related to each of the three emotions from the universal network and the language family network most relevant to our participants, namely, the Indo-European network and the Sino-Tibetan network. I then investigated how the observed Semantic profiles differed between the two distinct cultures, different stimuli formats (static vs dynamic), and two varying levels of emotion intensity.

3.3.1 Methods

Participants

The size of the participants’ sample was based on the results from our previous studies, suggesting that a minimum of 13 participants was required. Again, considering the fact that the study was conducted online we collected 20 participants for condition (dynamic/static).

Forty-five British participants were recruited from the University of East Anglia via the SONA System (7 males, 38 females; age ranged between 18-31 yrs., M = 20.2, SD = 2.92). Participants who did not identify themselves as British in the demographic questionnaire were excluded from the study. Forty Chinese participants were recruited from the Sun Yat-sen University, China (9 males; 31 females; age ranged between 18-28 yrs., M = 20.3, SD = 2.24). None of these participants had taken part in any previous studies of this project. All participants provided informed consent before taking part in the study and were debriefed at the end, receiving course credits as compensation.

Stimuli, Materials, and Tasks

The stimuli were the same as in previous studies of this project, including facial images or videos of 9 actors displaying 3 different emotions (Happy, Pain, Fear) at 2 intensities (High, Low) taken from the large MPI Facial Expression Database (Kaulard et al., 2012). The study consisted of a Semantic profiling task and a Similarity rating task. The similarity rating task was identical to that used in Study 3 (see Figure 26b).

For the Semantic profiling task, each trial started with a fixation cross (1000ms), followed by the stimulus displayed on the left half of the screen (i.e., image for the Static task, video for the Dynamic task) and 6 different sliders on the right half of the screen. Participants were asked to rate each facial expression along six semantic dimensions by moving the handle of the sliders (see Figure 26a). Each slider was independent from the others and ranged from 0 to 100. The handles of all sliders were initially placed on 0. To ensure participants had sufficient time to provide their answers without overthinking, preventing the interference of more high-level processes, the response screen had a time-limit of 20 seconds. Dynamic stimuli were presented on a loop until a decision was made or the time limit was reached. If the time limit was reached, the experiment moved on to the next trial. Similar to the emotion profiling task in Study 3, there were 108 trials in total (= 3 emotions * 2 intensity levels * 9 actors * 2 repetitions).

The six semantic dimensions differ depending on the facial emotion displayed by the stimulus (Happy, Fear, Pain). The selection of semantic dimensions was based on the work of Jackson et al. (2019). In their study, researchers investigated emotion semantics across 2474 spoken languages by analysing colexification, a phenomenon where similar concepts are expressed using the same word within a language. They aggregated colexifications into a database of cross-linguistic colexification featuring 2474 languages and 2439 distinct concepts, including 24 emotion concepts (Rzymski et al., 2020), and generated a universal colexification network, as well as networks for 20 specific language families. From the universal network of emotion and the language family networks most relevant to our participants, namely the Indo-European network and the Sino-Tibetan network, we extracted a list of 5 concepts that are semantically related to each of the three target emotions used in the present study: Happy, Fear, and Pain. Specifically, when a Happy expression was displayed, the dimensions shown were: "Happy, Joy, Want, Desire, Like, Love." For a Pain expression, the dimensions shown were: "Pain, Sick, Grief, Ache, Hurt, Anxiety." Finally, when a Fear expression was displayed, the dimensions shown were: "Fear, Anxiety, Envy, Grief, Surprise, Regret."

The same translation process employed in the previous study was followed, aiming to ensure accurate translation and maintain consistency across languages. Also in this case, it is crucial to recognize the potential for semantic differences between languages, which may result in subtle variations in interpretation.

Finally, the same measures taken in previous studies were implemented with the aim to reduce, and account for, potential participants' distractions or disengagement. In particular, (1) access to the study was restricted to PCs or laptops; (3) participants were required to self-report the reliability/usefulness of their data (e.g., due to lacked attention) at the end of the experiment, (4) a time-limit was imposed on each screen, and participants exceeding the maximum allotted time for completing the experiment were excluded from the study; (5) control trials were included in the study, where participants were asked to rate the similarity of identical images.

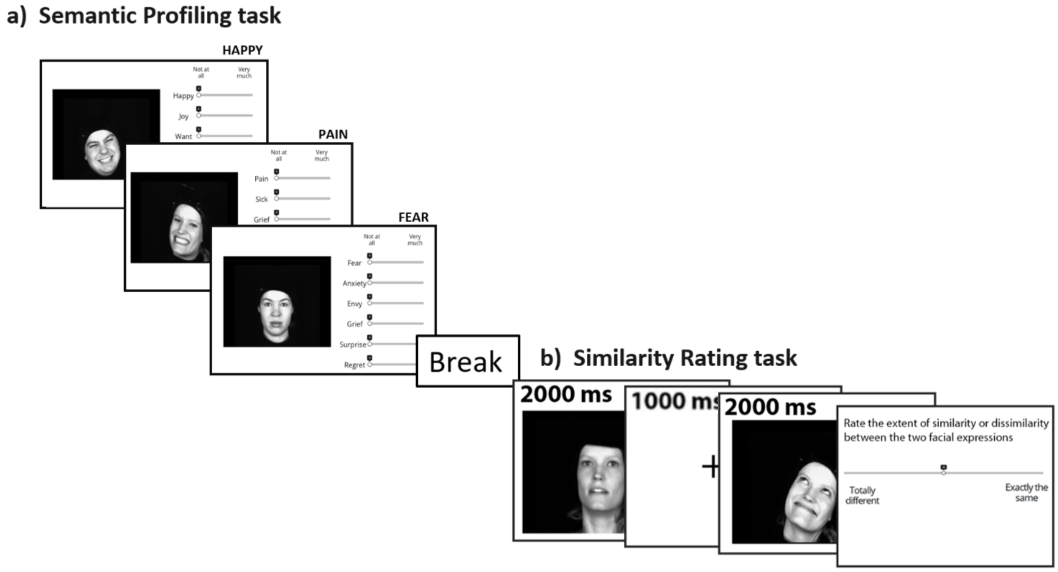

Figure 26. **Semantic profiling and similarity rating tasks**. *a) In the Semantic profiling task participants rated each facial expression along 6 semantic dimensions. The dimensions were specific to the emotion displayed by the stimulus (e.g., for Happy: Joy, Desire, Love, etc., for Pain: Grief, Anxiety, Hurt, etc.) b) In the similarity rating task, participants saw two facial emotions in a row. Then, using a slider ranging from 0 to 100, they rated the*

Procedure

Participants performed both Semantic profiling and similarity rating task online through the Gorilla platform using their desktop computer or laptop (i.e., no tablet or phone access was allowed). The general procedure was similar to Study 3 except that the Emotion profiling task was replaced by the Semantic profiling task. Half of the participants were randomly allocated to the Dynamic task, with movies as stimuli, while the other half to the Static task, with images as stimuli. Participants completed the Semantic profiling task first, then, after a 5-minute break, they went through the similarity rating task (see Figure 26). Three practice trials were given before starting each task. They had a second break of 5-minute halfway through the Similarity rating task and the whole study took about 1 hour to complete. At the end of the study, they filled out a self-report questionnaire regarding the reliability of their data and were then debriefed.

3.3.2 Results and Discussion

Applying the same exclusion criteria as in previous studies, one British participant was excluded from the following data analysis, resulting in a final sample size of 44 British participants and 40 Chinese participants.

Facial expressions of emotions are linked to more semantic concepts than the target emotion concept

To investigate whether participants attribute different related semantic terms to the observed facial expressions, I tested whether their responses to the Semantic profiling task exhibited a dominant score to the target emotion concept accompanied by near-zero scores for the other dimensions, or a more diverse profile with multiple prominent dimensions. Figure 27 shows the overall responses averaged across culture, stimulus types (static-dynamic), and intensity levels (low/high), which reveals that participants employed multiple semantics to interpret facial expressions of Happy, Fear, and Pain (for details of semantic profiles before averaging, see Appendix B, Figure B.3).

In particular, facial expressions of Pain are characterized by higher scores in the target dimension of pain (43%, SD = 16.6, paired t-test comparing to other dimensions, all $ts(88) \geq 3.48$, all $ps \leq .001$), and co-occurring high scores in other semantic dimensions, such as Hurt (37%, SD = 17.8), Ache (36%, SD = 15.4), Grief (26%, SD = 16.3), Anxiety (21%, SD = 15.6), and Sick (19%, SD = 16.5). For facial expressions of Fear, I found higher scores for Surprise (53%, SD = 16.5) than the target dimension of Fear (31%, SD = 13.8, paired t-test comparing target to other dimensions, all $ts(88) \geq 4.22$, all $ps < .001$), and co-occurring high scores in Anxiety (26%, SD = 17.6), Regret (23%, SD = 17.3), Envy (16%, SD = 13.4), and Grief (12%, SD = 12.6). Finally, for Happy, higher scores to the target dimension of Happy (45%, SD = 17.3, paired t-test comparing to other dimensions, all $ts(88) \geq 4.02$, all $ps < .01$) are accompanied by high scores in Joy (40%, SD = 17.8), Like (34%, SD = 17.5), Desire (27%, SD = 17.4), Want (27%, SD = 17.9), and Love (25%, SD = 17.8).

Figure 27. **Semantic Profiles of Happy, Fear and Pain.** *Colored lines represent averaged profiling responses across participants and conditions (i.e., facial motion, intensity, culture). Gray lines represent semantic profiles for individual participants.*

Semantic Profiles are sensitive to Facial Motion, Emotion Intensity and Culture

To test whether semantic profiles were affected by facial motion (i.e., Dynamic vs Static stimuli), emotion intensity (i.e., High vs Low) and participants' cultural background (i.e., Chinese vs British participants), I obtained and compared contrast profiles between corresponding conditions (for details of semantic profiles before averaging, see Appendix B, Figure B.3). Resulting semantic profiles for each facial expression were analysed using a 2 (British vs Chinese / Dynamic vs Static stimuli / High vs Low emotion intensity) by 6 (semantic dimensions) ANOVA.

***British vs Chinese participants' responses*.** For facial expressions of Happy (Figure 28a, top panel), there was a main effect of response dimensions, $F(5,435) = 59.22$, $p < .001$, $\eta p^2 = .405$. This indicates that participants provided significantly different scores for the six semantic dimensions. There was no significant difference in overall responses (across semantic dimensions) between British vs Chinese participants, $F(1,87) = 1.41$, $p = .239$, $\eta p^2 = .016$. Importantly, the interaction between Culture and response dimensions was significant,

$F(5,435) = 5.62$, $p < .001$, $\eta p^2 = .061$. Post-doc independent t-tests for each semantic dimension showed significantly higher responses to the dimensions of Happy ($t(87) = 2.17$, $p = .032$) and Love ($t(87) = 2.39$, $p = .019$) in British than in Chinese participants, while responses to other dimensions did not differ significantly (all ts $\leq$ 1.80, all ps $\geq$ .074).

For facial expressions of Fear (see Figure 28a, middle panel) and Pain (see Figure 28a, bottom panel), similar results were obtained. The same ANOVA analysis showed a significant effect of response dimensions, for Fear, $F(5,435) = 167.43$, $p < .001$, $\eta p^2 = .658$, and for Pain, $F(5,435) = 74.55$, $p < .001$, $\eta p^2 = .461$. The main effect of participant group was not significant, for Fear, $F(1,87) = 2.09$, $p = .152$, $\eta p^2 = .023$; for Pain, $F(1,87) = .105$, $p = .747$, $\eta p^2 = .001$. The interaction between culture and response dimensions was significant for Fear, $F(5,435) = 7.89$, $p < .001$, $\eta p^2 = .083$, and for Pain, $F(5,435) = 5.87$, $p < .001$, $\eta p^2 = .063$. For Fear, follow-up independent t-tests showed significantly higher scores for Surprise in Chinese participants ($t(87)= 2.34$, $p = .021$) and for Anxiety in British participants ($t(87) = 3.05$, $p = .003$). They did not differ along other dimensions (all ts $\leq$ 1.86, all ps $\geq$.067). For Pain, independent t-tests showed no significant difference between British and Chinese participants (all ts $\leq$ 1.84, all ps $\geq$.068).

<u>High vs Low intensity facial emotions</u>. The main effect of response dimensions was significant for all three facial emotions (Figure 28b) [for Happy, $F(5, 440) = 56.2$, $p < .001$, $\eta p^2 = .390$; for Fear, $F(5, 440) = 152.1$, $p < .001$, $\eta p^2 = .633$; for Pain, $F(5, 440) = 70.92$, $p < .001$, $\eta p^2 = .446$]. The main effect of emotion intensity was significant for Happy, $F(1,88) = 89.3$, $p < .001$, $\eta p^2 = .504$, for Fear, $F(1,88) = 20.3$, $p < .001$, $\eta p^2 = .188$, but not for Pain, $F(1,88) = .247$, $p = .621$, $\eta p^2 = .003$. Finally, there was a significant interaction between Intensity and response dimensions for Happy, $F(1,440) = 30$, $p < .001$, $\eta p^2 = .254$, for Fear, $F(1,440) = 53.9$, $p < .001$, $\eta p^2 = .380$, but again not for Pain, $F(1,440) = 1.75$, $p = .121$, $\eta p^2 = .020$. Follow-up paired t-tests showed that, for Fear, high intensity stimuli induced significantly higher responses to dimensions of Fear, Anxiety, Regret and Grief and significantly lower responses to Surprise and Envy (all ts(87) $\geq$ 2.48, all ps $\leq$ .015). For Happy, high intensity stimuli elicited significant higher responses to dimensions of Happy, Love, Joy, and

Like (all ts(87) $\geq$ 4.79, all ps < .001) but not for Desire and Want (both ts(87) $\leq$ 2.48, all ps $\geq$ .065).

Dynamic VS Static facial expressions. There was a significant main effect of response dimensions for all three emotions (Figure 28c) [for Happy, $F(5, 435) = 55.828$, $p < .001$, $\eta p^2 = .391$, for Fear, $F(5, 435) = 170.6$, $p < .001$, $\eta p^2 = .662$, and for Pain, $F(5, 435) = 72.39$, $p < .001$, $\eta p^2 = .454$]. There was also a significant difference between overall responses to static and dynamic stimuli for all three emotions [for Happy, $F(1,87) = 6.18$, $p = .015$, $\eta p^2 = .066$; for Fear, $F(1,87) = 10.2$, $p = .002$, $\eta p^2 = .105$; for Pain, $F(1,87) = 7.87$, $p = .006$, $\eta p^2 = .083$], static facial emotions consistently induced stronger responses than dynamic facial emotions. The interaction between facial motion and response dimensions was significant for Fear, $F(5, 435) = 11.3$, $p < .001$, $\eta p^2 = .115$, for Pain, $F(5, 435) = 2.77$, $p = .018$, $\eta p^2 = .031$, but not for Happy, $F(5, 435) = 0.645$, $p = .665$, $\eta p^2 = .007$. Follow-up independent t-tests showed that, for Fear, responses to static facial expressions were significantly higher for all dimensions except Surprise (all ts(87) $\geq$ 2.62, all ps $\leq$ .010, except for Surprise, $t(87) = 1.94$, $p = .056$). Similarly, for Pain, scores were significantly higher for static stimuli compared to dynamic stimuli for the dimensions of Grief ($t(87) = 4.70$, $p < .001$), Pain ($t(87) = 2.24$, $p = .027$), and Anxiety ($t(87) = 2.29$, $p = .024$), but not for the other response dimensions (all ts(87) $\leq$ 1.69, all ps $\geq$ .095). Finally, for Happy, scores were significantly higher for static stimuli for the dimensions of Happy ($t(87) = 3.15$, $p = .002$), Joy ($t(87) = 2.15$, $p = .034$), and Like ($t(87) = 1.93$, $p = .05$) but not for the other response dimensions (all ts(87) $\leq$ 1.89, all ps $\geq$ .062).

To sum up, these results suggest that while semantic profiles are overall similar between British and Chinese participants, static and dynamic faces, and high and low intensity emotions, they are also sensitive to these factors showing fine-grained but significant differences. To have a deep look into these differences, see Appendix B, Figure B.3 for results not averaged across conditions.

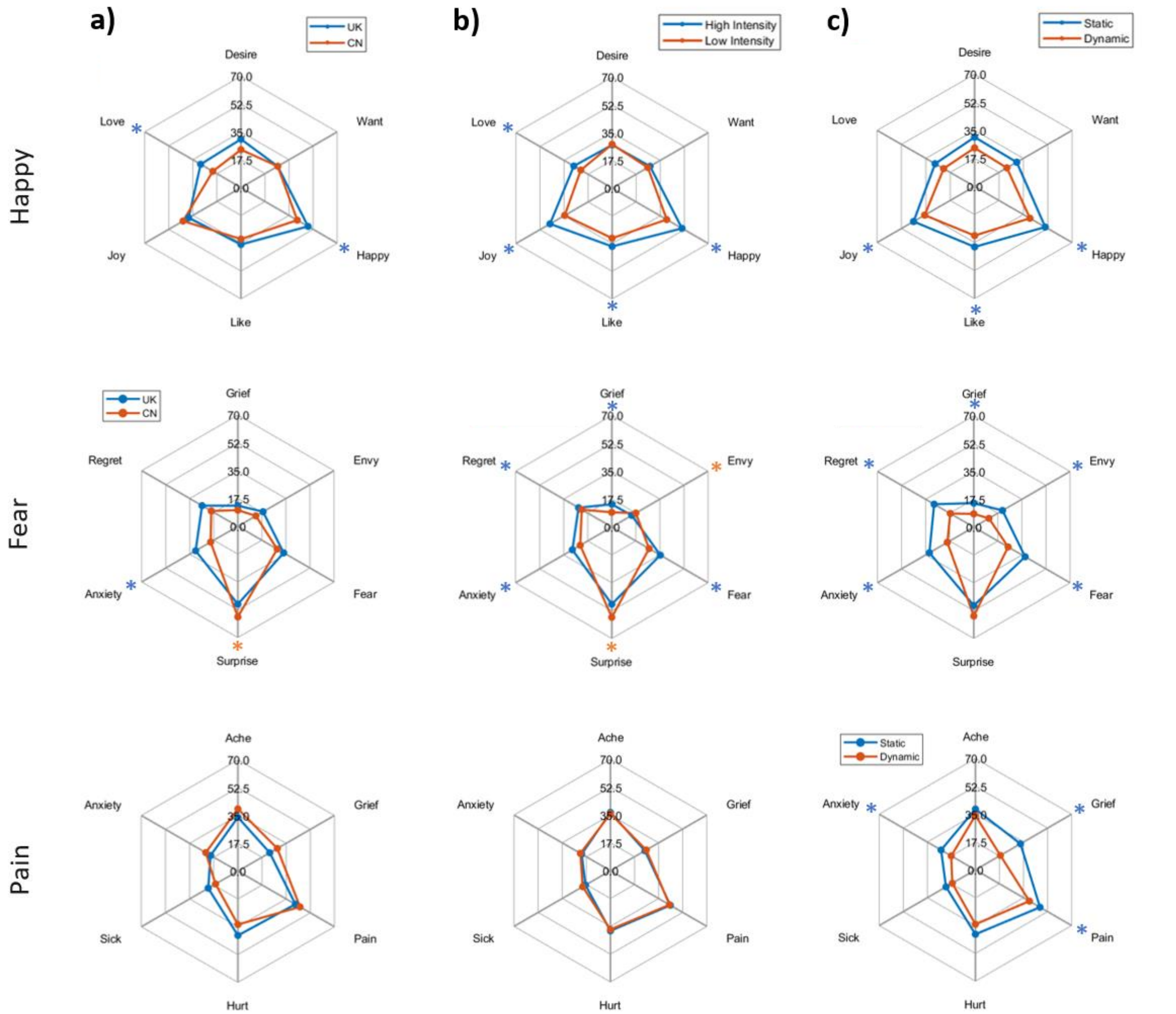

Figure 28. **Semantic profiles of Happy, Fear and Pain as a function of culture, facial motion and emotion intensity.** *a) Contrast between responses of British participants, in blue, and Chinese participants, in orange, to facial expressions of Happy, Fear and Pain. b) Contrast between responses to High and Low intensity emotion of Happy, Fear, and Pain. c) Contrast between responses to Dynamic and Static facial expressions of Happy, Fear, and Pain. Asterisks indicate significant differences.*

3.4 General discussion

In this Chapter, I presented two cross-cultural studies that demonstrate how facial expressions of emotions are perceived as high-dimensional emotional and semantic profiles. These profiles offer a better explanation for the representation of spontaneous facial emotions and their perceptual similarity. Moreover, such a more rich and sensitive representation of facial emotions also revealed the fine-grained impact of facial motion, emotion intensity and culture on emotion processing.

In Study 3, we employed an emotion profiling task to investigate whether participants perceived the emotional content of facial expressions as a single discrete category or as an emotion profile composed of a mixture of diversified emotion contents. Results showed that participants employed multiple emotion dimensions to define facial expressions of Happy, Fear, and Pain. For example, in facial expressions of Pain, higher scores to the target dimension of Pain were followed by high scores to the co-occurring dimensions of Disgust, Fear, and Surprise. Interestingly, for facial expressions of Fear, participants assigned higher scores to the non-target dimension of Surprise, followed by the target dimension (i.e., Fear). It is worth noting that stimuli used in our studies were validated by Kaulard et al. (2012) through a free-naming task, in which facial expressions depicting Fear received 71% of valid responses. While participants mostly recognized these expressions as conveying Fear when forced to assign a single label, adopting a profiling approach unveiled the perception of a significant component of Surprise. This result seems to suggest that, even for well-validated facial emotions, the commonly used label may not fully capture their closest meaning. Additional support to the emotion profiles comes from a Random Forest classification model and feature importance analysis of participants' responses. This analysis revealed that multiple emotion dimensions play a crucial role in enabling the model to distinguish between different facial emotions based on participants' profiling responses.

Importantly, results also indicate that participants adopted a more categorical approach in defining facial expressions of Happy compared to Fear and Pain. One potential explanation for this result may lie in the well-established "Happy advantage" which states that due to the perceptual and categorical distinctiveness of Happy facial expressions, Happy faces are recognized more quickly and accurately than other facial expressions (Calvo et al., 2012; Calvo & Beltrán, 2013; Leppänen & Hietanen, 2007). Hager and Ekman (1979) were the first to provide empirical evidence that, when viewed at a distance, facial expressions of happiness are more distinguishable than other expressions. Since then, the “Happy advantage” has been consistently observed in studies comparing the recognition of the six basic emotion categories (Calvo & Lundqvist, 2008; Palermo & Coltheart, 2004; Tottenham et al., 2009), adopting different stimulus sets such as the Karolinska Directed Emotional Faces (KDEF; e.g., Calvo & Lundqvist, 2008) or the Pictures of Facial Affect (Ekman & Friesen, 1976a; Leppänen & Hietanen, 2004), and with happy faces depicting either open mouths and exposed teeth or with closed-mouth smiles (Tottenham et al., 2009). However, my results reveal that, even if more categorically interpreted compared to Fear and Pain, facial emotions of Happy are still characterized by co-occurring emotional dimensions other than the primary one (e.g., Surprise, Disgust). Once again, facial expressions are better described as unique and rich profiles rather than a single discrete emotion category.

Study 4 further demonstrate that the perception of facial emotions carries a broader range of semantic concepts beyond the target emotion category, resulting in a coexisting Semantic profile. For example, when participants profiled facial expressions of Pain, alongside the target dimension (i.e., Pain) they consistently assigned high scores to other co-occurring semantic dimensions such as Hurt, Ache, Grief, Anxiety, and Sick. These semantics are strongly connected to the concept of Pain, as identified by Jackson (2019), and can all be simultaneously detected from facial expressions conveying Pain. These results further support that the perception of facial expression of emotions is rich, blended, diversified and high-dimensional, and it is better represented as fine-grained emotion/semantic profile than a discrete emotion category.

Through both studies we explored the role played by participants’ culture in how facial emotions were profiled and perceived. First, it was found higher variability in semantic profiles compared to emotion profiles within the same culture. This suggests that while the emotion profile of a facial expression is generally shared across participants, the way other emotion-relevant semantic concepts are extracted from faces is more susceptible to individual variability. This finding aligns with the view that our emotional experiences are shaped by culturally learned concepts and social rules (Barrett, 2017; Barrett et al., 2007; Matsumoto, Keltner, et al., 2008). For instance, mental states defined as “Fear” do not all look alike or feel alike. Variability in the frequency but also quality of the emotional experience has been observed within the same individuals over time, across individuals from the same culture, and across cultures (Barrett, 2009). Even though basic emotions may seem more universally defined, especially in traditional theories, the content beyond the same basic emotions can vary across cultures and individuals. For example, while in Russian, the concept of sadness is perceived as closer to physical agony, in USA, it is considered closer to the experience of loss (Wierzbicka, 2009). Similarly, while in USA, anger involves psychological distance from others, in Japan, it relates to increased proximity and closeness (Kitayama et al., 2006). Even within the same culture, different people may not necessarily experience or process emotions in the same way. we all define emotion using the same words, some people feel the heat of anger, the despair of sadness, the dread of fear, whereas others experience instead pleasant or unpleasant feelings with little specificity (Barrett, 1998, 2004; Feldman, 1995). Consistent with the literature, the present results show that, within participants from the same culture, while the structure and shape of the emotion profiles tend to be relatively stable, the semantic profiles beyond the target emotion concept exhibit significant variation. This suggests that the emotional experience is more diversified in terms of perceived fine-grained semantics than emotional labels.

Similarly, when comparing Chinese and British participants’ responses, the current results demonstrate that, although the overall emotion and semantic profiles are similar across participants’ cultural background, they still significantly differ from each other. Specifically, for emotion profiles, Chinese participants tend have a more diverse representation of the same facial emotion compared to British participants, attributing lower scores to the target dimension (e.g., happiness in happy expression) while detecting higher levels of co-occurring emotions dimensions. Consistent with this result, previous studies show that East Asians tend to experience multiple different emotions concurrently, while North Americans and Europeans are more likely to experience specific feelings. At the same way, Easterners are more inclined to perceive blends of emotions than Westerners (Grossmann et al., 2016; Miyamoto et al., 2010). In a study of Fang et al. (2018) Chinese observers endorsed non-target emotions more than Dutch observers and, interestingly, this difference was more pronounced for morphologically similar emotions than of dissimilar emotions. Jack et al. (2012) reconstructed a 3D dynamic model for the six basic emotions in Western Caucasian and East Asian cultures by using a FACS-based random facial expression generator and reverse correlation, and they found that while Western Caucasians represent each of the six basic emotions with a distinct set of facial muscles, East Asians show a considerable overlap between emotion categories.

However, while our finding that Chinese participants tend have a more diverse representation facial emotions compared to British, seem to be supported by previous literature, it is also important to consider that stimuli adopted in this project depicted Western no-professional actors, which may have led to an in-group advantage in processing facial expressions. People tend to be more accurate in judging emotional expressions of individuals within their own cultural group. Theories explained this phenomenon suggesting that in-group faces convey culturally specific elements of emotional expressions, that emotion is a universal language characterized by subtly different dialects, and that individuals may be less motivated to recognize the emotions of other individuals of foreign cultures (Elfenbein, 2015; Elfenbein & Ambady, 2002).Therefore, further investigations are needed to better understand the nature of the cultural differences detected in our studies.

Emotion and Semantic profiles have also been found to be sensitive to emotion intensity and facial motion. Regarding the intensity of the emotion displayed by the stimulus, participants' responses to the target emotion dimensions were higher for stimuli depicting high-intensity emotions compared to low-intensity emotions. These results align with the rating scores obtained from the Intensity rating task in Study 2, where facial expressions of high-intensity emotions were rated as more intense than facial expressions of low-intensity emotions. It is important to highlight that, unlike previous studies where different levels of intensity are usually obtained by directly controlling the physical muscle involved in that specific emotion, here these are elicited by specific emotional scenarios validated to evoke a high- or low- intensity emotion. Moreover, while the target emotion dimensions were always perceived as being higher when the emotion was displayed at high intensity, other emotion dimensions within the emotion profile modulated based on the emotion intensity. Some of these dimensions were perceived more strongly when stimuli depicted high-intensity emotions, while others were more pronounced when stimuli depicted low-intensity emotions. Similar results were observed in the semantic profiles.

When comparing emotion profiles in response to static and dynamic stimuli, it was observed that while the overall emotion profiles were quite similar, certain emotion dimensions were perceived more strongly with dynamic stimuli, while others were perceived more strongly with static stimuli. This difference was particularly noticeable in response to facial expressions of Fear and Pain. Interestingly, for the semantic profiles, most of the semantic dimensions were perceived more strongly with static facial emotions than with dynamic ones. To sum up, even though the general overall structure of emotion and semantic profiles is similar, they do show sensitivity to relevant factors such as perceivers' culture, emotional intensity, and whether facial expressions contain dynamic cues.

Results to both studies also provide compelling evidence that rich emotion profiles can be used to predict participants' perceptual similarity across cultural backgrounds and stimulus types. Supported on a group-, participant-, and trial level, these finding indicate that participants may in fact rely on the information contained in the emotion profiles to make judgments regarding the similarities between facial expression of emotions. Furthermore, analysis conducted at the participant level revealed a stronger association between perceived emotion profiles and participants' similarity judgments for dynamic stimuli compared to static stimuli, and for British participants compared to Chinese participants. The former suggests that when motion cues are available, participants tend to base their similarity judgments more on stimuli's emotional content (i.e., emotion profiles), compared to when facial emotions are static. Facial motion seems to convey relevant information for the extraction of emotional content that guide the perception of facial expressions. The latter, while may be related to the out-group disadvantage for face processing resulting in more confused responses to the profiling task, it may also suggest that, while Chinese and British participants are more similar in rating stimuli similarity, when they are required to define the emotion profile of a single face, Chinese participants tend to have a more nuanced and complex approach which is less adopted once asked to rate stimuli similarity. In support of this, it seems that differences in participants perceptual similarity judgments (i.e., perceived similarity matrices) depend more on the stimulus type (dynamic vs static) compared to culture. That is, pattern of responses for perceptual similarity of British and Chinese participant are highly similar. However, profile similarity matrices show an opposite tendency. For profile matrices it seems that pattern of responses within stimulus type and between cultures (e.g., British Dynamic and Chinese Dynamic) are more different than pattern of responses between stimulus type and within culture (e.g., British Dynamic and British static). This result suggests that, for profile similarity, the effect of culture is stronger compared to that of stimulus type. To sum up, it seems that while perceptual similarity is more influenced by the stimulus type (dynamic or static) than Culture, profile similarity is more influenced by Culture and less by the stimulus type.

I also investigated the relationship between profile and perceived similarity with stimuli physical similarity. Results showed that both perceptual and profile similarity correlated with physical similarity, for both Cultures, and at a group- and participant-level. Moreover, physical similarity has a slightly stronger correlation with profile similarity in comparison to perceived similarity, and Chinese participants show overall higher correlation between Physical and Perceptual similarity compared to British participants. These results suggest that Chinese participants integrate more physical information in their similarity ratings compared to British participants, which is consistent with previous results showing that emotion profile information is more integrated in perceptual similarity by British compared to Chinese participants. Profiling similarity is strongly correlated with physical similarity compared to perceived similarity. Yrizarry and colleagues (1998) have proposed that people preferentially endorse emotions that share the same facial components (i.e., morphological similarities). It has also been shown that the extent to which non-target emotions are perceived from a facial expression depends on their morphological similarity (Fang et al., 2018). This may explain why a fine-grained representation of the emotional content (i.e., emotion profiles) strongly correlate with stimuli-based physical properties.

Incorporating results from Study 1 and 2, I determined the sources of information that could better explained participants' perceived similarity. In particular, results showed that Physical similarity is the best individual predictor in explaining participants' judgments of stimuli similarity. However, the model performance significantly improves with the integration of further perceiver-based information, particularly Emotion profiling or Categorical similarity. Regarding Intensity similarity, while it slightly enhances the model performance for static stimuli, it does not improve the performance for dynamic stimuli, suggesting that the availability of dynamic cues make explicit intensity information became redundant. Importantly, when compared to Categorical models, Emotion Profiles better explain the way we perceive similarity between facial emotions, especially when dynamic cues are available. Categorical similarity became largely redundant once Profiling information is combined with Physical information. Overall, when dynamic cues are available, behavioural models (i.e., Profiling, Categorical and Intensity Similarity) can better explain perceptual similarity compared to their performance in response to static stimuli.

It is important to consider that, in line with previous studies reported in this work, participants' similarity scores to identical stimuli were not excluded from our analyses. Also in this case, despite being recognized as highly similar, identical stimuli were often not rated as identical. While this decision may have slightly inflated the scores resulting from our correlation analysis, we also conducted analyses excluding trials containing identical faces, and the results were not significantly different.

Finally, I also tested whether algorithms-based emotion categorization could generate emotion profiles similar to those observed in human responses. The computational model I trained could categorize our stimuli with an accuracy of the 67%, which is quite high considering that these facial expressions were validated with an overall accuracy of 60% (Kaulard et al., 2012). However, I found no significant correlation between Model-based similarity and participants perceived similarity, neither at a group- or participants-level. When specifically comparing Human vs Model-based emotion profiles, while both the machine learning model and participants gave higher scores to the target emotion dimensions, responses to other emotion dimensions significantly differ, especially for Pain and Fear.

Specifically, when examining the model's performance in recognizing happy facial expressions, it tends to assign higher scores to dimensions of fear and pain compared to human observers. Likewise, for expressions of pain, the model primarily attributes scores to happiness and, for intense emotions, fear, while humans tend to score higher on disgust and surprise. In the case of fearful expressions, humans often perceive a significant element of surprise, even stronger than the fear itself, along with disgust. However, the model fails to detect surprise and assigns higher scores to pain and happiness instead. These disparities in detection could be attributed to the model relying solely on physical, image-based information, while humans likely incorporate a more nuanced understanding of emotion. For instance, facial expressions of happiness and pain often involve similar configurations of facial muscles and this shared similarity makes it challenging to distinguish between intense facial displays of happiness or pain when these are presented without contextual information (Barrett et al., 2011). This physical similarity might explain why the model attributes relevant scores to happiness when processing pain, and vice versa. However, it's crucial to acknowledge the specific limitations due to the small training set Although each emotion category in the training set contained 70 images, the proportion of images from different datasets varied slightly based on availability. For example, while the happy, fear, and pain sets included 18 images from the MPI dataset (the same dataset as the test images), surprise, disgust, and sadness contained only 12 images each, and anger none. Images from the same dataset tend to be more physically similar, and this imbalance across categories may cause the model to overrepresent certain emotions due to similarities unrelated to facial expression. This might explain why the model seldom attributes higher scores to surprise, anger, sadness, and disgust—categories with fewer MPI dataset images. Future analyses should consider using a larger training sample to address these issues effectively.

When recognizing posed facial emotion expressions, performance of machine models is often comparable or superior to human performance. However, many machine-based systems are mostly trained on a limited number of posed datasets (Pantic & Bartlett, 2007), and perform less effectively with non-stereotypical subtle expressions (Yitzhak et al., 2017) or expressions produced by lay people in a laboratory (Stöckli et al., 2018). Recently, efforts have emerged to develop algorithms that can process spontaneous non-stereotypical facial expressions, with some models achieving performance comparable to that of humans (Krumhuber et al., 2021). Our results suggest that computational models can achieve human-level performance when trained with non-stereotypical stimuli on a categorical one-label task. However, a fine-grained profiling approach reveals how human and computational models extract emotional content (i.e., emotion profiles) differently from the same faces.

In summary, Chapter 3 shows how perceivers extract much more than one target emotional label in facial expressions of emotions, and that a richer representation of facial emotions (i.e., emotion and semantic profiles) can reveal the impact that facial motion, emotion intensity and culture have on emotion perception. In this studies, like many others in the literature, we have examined emotional experiences without considering contextual information. However, a substantial body of evidence suggests that contextual information plays a significant role in shaping how we express and perceive emotions (Barrett et al., 2011; Greenaway et al., 2018; Wieser & Brosch, 2012). In Chapter 4, I present two behavioural studies conducted to specifically investigate the role of physical and social context scenarios in the perception of facial expression of emotion.

Chapter 4

The role of context in profiling facial expressions of emotions

4.1 Introduction

The studies reported in Chapter 3 have highlighted that facial expressions of emotions are better characterized as high-dimensional emotion and semantic profiles. People tend to establish a complex and rich representation of the emotional content conveyed by a facial expression, which is in turn linked to multiple related semantic dimensions that shape our personal experience of the emotion perceived. These studies have shown that both perceiver-based and stimulus-based information are important in forming complex emotion profiles and in processing and distinguish between facial emotions. However, it's important to note that these facial expressions of emotions are typically perceived in specific contexts. In our day-to-day lives, both the facial expression of emotion and the perceiver who interpret them are embedded in specific context which significantly contributes to how emotion is expressed, perceived, and regulated. The emotional meaning of facial actions is sometime constructed based on cues external to both the stimulus and the perceiver, originating from the surrounding environment (e.g., visual scene, other faces, social situations).

In previous literature, both functional and constructivist theories of emotion perception have highlighted the importance of context in processing facial expressions of emotion. Functional theories propose that emotions serve adaptive functions, evolved to optimize adjustment to physical (e.g., avoiding danger) and social environments (e.g., creating relationships) (Barrett & Campos, 1987; Johnson-Laird & Oatley, 1998; Levenson, 1994). Similarly, constructivist approaches suggest that emotional experiences, including valence and arousal, are constructed based on the interpretation of situational contexts (Barrett & Bliss-Moreau, 2009; Russell, 2003). Nonetheless, the predominant traditional basic emotion view states that the extraction of emotional content from faces is the result of automatic categorization of universal expressions (Ekman, 1992). Consequently, facial expressions are often studied in decontextualized, static pictures that maximize the distinction between emotion categories. According to functionalist theories, nowadays facial expressions of emotions would mostly be, as defined by Darwin, “serviceable associated habits”, which means that while they vestigially reflect their original adaptive functioning, they are now used for their functions as intra-individual regulation of thoughts and inter-individual regulation of social interactions. Thus, facial expressions of emotions are expected to be strongly associated with situational context that are in line with the function they developed from. Despite numerous studies exploring the role of context, particularly in the categorization of facial expressions of emotions, the distinction between contexts closely linked to evolutionary-based functions (referred in this work as "physical contexts") and those evolved to meet social situations (referred in this work as "social contexts") has not been thoroughly explored. In this chapter, I aimed to investigate how the perception of facial expressions of emotion differs when presented with and without contexts, how it varies when presented with congruent and incongruent contextual information and physical and social scenarios.

Numerous studies have consistently demonstrated the significant impact of contextual information on emotion perception (for relevant reviews, see Barrett et al., 2011; Greenaway et al., 2018; Wieser & Brosch, 2012). In a study by Carroll and Russell (Carroll & Russell, 1996), participants were asked to read stories setting a specific social situation as a context and were then required to rate different facial expression. Their findings indicated a strong influence of the context, leading participants to perceive the emotion conveyed by the story from the facial expression, even when the face expressed a different emotion (e.g., when the story described a painful situation, participants judged a face displaying fear as being in pain). In line with these results, Kim et al. (2004) demonstrated that brain responses to ambiguous emotional faces were modified by verbal descriptions of contextual conditions, illustrating context-dependent neural processing of the very same emotional face. Moreover, evidence has shown that perceivers' judgments of facial actions are influenced by accompanying body postures. For instance, facial portrayals of anger were more likely to be perceived as displaying disgust when combined with a body posture involving a soiled object. Similarly, facial portrayals of disgust were perceived as expressions of proud when combined with a muscled body whose arms are raised in triumph (Aviezer et al., 2008). In a series of studies, Righart and de Gelder (2006, 2008), investigated the impact of visual context on facial emotion recognition and its neural processing. They examined event-related brain potentials in response to fearful or neutral faces embedded in fearful or neutral visual scenarios. Their results revealed that information from facial expressions was integrated with the visual context during the early stages of face processing. Specifically, N170 amplitudes were significantly increased for fearful faces in fearful scenes compared to those in happy scenes, demonstrating that these contextual influences were perceived early and automatically during emotion processing. In a follow-up study, their participants categorized facial expressions embedded in emotionally congruent or incongruent visual scenes. They found that participants were faster at categorizing facial expression presented in congruent contexts, and this advantage remained unaffected even when the task load increased. Likewise, using videos as visual context, positive and negative contexts elicited significantly different ratings of faces compared to those presented in neutral contexts (Mobbs et al., 2006). Together, these finding suggest that information conveyed by facial expressions is automatically integrated with the information conveyed by its context during emotion processing, highlighting the relevant role of context in shaping the way we perceive and represent facial emotions.

The studies reviewed above primarily focused on examining how contextual information affects recognition or categorization of facial emotions. However, considering the findings from my previous studies, suggesting that facial emotions are often perceived as rich and complex emotion profiles, the impact that contextual information may have on the emotional profiles conveyed by facial expressions remains unexplored. Similarly, we have limited knowledge of how contextual information may differently influence emotion perception across facial motion, emotion intensity, and cultural background. To gain a more holistic understanding of the complexity of our daily emotion experiences, I conducted two cross-cultural studies to (1) assess the strength of the association between a given facial expression of emotion and congruent/incongruent emotional scenarios, and (2) explore how the profiles of emotion perception are influenced when these expressions are presented within emotionally congruent/incongruent scenarios.

4.2 Study 5. Are facial emotions strongly associated with congruent emotional contexts?

The aim of Study 5 was to investigate whether and to what extent facial expressions of emotions are associated to their congruent emotional contexts and to assess whether facial motion, emotion intensity and culture may influence these associations. According to functional theories of emotion processing, emotions have evolved as adaptive responses to environmental challenges by prioritizing and organizing human behaviour that optimise the expresser's adjustment to the demands of the physical and social environment. Emotions originate in addressing survival-relevant problems including forming attachments, maintaining cooperative relations, or avoiding physical threats (Barrett & Campos, 1987; Ekman, 1992; Johnson-Laird & Oatley, 1998; Lazarus, 1991; Levenson, 1994; Oatley & Jenkins, 1992). As a result, nowadays facial expressions of emotions would mostly be, as defined by Darwin, "serviceable associated habits", which means that while they vestigially reflect their original adaptive functioning, they are now used for their functions as intra-individual regulation of thoughts and inter-individual regulation of social interactions. For instance, Shariff and Tracy (2009) found a robust implicit association between the expression of pride and the concept of high-status, suggesting that facial expressions of pride send a functional signal about a social group member's increased social status, which may serve to intimidate potential challengers in competitive contexts. According to the functionalist account of the origins and functions of emotions, facial expressions of emotions are expected to be strongly associated with situational context that are in line with the function they developed from.

The assumption of a strong association between facial expressions of emotions and their corresponding contexts has been widely adopted, explicitly or implicitly, in the field of emotion research. This is particularly evident when researchers aim to induce or recreate specific facial emotions in participants by employing situational contexts. This has been achieved by asking participants to imagine themselves in a specific emotional situation or by actually reproducing the situation and making them experience the emotion live. For example, emotion of anger has been elicited by staging a situation in which participants are insulted (Harmon-Jones & Sigelman, 2001) or by instructing participants to imagine being insulted (Keltner et al., 1993). Similarly, different facial expressions datasets have been obtained by asking participants to reproduce the facial emotions elicited by specific scenarios (Kaulard et al., 2012). All these approaches are rooted in the assumption that we consistently associate specific contextual scenarios with specific emotional experiences and corresponding facial expressions.

To validate the association between contextual scenarios and facial expression of emotions, previous studies have often examined how consistently target emotions can be recognized by other observers (i.e, are happy faces consistently perceived as showing happy emotion across participants?). However, there has been relatively little research into how effectively the produced facial expressions can be linked back to their triggering contextual scenarios. Here, I employed the same stimuli as in my previous studies, which were specifically chosen to convey spontaneously evoked facial expressions of emotions, to address this question. Similar to the studies in previous chapters, I obtained a context profile for each facial emotion to explore the rich context information that is connected to these facial emotions. To do so, I had participants rate the likability of specific facial expressions displayed within a range of context scenarios. The scores provided for each context dimension formed participants' unique response profile (i.e., context profile). In particular, I selected various emotional scenarios that elicited intended emotion through either a physical stimulation (e.g., "You're eating your favourite food in your favourite place") or social stimulation (e.g., "You won the first prize of a big competition"). I then investigated the extent to which participants associated facial expressions of emotions with different congruent or incongruent emotional scenarios and how facial motion, emotion intensity and culture may influence these associations. Finally, I tested whether context information is related to the way we perceive differences and similarities between facial expressions of emotions and, if so, whether this correlation is stronger within social or physical scenarios.

4.2.1 Methods

Participants

The size of the participants' sample was based on the results from our previous studies, suggesting that a minimum of 13 participants was required. Again, considering the fact that the study was conducted online we collected 20 participants for condition (dynamic/static).

Forty-two British participants were recruited from the University of East Anglia via the SONA System (7 males, 35 females; age ranged between 18-53 yrs., M = 22.7; SD = 8.36). Participants who did not identify themselves as British in the demographic questionnaire were excluded from the study. Forty Chinese participants were recruited from the Sun Yat-sen University in China (6 males, 34 females; age ranged between of 19-27 yrs., M = 21.5; SD = 2.12). All participants were naïve to the purpose of the investigation and had not taken part in any previous studies of this project. All participants provided informed consent before taking part in the study and were debriefed at the end, receiving course credits as compensation. The study's experimental procedure was approved by the Ethics Committee of the School of Psychology at UEA.

Stimuli, Materials, and Tasks

The stimuli were the same as in previous studies of this project, including images or videos of 9 actors displaying 3 different emotions (Happy, Pain, Fear) at 2 intensities (High, Low) taken from the MPI Facial Expression Database (Kaulard et al., 2012). The study had a Context profiling task and a similarity rating task. In the first half of the Context profiling task, all 54 facial emotions (i.e., 3 emotions * 2 intensities * 9 actors) were shown randomly and rated along 6 dimensions, i.e., physical scenarios. In the second half of the task, the same 54 facial emotions were presented again, in a random order, and were rated along other 6 dimensions, i.e., social scenarios. Thus, the first task presented a total of 108 trials, 18 trials for each combination of emotion/intensity (e.g., Happy emotion high intensity for 9 actors presented twice). The materials and structure of the Similarity Rating Task were the same as in previous studies, consisting of a total of 160 trials (see Figure 29b).

The Context profiling task was created using the *Gorilla* experiment builder. Each trial started with a fixation cross (1000ms), followed by the stimulus displayed on the left half of the screen (i.e., image for the Static task, video for the Dynamic task) and 6 different sliders on the right half of the screen. Participants were instructed to assess the likelihood of the displayed facial expression occurring in each of the six specified scenarios, which were labeled above each slider, by adjusting the position of the sliders accordingly (see Figure 29a). Each slider was independent from the others and ranged from 0 to 100, the handles of all sliders were initially placed on 0. To ensure participants had sufficient time to provide their answers without overthinking, preventing the interference of more high-level processes, the response screen had a time limit of 35 seconds. Dynamic stimuli were presented on a loop until a decision was made or the time limit was reached. If the time limit was reached, the experiment moved on to the next trial. The context scenarios above the 6 sliders were presented in two different groups, 6 Social scenarios and 6 Physical scenarios. During the first half of the Context profiling task, participants rated the stimuli according to the physical scenarios, then after a 5-minute break, they rated the same stimuli according to the social scenarios.

The same translation process employed in the previous study was followed, aiming to ensure accurate translation and maintain consistency across languages. Also in this case, it is crucial to recognize the potential for semantic differences between languages, which may result in subtle variations in interpretation.

Finally, the same measures taken in previous studies were implemented with the aim to reduce, and account for, potential participants' distractions or disengagement. In particular, (1) access to the study was restricted to PCs or laptops; (3) participants were required to self-report the reliability/usefulness of their data (e.g., due to lacked attention) at the end of the experiment, (4) a time-limit was imposed on each screen, and participants exceeding the maximum allotted time for completing the experiment were excluded from the study; (5) control trials were included in the study, where participants were asked to rate the similarity of identical images.

Emotion-evoking contexts

To formulate a list of emotion evoking contexts, I conducted a literature review with the aim of identifying validated emotional scenarios, episodes, or topics intended to evoke an emotional state of happiness, fear, or pain. The emotion story method, where critical contextual information is added around emotion word concepts (e.g., "You have just met your friend and feel very happy that your friend is here") is widely used in cross-cultural studies where the exact translation of single emotion words can be challenging (Russell & Sato, 1995). I reviewed, specifically for the three emotions of interest, recent cross-cultural studies where this methodology was successfully applied (Cordaro et al., 2016, 2018; Sauter et al., 2010; Simon-Thomas et al., 2009). Furthermore, to have a deeper insight into the prototypical characteristics that people associate with emotional events I explored descriptions of emotional episodes by participants' direct experience. For instance, in a study by Shaver et al. (Shaver et al., 1987), subjects were asked to write descriptions of actual episodes in which they experienced fear, sadness, anger, joy, and love, resulting in a list of prototypical features that characterized each emotion category (e.g., Fear "being alone, threat of social rejection, sweating, etc.").

Based on the literature review, I formulated an initial list of 33 text-based contexts following 3 guiding criteria: (a) conciseness—all stories were one sentence long, (b) simplicity—the stories described everyday events, and (c) thematic universality—the stories centered upon simple events likely to occur across different cultures. Then, I ran a pilot study with the aim to select the 12 sentences that were better recognized as evoking the intended emotions at the intended level of intensity. I asked 10 English volunteers, from the University of East Anglia, and 10 Chinese volunteers, from the University of Sun Yat-sen, to select the emotion, or emotions, that would be induced by each scenario and rate its intensity. Participants could choose one or more of the following options: "Happy", "Fear," "Pain", "None of the above", and rate the intensity by clicking on a 7-points scale. We selected the scenarios that were mostly recognized to convey the intended emotion at a level of intensity that allow to significantly differentiate between high/low intensity. Importantly, each scenario would induce the target emotion either due to physical stimulation, or a socially meaningful event (e.g., “You're eating your favourite food in your favourite place” or “You won the first prize of a big competition”), and at two levels of intensity (e.g., “You hear a strange sound while walking in the woods” or “You find a snake slithering into your sleeping bag”). This led us to a total of 12 scenarios [3 emotions * 2 intensities * 2 stimulation types (social vs physical)]. A complete list of the 12 scenarios can be found in Appendix C, Figure C.1.

Procedure

Participants performed both the Context profiling and Similarity rating task online through the Gorilla platform (i.e., https://gorilla.sc/). They gained access through a URL link using their desktop computer or laptop (i.e., no tablet or phone access were allowed). The study followed the same general procedure for all participants. Once consent was given, participants were directed to a demographic questionnaire. Closed-ended questions asked for their hand dominance (right-handed/left-handed) and the gender they identify with (male/female/other); while open-ended questions asked for their age and nationality. Half of the participants were then randomly allocated by the system to the Dynamic version of the study, with movies as stimuli, while the other half to the Static version, with images as stimuli. Participants completed the Context profiling task first, rating each stimulus on the 6 Physical scenarios, and, after a break, on the 6 Social scenarios. Then, they went through the Similarity rating task (see Figure 29). They had a few minutes break between the two tasks and again in the middle of the Similarity rating task. Detailed instructions and three example trials were given before starting both tasks. The whole study took about 1 hour to complete. At the end of the study, they filled out a self-report questionnaire regarding the reliability of their data and were then debriefed.

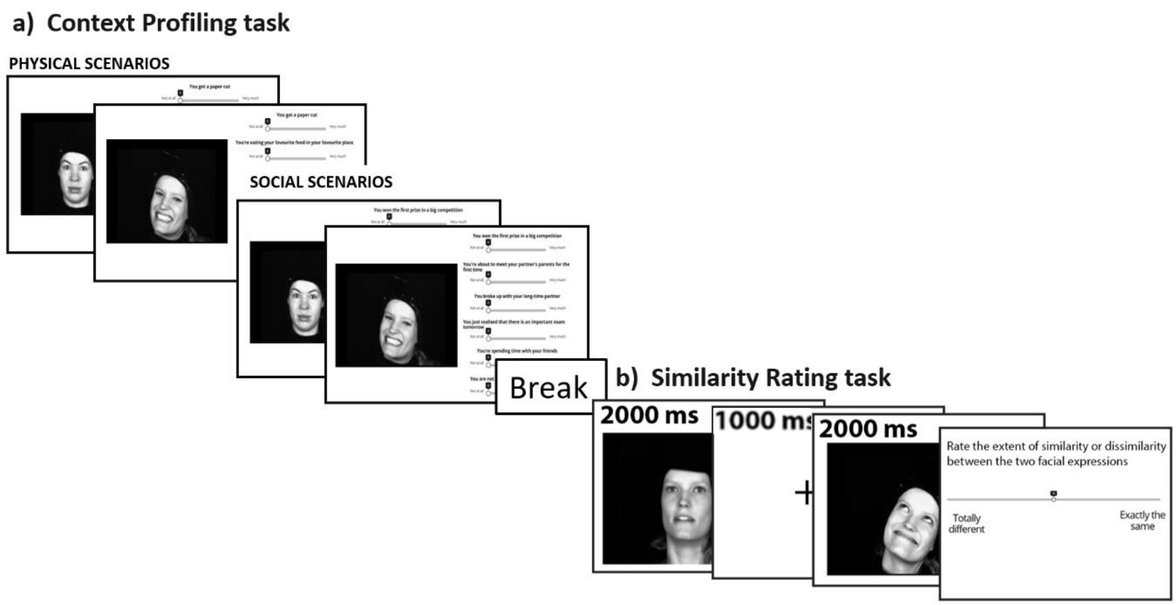

Figure 29. **Context profiling and similarity rating tasks**. *a) In the Context profiling task, participants first rated how likely it was for each facial emotion to appear in 6 different Physical scenarios (e.g., The person is eating their favorite food in their favorite place). Then, they rated the same stimuli along 6 Social scenarios (e.g., The person won the first prize of a big competition). b) In the similarity rating task, participants saw two facial emotions in a row. Then, using a slider ranging from 0 to 100, they rated the degree of similarity between them.*

4.2.2 Results and Discussion

Prior to commencing data analysis, we applied the same exclusion criteria as in previous studies to address potential limitations associated with online testing. Specifically, participants were excluded if they failed to respond to more than 50% of the trials in the Context profiling task, rated more than half of identical facial emotions as below 50 (on a 100-point scale for similarity rating) or reported data unreliability in the final questionnaire. Applying these criteria, one British participant was excluded, resulting in a final sample of 41 British participants and 40 Chinese participants.

Facial emotions had a stronger association with physical scenarios than with social scenarios.

To investigate the strength of the association between facial emotions and congruent physical or social emotional scenarios, I examined whether responses to the six context dimensions in the Context profiling task align with a categorical mapping, characterized by high scores for congruent contexts and near-zero scores for incongruent contexts, or if they reflect a more diverse and comprehensive profile with higher scores across multiple context dimensions.

Overall responses to the 6 emotional scenarios averaged across culture (British/Chinese), stimulus types (Static/Dynamic), and intensity levels (Low/High) are shown in Figure 30a (for details of context profiles before averaging, see Appendix C, Figure C.1 and C.2). The results revealed that participants employed multiple contextual dimensions to interpret facial expressions of Happy, Fear, and Pain. Notably, the context profiling responses varied depending on whether the scenarios induced emotions through physical or social stimulation, with response profile for social scenarios demonstrated a greater richness and diversity compared to the response profile of physical scenarios, which is mostly characterized by prominent scores in target congruent contexts.

In particular, when judging how likely it is for facial expressions of Pain to be shown in 6 different physical contexts, participants gave higher scores to scenarios evoking Pain (Pain scenarios, for high 47%, SD =17.6; for low 45%, SD = 17.6), followed by high and low intensity Fear scenarios (for high 40%, SD =19.4; for low 29%, 18.9), and lower, but significantly above zero (all *t*s(80) ≥ 8.59, p <.001) scores for high and low intensity Happy scenarios (for high 12%, SD = 8.38; for low 8%, SD = 8). In contrast, when judging the same Pain expression in response to social scenarios, the higher score was given to the scenario evoking high-intensity Fear (49%, SD = 19.2), followed by low and high intensity Pain (for low 39%, SD = 17.9; for high 36%, SD = 18.9), low-intensity Fear (22%, SD = 16.6), and lower but significantly above zero (all *t*s(80) ≥ 12.08, p <.001) scores for high and low-intensity Happy (for high 20%, SD = 12.3; for low 16%, SD = 12).

To compare the two context profiles for Pain, I conducted a 2 (stimulation type: physical vs social contexts) by 6 (context response dimensions) repeated measure ANOVA. Results showed a significant main effect of response dimensions, $F(5,400) = 168.42$, $p < .001$, $\eta p^2 = .678$, indicating that participants provided significantly different scores for the six dimensions. There was no significant difference in overall responses (across emotion dimensions) between Physical vs Social scenarios, $F(1,80) = 0.05$, $p = .828$, $\eta p^2 = .001$, while, most importantly, the interaction between stimuli type and response dimensions was significant, $F(5,400) = 40.9$, $p < .001$, $\eta p2 = .338$. Paired t-tests indicated a significant difference in all dimensions between responses to Social vs Physical scenarios (all *t*s(80) ≥ 3.37, all *p*s ≤ .001), with higher scores given to Physical than to Social pain scenarios.

Similar results were found in response to facial expressions of Fear and Happy. For Fear, responses to Physical scenarios showed higher scores for high and low intensity Fear scenarios (for high 51%, SD = 17.6; for low 57%, SD = 18), which were also characterized by relatively high score for Pain scenarios (for high 24%, SD = 19; for low 28%, SD = 17.7), and lower but significantly above 0 (all *t*s(80) ≥ 7.38, p <.001) scores to Happy scenarios (for high 8%, SD = 8.54; for low 5%, SD = 6.04). Again, the response profile to Social scenarios was richer and more complex. The higher score was observed for the high-intensity Fear scenario (61%, SD = 18.1), followed by scenarios intended to evoke low-intensity Pain (48%, SD = 19.9), low-intensity Fear (27%, SD = 17.9), high-intensity Happy (25%, SD = 15.3), high-intensity Pain (25%, SD = 18.5), and, finally, low but significantly above 0 (all *t*s(80) ≥ 6.71 , all *p*s≤. .001) score to the low-intensity Happy scenario (10% SD = 12.9).

The same 2 (stimulation type: physical vs social contexts) by 6 (response dimensions) ANOVA showed a significant main effect of response dimensions, $F(5,400) = 290.1$, $p < .001$, $\eta p^2 = .784$, a significant interaction between stimulation type and response dimensions, $F(5,400) = 127.8$, $p < .001$, $\eta p^2 = .615$, and, in this case, a significant difference in overall responses between Physical vs Social scenarios, $F(1,80) = 16.4$, $p < .001$, $\eta p^2 = .017$. Paired t-tests indicated a significant difference in all dimensions between responses to Social vs Physical scenarios (all ts(80) ≥ 5.65 , all ps ≤ .001) except for the high-intensity Pain scenario ($t(80) \geq 0.58$, $p = .560$).

Finally, for Happy, response to Physical scenarios showed higher scores to Happy scenarios (for high-intensity 61%, SD = 16.3; for low-intensity 57%, SD = 18.4), and much lower but above-zero (all ts(80) ≥ 6.12 , all ps≤ .001) scores to other scenarios (for low-intensity Fear 8%, SD = 11.2; for low-intensity Pain 7%, SD = 9.49; for high-intensity Fear 6%, SD = 8.25; for high-intensity Pain 5%, SD = 6.08). In contrast, responses to Social scenarios showed high score to the target Happy scenarios (for high-intensity 53%, SD = 18.5; for low-intensity 58%, SD = 14.7), followed by relatively high score to scenarios intended to evoke Fear (low-intensity, 30%, SD = 18.7; high-intensity, 12%, SD = 11.2), and Pain (for high-intensity, 9%, SD = 9.8; for low-intensity, 11%, SD = 9.90).

A 2 by 6 repeated measures ANOVA showed a significant main effect of response dimensions, $F(5,400) = 519$, $p < .001$, $\eta p^2 = .866$, a significant difference in overall responses between Physical vs Social scenarios, $F(1,80) = 55.2$, $p < .001$, $\eta p^2 = .040$, and a significant interaction between stimulation type and response dimensions $F(5,400) = 43$, $p < .001$, $\eta p^2 = .350$. Also in line with previous results, paired t-tests indicated a significant difference in all dimensions between responses to Social vs Physical scenarios (all ts(80) ≥ 4.90, all ps≤ .001), except for low-intensity Happy ($t(80) = .08$, $p = .936$).

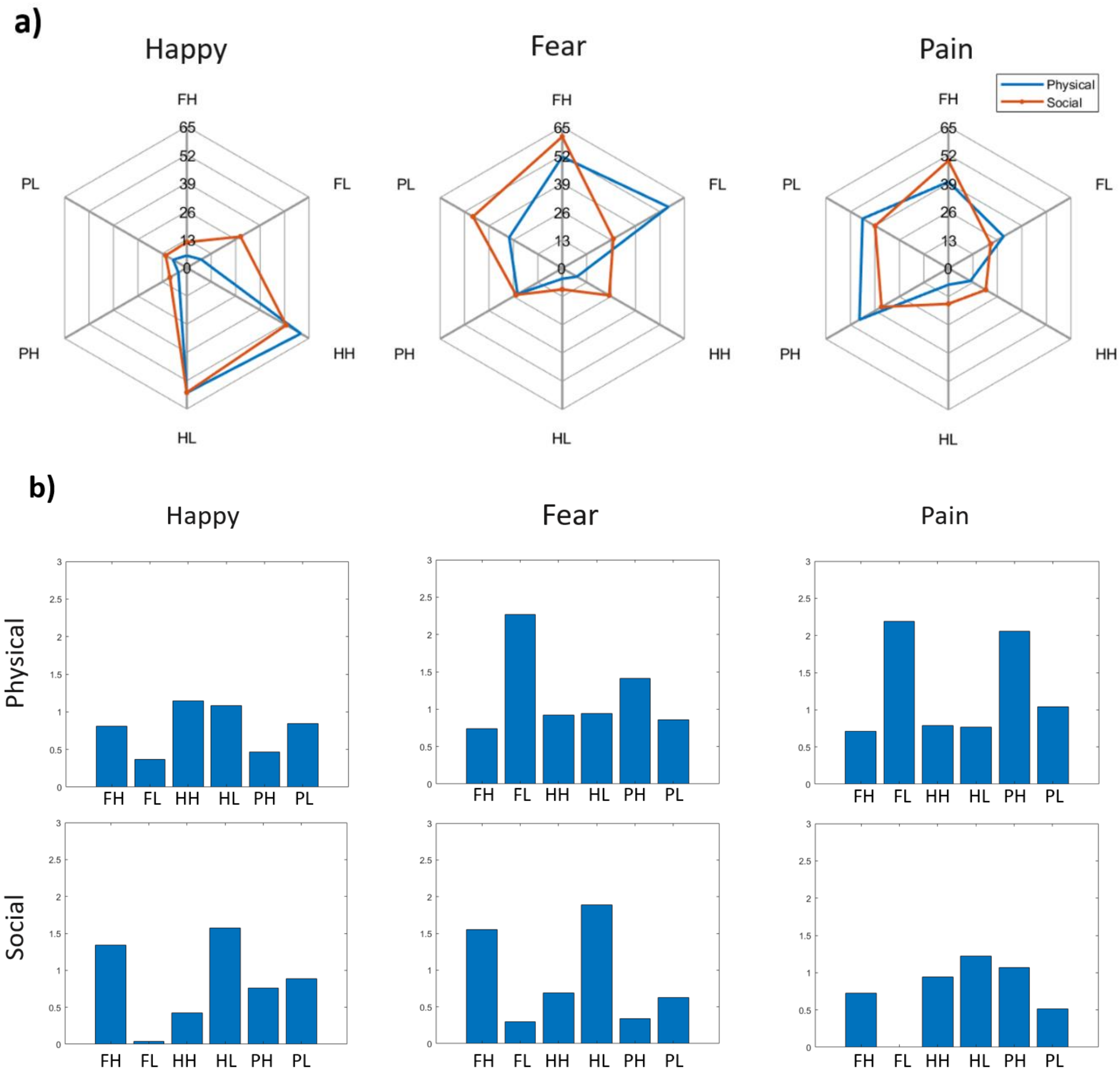

Figure 30. **Context Profiles and Feature importance in perception of facial emotions of Happy, Fear and Pain.** *a) Averaged profiling responses to physical (blue) and social (orange) contexts. The dimension of context was denoted by its emotions (happy H-, fear F-, pain P-) and intensity, high or low (-H, -L). b) Feature (i.e., dimension) importance, estimated based on the decrease in prediction accuracy of a classifying model (i.e., Random Forest) when the values of a particular feature are randomly permuted while keeping other features unchanged. Results are the averaged values obtained from 10 iterations. It is worth noting that feature importance represents the relative contribution of each feature in the model's decision-making process, and that the units of measurement are arbitrary.*

To further investigate how perception of facial emotions is linked to its evoking contexts, a Random Forest machine learning model was utilized, following the same methods used in Chapter 3. The model was trained to classify participants' context profiles into the three emotion categories, and a feature importance algorithm (Out-of-Bag feature importance) was employed to identify the dimensions that significantly contribute to the model's predictions. The output represents the relative contribution of each dimension to the model's decision-making process, and the units of measurement are arbitrary. To provide a more robust and representative estimation of the importance of each dimension I have ran the feature importance analysis 10 times and averaged the results, this process helps to mitigate the potential impact of random variations in the model training process. The rankings across different iterations remained relatively stable, suggesting the consistency and reliability of the output values (all SDs ≤ 0.20). This approach allowed us to determine which dimensions primarily characterize a profile and differentiate it from others. It is worth noting that important dimensions may not necessarily exhibit higher scores in participants' responses, but rather possess a consistent score that distinguishes them from the other profiles. The results are shown in Figure 30b.

The results indicate that context profiles generated in response to the facial emotions of Happy, Fear, and Pain are characterized by multiple dimensions. Notably, the importance of different dimensions for the same emotion category varied depending on whether the contextual scenarios were social or physical. Specifically, when examining physical profiles generated in response to facial expressions of Fear, the dimension of low intensity Fear (2.26) emerged as the most crucial predictor for distinguishing context profiles of Fear from those to other emotion categories. The importance of high intensity Pain (1.4) was also highlighted, with the remaining dimensions displaying relatively lower importance to the model's predictions. However, for social scenarios, low intensity Fear is detected as the least important (0.04) while low intensity Happy emerged as the most important dimension (1.57), followed by high intensity Fear (1.34), low and high intensity Pain (0.89 and 0.76, respectively), and, finally, high intensity Happy (0.42).

For facial expressions of Happy, within physical scenarios, higher importance scores were observed for the dimensions of high and low intensity Happy (1.15, 1.08, respectively), followed by the dimensions of low intensity Pain and high intensity Fear (0.84, 0.80, respectively), and finally, high intensity Pain and low intensity Fear (0.47, 0.37, respectively). However, again, the pattern of importance slightly changed in social scenarios. While high intensity Happy remained the most important dimension (1.89), this was followed by high intensity Fear (1.55), then, with lower scores to high intensity Happy, low intensity Pain (0.69, 0.63, respectively), and, with close to zero scores for high intensity Pain and low intensity Fear (0.34, 0.29, respectively).

Lastly, for facial expressions of Pain, physical scenarios also exhibited different importance scores compared to social scenarios. For physical scenarios, higher importance scores were attributed to the dimensions of high intensity Fear (2.19) and high intensity Pain (2.05), followed by low intensity Pain (1.03), high and low intensity Happy (0.79 and 0.76, respectively) and high intensity Fear (0.71). For social scenarios, higher score was detected for low intensity Fear (1.22) and high intensity Pain (1.06), followed by high intensity Happy (0.94), high intensity Fear (0.72), low intensity Pain (0.51) and a negative score for low intensity Fear (-0.01), which suggests the irrelevance of this last dimension for the model's performance.

Context Profiles were sensitive to Facial Motion, Emotion Intensity and Culture

To test whether Context Profiles are affected by facial motion (i.e., Dynamic vs Static stimuli), emotion intensity (i.e., High vs Low) and participants' cultural background (i.e., Chinese vs British participants), I obtained and analysed contrasting context profiles for physical and social scenarios in response to the three facial emotions. As in the previous studies, these profiles were derived by averaging participants' responses across all conditions except the condition being contrasted (for details of context profiles before averaging, see Appendix C, Figure C.1 and C.2). The obtained context profiles were then submitted to a 2 (British vs Chinese / Dynamic vs Static stimuli / High vs Low emotion intensity) by 6 (response dimensions) ANOVA. Despite an overall similarity across Dynamic/Static stimuli, High/Low emotion intensity, and British/Chinese participants, context profiles showed a fine-grained sensitivity to these conditions.

British vs Chinese participants' responses. For facial expressions of Happy, responses to Physical scenarios (Figure 31a top panel) showed a significant effect of response dimensions, $F(5,395) = 518$, $p < .001$, $\eta p^2 = .868$, indicating that participants provided significantly different scores to the six scenarios. There also was a significant difference in overall responses (across emotion dimensions) between British vs Chinese participants, $F(1,79) = 7.45$, $p = .008$, $\eta p^2 = .086$, and a significant interaction between Culture and response dimensions, $F(5,395) = 3.26$, $p = .007$, $\eta p^2 = .040$. Post-doc independent t-tests showed that, compared to British participants, Chinese participants gave higher scores to non-target contexts related to Pain and Fear (all ts(79) $\geq$ 2.44, all ps $\leq$ .017), while there was no difference in responses given to the target dimensions, high and low intensity Happy (both ts(79) $\leq$ 1.75, both ps $\geq$ .083).

Similarly, responses to Social scenarios (see Figure 32a top panel) showed a significant effect of response dimensions, $F(5,395) = 287.39$, $p < .001$, $\eta p^2 = .784$, a significant difference in overall responses between British vs Chinese participants, $F(1,79) = 4.69$, $p = .033$, $\eta p^2 = .056$, and a significant interaction between Culture and response dimensions, $F(5,395) = 7.96$, $p < .001$, $\eta p^2 = .092$. Again, post-doc independent t-tests revealed that, compared to British participants, Chinese participants' responses were higher for the non-target dimensions of Pain and Fear (all ts(79) $\geq$ 2.09, all ps $\leq$ .039). Also, while no significant difference was found between British and Chinese responses to high intensity Happy scenario ($t(79) = 0.49$, $p = .622$), British participants gave significantly higher score to low intensity Happy scenario ($t(79) = 2.34$, $p = .021$).

For facial expressions of Pain, for Physical scenarios (see Figure 31a bottom panel), there was an effect of response dimensions, $F(5,395) = 172.21$, $p < .001$, $\eta p^2 = .686$, a significant effect of culture, $F(1,79) = 5.06$, $p = .027$, $\eta p^2 = .060$, but no significant interaction between response dimensions and Culture, $F(5,395) = .636$, $p = .673$, $\eta p^2 = .008$. Chinese participants gave overall higher scores compared to British participants. Independent t-tests revealed that particularly for the dimensions of high and low intensity Happy, Chinese participants showed significantly higher scores than British participants (both ts(79) $\geq$ 3.24, both ps $\leq$ .002).

For social scenarios (see Figure 32a bottom panel), it was found an effect of response dimensions, $F(5,395) = 91.17$, $p < .001$, $\eta p^2 = .536$, a non-significant effect of culture $F(1,79) = 1.05$, $p = .308$, $\eta p^2 = .013$, and a significant interaction between culture and response dimensions $F(5,395) = 3.34$, $p = .006$, $\eta p^2 = .041$. While overall responses to all dimensions were similar between Chinese and British participants (all ts(79) $\leq$ 1.69, all ps $\geq$ .094), scores assigned to high and low intensity Happy scenarios were again significantly higher for Chinese compared to British participants (both ts(79) $\geq$ 2.56, both ps $\leq$ .012).

Finally, for facial expressions of Fear (see Figure 31a and Figure 32a, middle panels), I found an effect of response dimensions for response to both physical, $F(5,395) = 304.33$, $p < .001$, $\eta p^2 = .794$, and social scenarios, $F(5,395) = 173.60$, $p < .001$, $\eta p^2 = .687$. However, differently from responses to Happy and Pain, there was no significant effect for culture (for Physical scenarios, $F(1,79) = 3.21$, $p = .077$, $\eta p^2 = .039$; for Social scenarios, $F(1,79) = 1.17$, $p = .282$, $\eta p^2 = .015$), nor for the interaction between culture and response dimensions (for Physical scenarios, $F(5,395) = .160$, $p = .977$, $\eta p^2 = .002$; for Social scenario, $F(5,395) = .829$, $p = .529$, $\eta p^2 = .010$).

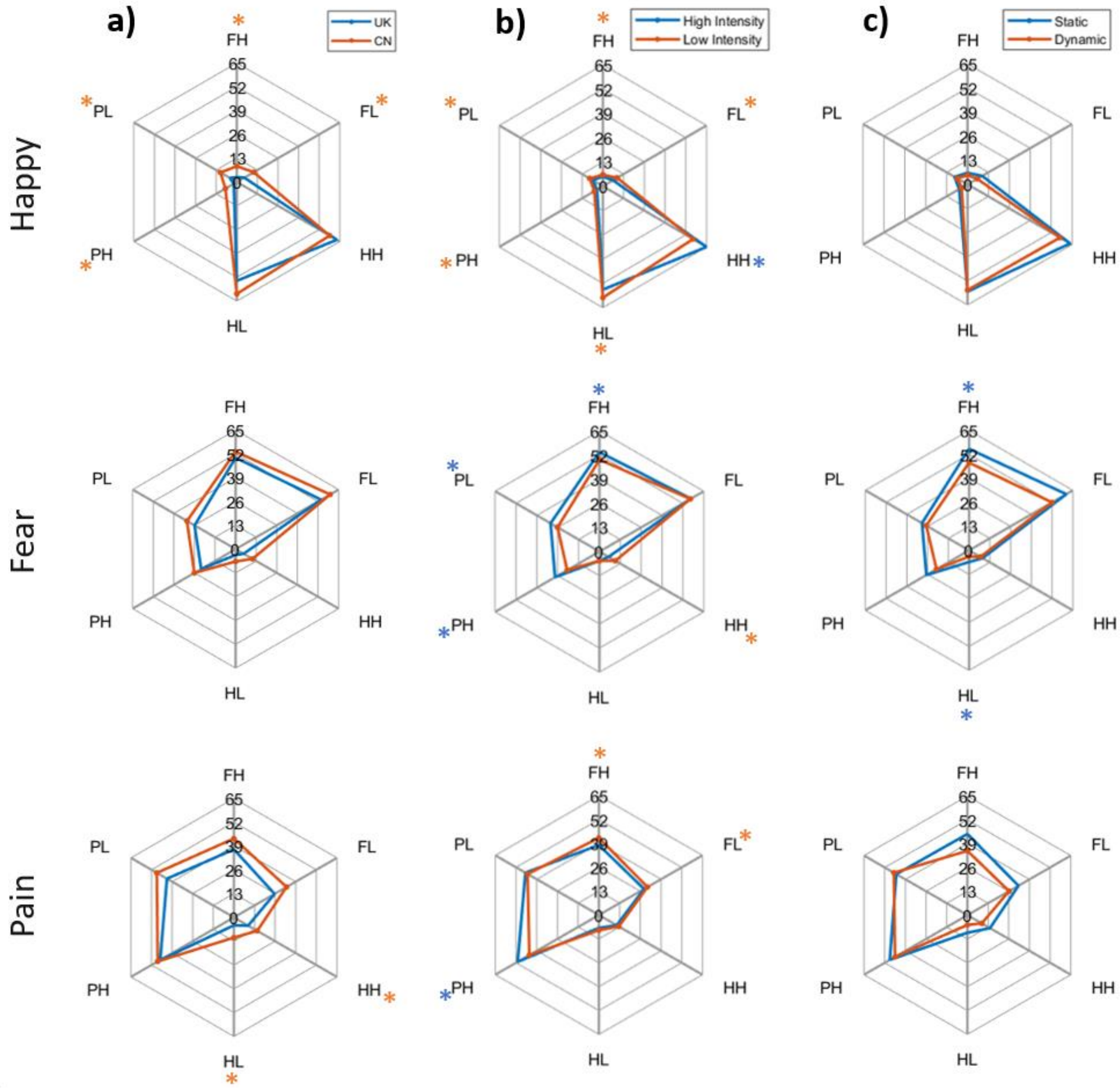

Figure 31. **Physical context-profiles as a function of Culture, Facial Emotion and Emotion Intensity.** *a) Responses of British (blue) and Chinese (orange) participants to facial expressions of Happy, Fear and Pain. b) Responses to High and Low intensity facial emotions of Happy, Fear, and Pain. c) Responses to dynamic and static facial expressions of Happy, Fear, and Pain. Asterisks indicate significant differences.*

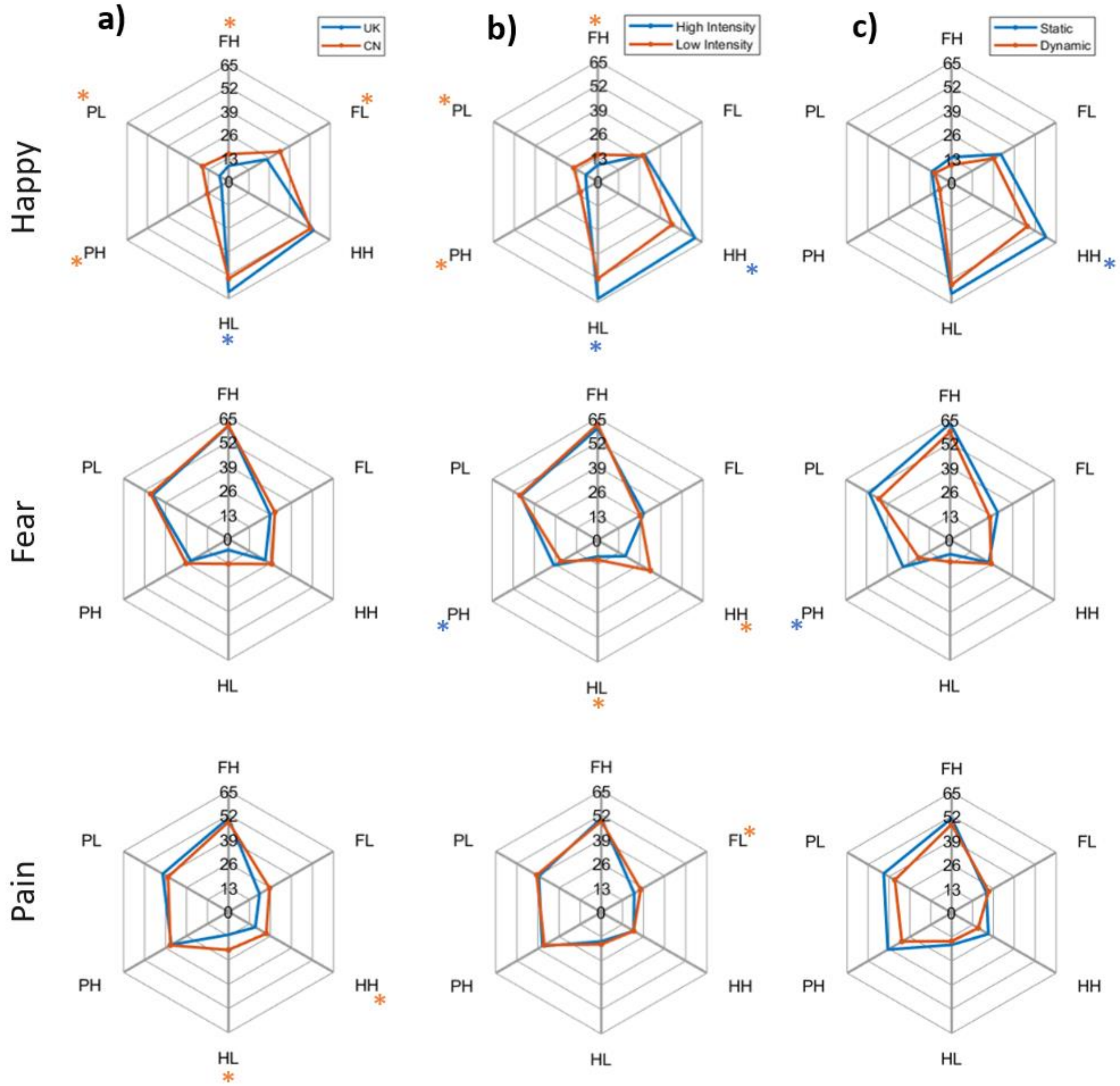

Figure 32. **Social context-profiles as a function of Culture, Facial Emotion and Emotion Intensity.** *a) Responses of British (blue) and Chinese (orange) participants to facial expressions of Happy, Fear and Pain. b) Responses to High and Low intensity facial emotions of Happy, Fear, and Pain. c) Responses to dynamic and static facial expressions of Happy, Fear, and Pain. Asterisks indicate significant differences.*

Static vs Dynamic facial expressions. For facial expressions of Happy (see Figure 31c top panel), responses to physical scenarios showed a significant effect of response dimensions, $F(5,395) = 497.16$, $p < .001$, $\eta p^2 = .863$, a non-significant effect of facial motion $F(1,79) = 2.29$, $p = .134$, $\eta p^2 = .028$, and a non-significant interaction between facial motion and response dimensions, $F(5,395) = .684$, $p = .636$, $\eta p^2 = .009$. However, responses to social scenarios (see Figure 32c top panel) showed, besides a significant effect of response dimensions, $F(5,395) = 262.91$, $p < .001$, $\eta p^2 = .769$, a significant effect of facial motion, $F(1,79) = 6.54$, $p = .012$, $\eta p^2 = .076$. There were overall higher scores in response to static than dynamic facial expressions. There was no interaction between facial motion and response dimensions, $F(5,395) = 1.58$, $p = .164$, $\eta p^2 = .020$.

For facial expression of Fear, responses to Physical scenarios (see Figure 31c middle panel) showed a significant effect of response dimensions, $F(5,395) = 304.04$, $p < .001$, $\eta p^2 = .794$, an effect of facial motion, $F(1,79) = 3.83$, $p = .054$, $\eta p^2 = .046$, but a non-significant interaction between response dimensions and facial motions, $F(5,395) = 1.52$, $p = .184$, $\eta p^2 = .019$. Again, there were overall higher scores in response to static than dynamic facial expressions. Responses to social scenarios (see Figure 32c middle panel), showed an effect of response dimensions, $F(5,395) = 176.19$, $p < .001$, $\eta p^2 = .690$, a non-significant effect of facial motion, $F(5,395) = 3.41$, $p = .005$, $\eta p^2 = .041$, and a significant interaction between response dimensions and facial motion $F(1,79) = 1.30$, $p = .258$, $\eta p^2 = .016$. Independent t-tests revealed a significant difference in the dimension of high intensity Pain ($t(79) = 2.42$, $p = .018$), with scores being higher for static stimuli, but not in other dimensions (all $ts(79) \leq 1.43$, all $ps \geq .155$).

Finally, for facial expressions of Pain, regarding both physical and social scenarios (see Figure 31c bottom panel, and see Figure 32c bottom panel), there was an effect for response dimensions (for physical scenarios, $F(5,395) = 174.49$, $p < .001$, $\eta p^2 = .688$; for social scenarios, $F(5,395) = 88.51$, $p < .001$, $\eta p^2 = .528$), but no significant effect of facial motion (for physical scenarios, $F(1,79) = 2.94$, $p = .090$, $\eta p^2 = .036$; for social scenarios, $F(1,79) = 2.67$, $p = .107$, $\eta p^2 = .033$) nor interaction between response dimensions and facial motion (for physical scenarios, $F(5,395) = 1.86$, $p = .100$, $\eta p^2 = .023$; for social scenarios, $F(5,395) = 1.87$, $p = .099$, $\eta p^2 = .023$).

***High vs low intensity facial expressions*.** For facial expressions of Happy, responses to physical scenarios (see Figure 31b top panel) showed a significant effect of response dimensions, $F(5,400) = 504.40$, $p < .001$, $\eta p^2 = .863$, a non-significant difference between overall responses to high low intensity stimuli, $F(1,80) = 1.54$, $p = .219$, $\eta p^2 = .019$, and a significant interaction between response dimensions and intensity $F(5,400) = 19.23$, $p < .001$, $\eta p^2 = .194$, Follow-up t-tests revealed that participants gave significantly higher scores to high intensity Happy scenarios when the stimuli were depicting facial expression of high intensity compared to stimuli of low intensity Happy ($t(80) = 5.93$, $p < .001$). Responses to all other dimensions were also significantly different, with higher scores in response to scenarios of low intensity Happy [for the dimensions of low intensity Happy ($t(80) = 2.67$, $p = .009$), High and low intensity Pain (for high, $t(80) = 2.70$, $p = .009$; for low, $t(80) = 2.33$, $p = .022$), and high and low intensity Fear (for high, $t(80) = 1.94$, $p = .056$; for low, $t(80) = 4.24$, $p < .001$)].

Results of response to Social scenarios (see Figure 32b top panel) showed a significant effect of response dimensions, $F(5,400) = 265.4$, $p < .001$, $\eta p^2 = .768$, intensity, $F(1,80) = 11.9$, $p < .001$, $\eta p^2 = .129$, and its interaction $F(5,400) = 54.2$, $p < .001$, $\eta p^2 = .404$. Follow-up paired t-tests found significantly different scores to high versus low intensity Happy expressions in all response dimensions except for low-intensity Fear scenario (all ts(80) > 4.77, all ps < .001; for low intensity Fear $t(80) = .727$, $p = .469$). In particular, response profile to high intensity stimuli showed higher scores to the target Happy scenarios, and lower scores to the remaining non-target dimensions.

For facial expressions of Fear, for both physical and social scenarios, the results showed a significant effect of response dimensions (for physical scenarios, $F(5,400) = 307.6$, $p < .001$, $\eta p^2 = .794$; for social scenarios, $F(5,400) = 174.3$, $p < .001$, $\eta p^2 = .685$), a significant effect of intensity (for physical scenarios, $F(1,80) = 18.3$, $p < .001$, $\eta p^2 = .0186$; for social scenarios, $F(1,80) = 16.0$, $p < .001$, $\eta p^2 = .167$), and a significant interaction between response dimensions and intensity (for physical scenarios, $F(5,400) = 15.1$, $p < .001$, $\eta p^2 =$

.158; for social scenarios, $F(5,400) = 36.1$, $p < .001$, $\eta p^2 = .311$). Follow-up t-tests revealed that for physical scenarios (see Figure 31b middle panel), stimuli depicting high intensity Fear generated significantly higher scores in high intensity Fear and Pain scenarios (all ts(80) ≥ 2.91, all ps $\leq .005$), and significantly lower scores to high intensity Happy scenario ($t(80) = 5.36$, $p < .001$). For social scenarios (see Figure 32b middle panel), stimuli depicting high intensity Fear generated significantly higher response to high intensity Pain scenario ($t(80) = 3.50$, $p < .001$), and significantly lower response to Happy scenarios (both ts(80) ≥ 2.12, both ps $\leq .030$).

Finally, for facial expressions of Pain, the results showed a significant effect of response dimensions (for Physical scenarios, $F(5,400) = 173.1$, $p < .001$, $\eta p^2 = .684$; for Social scenarios, $F(5,400) = 265.4$, $p < .001$, $\eta p^2 = .768$), a no significant effect of intensity (for Physical scenarios, $F(1,80) = .000$, $p = .985$, $\eta p^2 = .000$; for Social scenarios, $F(1,80) = 1.67$, $p = .200$, $\eta p^2 = .020$), and a significant interaction between response dimensions and intensity (for Physical scenarios, $F(5,400) = 11.0$, $p < .001$, $\eta p^2 = .120$; for Social scenarios, $F(5,400) = 3.04$, $p = .011$, $\eta p^2 = .037$). Follow-up t-tests revealed that, for physical scenarios (see Figure 31b bottom panel), stimuli depicting high intensity Pain generated significantly higher scores to high intensity Pain scenario ($t(80) = 5.55$, $p < .001$), and significantly lower scores in Fear scenarios (both ts ≥ 1.94, both ps $\leq .05$). For social scenarios (see Figure 32b bottom panel), stimuli depicting high intensity Pain generated significantly lower scores to low intensity Fear scenario ($t(80) = 4.52$, $p < .001$).

To sum up, while Context profiles exhibit remarkable similarity across Chinese and British participants, dynamic and static stimuli, and high and low intensity emotions, some subtle but significant differences do emerge. Notably, Chinese participants tend to assign higher scores to non-target dimensions compared to British participants, for both social and physical scenarios. Additionally, it seems that all participants tend to attribute higher scores to static than dynamic facial emotions, regardless of the specific dimensions involved. Finally, intensity of facial emotion only slightly modulates context profiles, with a tendency to perceive the target dimension as higher when the stimulus conveys high intensity emotions.

Context profiles to physical scenarios and dynamic facial emotions showed stronger correlation with perceptual similarity than those to social scenarios and static facial emotions

The above initial results revealed that context profiles for social scenarios were richer and more diversified compared to those generated for physical scenario, which were mostly characterized by prominent scores in the target contextual scenarios (e.g., Fear scenarios when judging facial expressions of Fear). Next, I tested whether context profiles are related to the perceived differences and similarities between facial expressions of emotions, and, if so, whether this correlation is different between social or physical context profiles.

I followed the same procedure as in Chapter 3. I first computed a profile similarity measure using the Cosine distance between response vectors for each pair of stimuli presented in the similarity rating task, for a total of 160 trials, for both social and physical scenarios. Then, I averaged values across each combination of facial emotions displayed by the stimuli (i.e., High- and Low-intensity facial expressions of Happy, Fear and Pain) and created a similarity matrix for social contexts and one for physical contexts. This procedure was repeated for each participant, resulting in a total of 81 matrices for social and 81 for physical scenarios. These were then averaged across cultures (i.e., British vs Chinese participants) and stimuli type (i.e., dynamic vs static stimuli) resulting in 8 matrices (i.e., 4 for social and 4 for physical scenarios), which represented the distance between context profiles of facial emotions, an indirect measure of perceptual similarity, see Figure 33 b and c.

Participants' responses to the similarity rating task were also averaged across each combination of facial emotion (i.e., High intensity and Low intensity facial expressions of Happy, Fear and Pain) to produce a perceptual similarity matrix for each participant. Again, these were then averaged for each culture (i.e., British vs Chinese participants) and each stimuli type (i.e., dynamic vs static stimuli), resulting in 4 matrices representing a direct measure of perceptual similarity (**Error! Reference source not found.**Figure 33a). To prevent diagonal-line values from inflating our results, I extracted and computed the Spearman rho coefficient using the upper triangle of the matrices. Since the direct perceptual similarity matrix lacks symmetry along the diagonal, the mirrored cells in the upper and lower triangles were averaged.

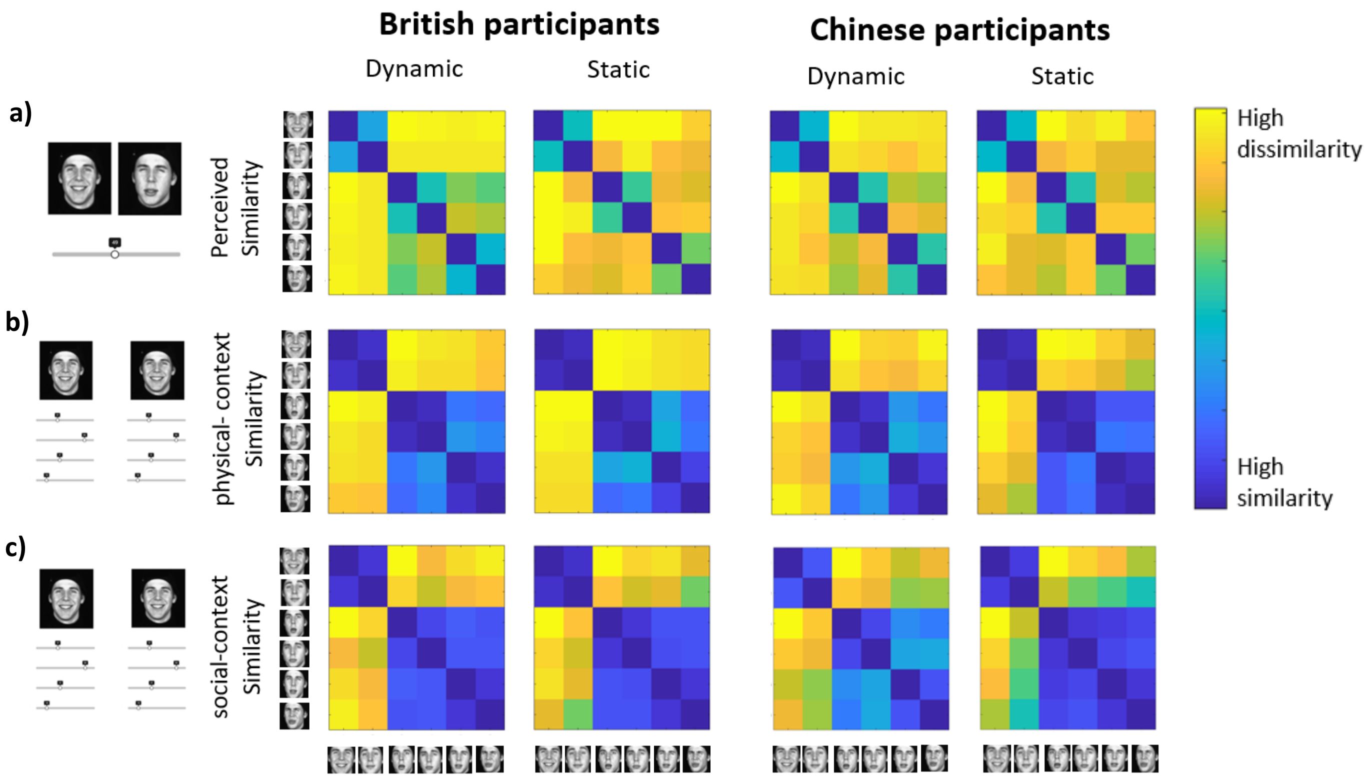

Figure 33. **Similarity matrices of Perceptual similarity and Context similarity for Social and Physical scenarios.** *a) Perceptual similarity directly obtained from participants' responses to the Similarity rating task. b) Profile similarity computed as the cosine distance between participants' vector responses to the Physical Context profiling task. c) Profile similarity computed as the cosine distance between participants' vector responses to the social context profiling task. Along the diagonal values obtained comparing profiles in response to the same facial emotions at the same intensity. Facial emotions from top to bottom (Y axis) and left to right (X axis) depicting high intensity Happy, low intensity Happy, high intensity Fear, low intensity Fear, high intensity Pain, low intensity Pain.*

As shown in see Figure 33, for British participants, results showed a strong correlation between perceptual similarity matrices and profile similarity matrices obtained with both social and physical scenarios, particularly for dynamic stimuli (for Physical scenarios r = .971, p < .001; for Social scenarios r = .932, p < .001) compared to static (for Physical scenarios r = .792, p < .001; for Social scenarios r = .758, p < .001). Similar results were also found for Chinese participants, with strong correlation between perceptual and physical-context profile similarity, particularly for dynamic stimuli (for Physical scenarios r = .945, p < .001; for Social scenarios r = .930, p < .001) compared to static (for Physical scenarios r = .756, p < .001; for Social scenarios r = .727, p < .001).

To assess the consistency of these results across individual participants, I analysed data at a participant-level by calculating the Spearman's rho coefficient between each participants' context profile matrices, for both social and physical scenarios, and their corresponding perceptual similarity matrix. The results are presented in Figure 34. For British participants who completed the Dynamic version of the study, all the profile matrices obtained with physical scenarios exhibited a significant correlation with their respective perceptual similarity matrices except for one participant (all rs ≥ .53, all ps ≤ .04; except for r = .33, p = .22). Context profile matrices obtained with social scenarios also significantly correlated with their respective perceptual similarity matrices except for two participants (all rs ≥ .56, all ps ≤ .03; except for two, both rs ≤ .24, both ps ≥ .38). For British participants who completed the Static version of the study, perceptual similarity matrices were significantly correlated with all the profile matrices obtained with physical scenarios except for three participants (all rs ≥ .49, all ps ≤ .05; except for three, all rs ≤ .50, all ps ≥ .06), and with all profile matrices obtained with social scenarios except for four participants (all rs ≥ .51, all ps ≤ .05; except for four all rs ≤ .47, all ps ≥ .07).

Similar results were found for Chinese participants. For those who completed the Dynamic task, all the physical context profile matrices exhibited a significant correlation with their respective perceptual similarity matrices except for one participant (all rs ≥ .61, all ps ≤ .01; except for r = .49, p = .06), and all the social context profile matrices significantly correlated with their respective perceptual similarity matrices except for three participants (all rs ≥ .51, all ps ≤ .05; except for three, all rs ≤ .40, all ps ≥ .13). For those who completed the Static task, all the context profile matrices obtained with physical scenarios exhibited a significant correlation with their respective perceptual similarity matrices except for five participants (all rs ≥ .54, all ps ≤ .03; except for five, all rs ≤ .49, all ps ≥ .06), and context matrices obtained with social scenarios significantly correlated with their respective perceptual similarity matrices except for seven participants (all rs ≥ .53, all ps ≤ .03; except for seven all rs ≤ .48, all ps ≥ .07).

Interestingly, as shown in Figure 34, correlation coefficients tended to be higher in the Dynamic compared to the Static conditions for both cultures, and context profile matrices regarding physical scenarios tended to have higher correlation coefficients compared to those generated with social scenarios. To investigate this observation further, I conducted a 2 (British/Chinese) by 2 (Static/Dynamic) by 2 (Physical/Social scenarios) mixed model ANOVA on the correlation coefficients obtained. The analysis revealed a significant effect of stimulus type (i.e., Dynamic vs. Static), $F(1, 77) = 13.18$, $p < .001$, a significant effect of Scenario, $F(1, 77) = 47.31$, $p < .001$, a non-significant effect of culture $F(1, 77) = 3.68$, $p = .06$, and no significant interactions among the three factors (all rs ≤ 3.12 , all ps ≥ .08). These results provide statistical support for the observations mentioned above.

In sum, while context profile similarity correlates with participants' perceptual similarity across cultural backgrounds, stimulus types, and types of contextual scenarios, a participant-level analysis revealed that this correlation is overall stronger in response to dynamic compared to static stimuli and is stronger for physical compared to social scenarios.

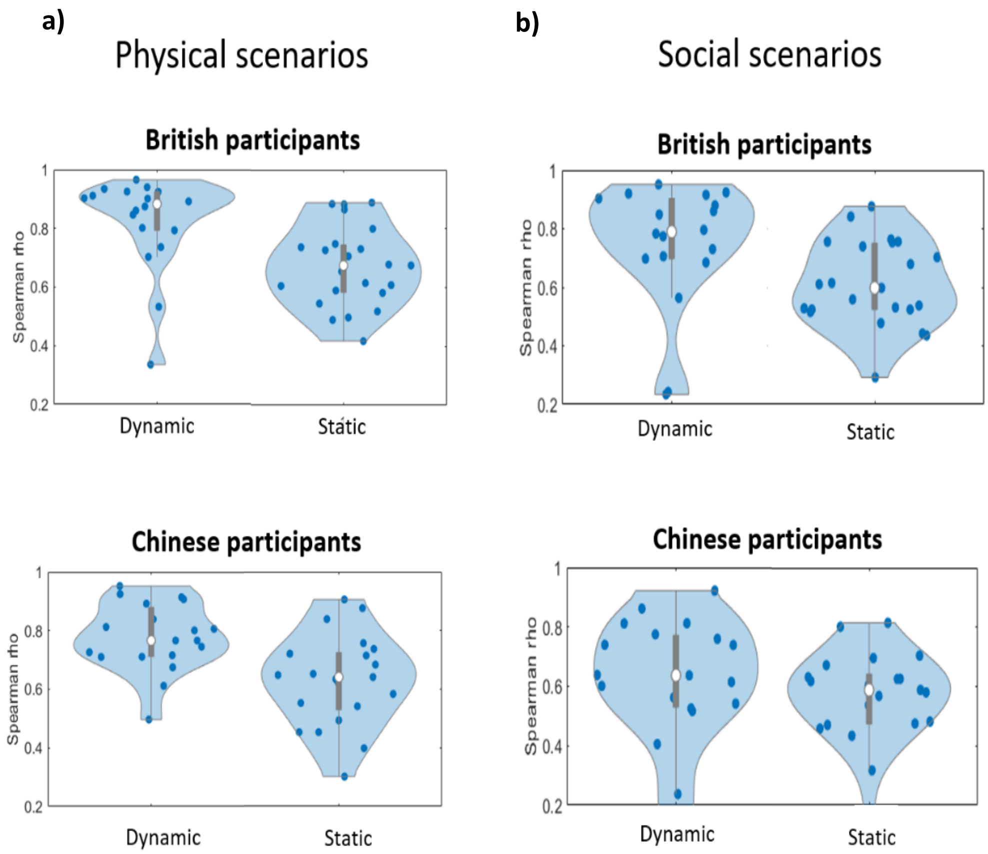

Figure 34. **Correlation coefficients between individual participants' perceptual similarity and their corresponding context profile similarity obtained with responses to physical and social scenarios.** *a) Results for Physical context scenarios for British and Chinese participants b) Results for Social context scenarios in British and Chinese participants.*

4.3 Study 6. How does contextual information affect the processing of facial expressions of emotions?

Study 5 has shown that facial expressions of emotions are associated with a wide array of social and physical contexts that are both congruent and incongruent to the facial emotion being expressed. Notably, context profiles elicited by physical stimulation exhibit a stronger association with perceived similarity between facial emotions compared to context profiles elicited by social stimulation. Moreover, context profiles and their associations with perceptual similarity between facial emotions can be influenced by facial motion, emotion intensity and perceivers' culture background. While these findings provide compelling evidence of the strong link between facial expressions of emotions and contextual information, they do not delve into how the presence of physical and social context information might impact our perception of facial emotions, especially in terms of their rich meanings as measured by emotion profiles. This question is addressed in Study 6.

In Study 6, I investigated whether presenting facial emotions with congruent or incongruent social and physical scenarios influences how we construct profiles of the emotion content perceived in spontaneous facial expressions of emotions. Additionally, consistent with previous studies reported in this thesis, I also examined the role played by facial motion, emotion intensity and culture in this potential modulation process. To do so, using the same stimuli of Study 5, participants were asked to perform an emotion profiling task (i.e., same as in Study 3) while simultaneously being presented with a facial expression of emotion, and one of the six possible social and physical context scenarios (i.e., the same as used in Study 5). Given that the existing literature has often revealed the impact of situational context on how we perceive emotion categories from stereotypical facial expressions (Carroll & Russell, 1996; Kim et al., 2004; Mobbs et al., 2006; Righart & de Gelder, 2008), I expected to see similar context influences also when facial expressions are perceived as rich and multidimension emotions profiles. By combining data from Study 3, where participants performed the emotion profiling task without any context, the results of the present study also enabled me to investigate whether emotion perception is *enhanced* when presented with congruent social/physical contexts, and whether it is *attenuated* when presented with incongruent contexts.

4.3.1 Methods

Participants

The size of the participants' sample was based on the results from our previous studies, suggesting that a minimum of 13 participants was required. Again, considering the fact that the study was conducted online we collected 20 participants for condition (dynamic/static).

Forty-three British participants were recruited from the University of East Anglia via the SONA System. Participants who did not identify themselves as British in the demographic questionnaire were excluded from the study. However, seven of them did not complete the second part of the experiment, resulting in a final sample of thirty-six participants (5 males, 30 females, 1 other; age ranged between 18-31 yrs., M = 20.3, SD = 2.69). Forty-four Chinese participants were recruited from the Sun Yat-sen University, China (22 males, 22 females); age ranged between 18-27 yrs., M = 20.6, SD = 1.69). All participants were naïve to the purpose of the investigation and had not taken part in any previous studies of this project. All participants provided informed consent before taking part in the study and were debriefed at the end, receiving course credits as compensation.

Stimuli, Materials, and Tasks

The stimuli were the same used in Study 5, including images or videos of 9 actors displaying 3 different emotions (Happy, Pain, Fear) at 2 intensities (High, Low) taken from the large MPI Facial Expression Database (Kaulard et al., 2012). The Emotion profiling in Context Task was similar to the Emotion profiling task employed in Study 3. However, in this case, stimuli were displayed together with a specific context. The context was presented as a text-based scenario intended to induce a target emotion of high intensity Happy, Fear or Pain based on physical stimulation or a socially meaningful event. The study consisted of two sessions, one with Social contexts and one with Physical contexts, with both sessions completed by participants about 12 hours apart.

social and physical scenarios were the same as in Study 5 but were presented in a third person perspective. The first session used physical scenarios as context, e.g. “The person is eating their favorite food in their favorite place”, while the second session used social scenarios, e.g. “The person won the first prize in a big competition”. Only the scenarios used to evoke high intensity emotions were selected, for a total of 6 scenarios (i.e., 2 context type: social vs physical * 3 emotions: Happy vs Fear vs Pain, see Appendix C, Figure C.1 for a complete list). Depending on the study version, dynamic or static, stimuli were either videos or images. In both sessions, each of the 54 stimuli (i.e., 9 actors * 3 emotions * 2 intensity) was combined in random order with all the three context scenarios, for a total of 162 trials per session. In both sessions, each trial started with a fixation cross (1000ms) followed by a screen displaying the stimulus on the left side of the screen (i.e., image for the Static version, or video for the Dynamic version of the Task), and 8 different sliders on the right side of the screen, a sentence above the stimulus described the scenario that elicits the facial expression displayed by the stimulus, (see Figure 35). Participants were asked to rate the facial expressions along multiple dimensions considering the scenario where the expression was shown. They needed to indicate how much the facial expression was displaying each of 8 possible emotions (Happy, Surprise, Sad, Disgust, Neutral, Anger, Fear, Pain) by moving the handles of the 8 sliders, that went from 0 to 100. Each slider was independent from the others, and the handles of all sliders were initially placed at 0. To ensure participants had sufficient time to provide their answers without overthinking, preventing the interference of more high-level processes, the response screen had a timelimit of 30 seconds. Dynamic stimuli were presented on a loop until a decision was made or the time limit was reached. If the time limit was reached the experiment moved them to the next trial. Participants completed both sessions of the study, however the link to the second session of the study became available only 12 hours after the completion of the first part.

The same translation process employed in the previous study was followed, aiming to ensure accurate translation and maintain consistency across languages. Also in this case, it is crucial to recognize the potential for semantic differences between languages, which may result in subtle variations in interpretation.

Finally, the same measures taken in previous studies were implemented with the aim to reduce, and account for, potential participants' distractions or disengagement. In particular, (1) access to the study was restricted to PCs or laptops; (3) participants were required to self-report the reliability/usefulness of their data (e.g., due to lacked attention) at the end of the experiment, (4) a time-limit was imposed on each screen, and participants exceeding the maximum allotted time for completing the experiment were excluded from the study; (5) control trials were included in the study, where participants were asked to rate the similarity of identical images.

Procedure

Participants performed the Emotion profiling in Context task online through the Gorilla platform. They gained access through a URL link using their desktop computer or laptop (i.e., no tablet or phone access were allowed). The study followed the same general procedure for all participants. Once consent was given, participants were directed to a demographic questionnaire. Closed-ended questions asked for their hand dominance (right-handed/left-handed) and the gender they identify with (male/female/other); while open-ended questions asked for their age and nationality. Half of the participants were then randomly allocated by the system to the Dynamic version of the study, with movies as stimuli, while the other half to the Static version, with images as stimuli. The assigned allocation was kept for both sessions of the study. During the first session of the study, the Emotion profiling in Physical Context task, participants rated each stimulus while considering the physical scenario where the facial expression occurred. They had a break halfway through the task of no more than 5 minutes. After about 12 hours from the first session, participants completed the second session of the study, the Emotion profiling in Social Context task (see Figure 35). In this second session they rated each stimulus considering the social scenario where the facial expression occurred. Each session took about 40 minutes to complete. Detailed instructions and three practice trials were given before starting each session. At the end of the study, participants filled out a self-report questionnaire regarding the reliability of their data and were debriefed.

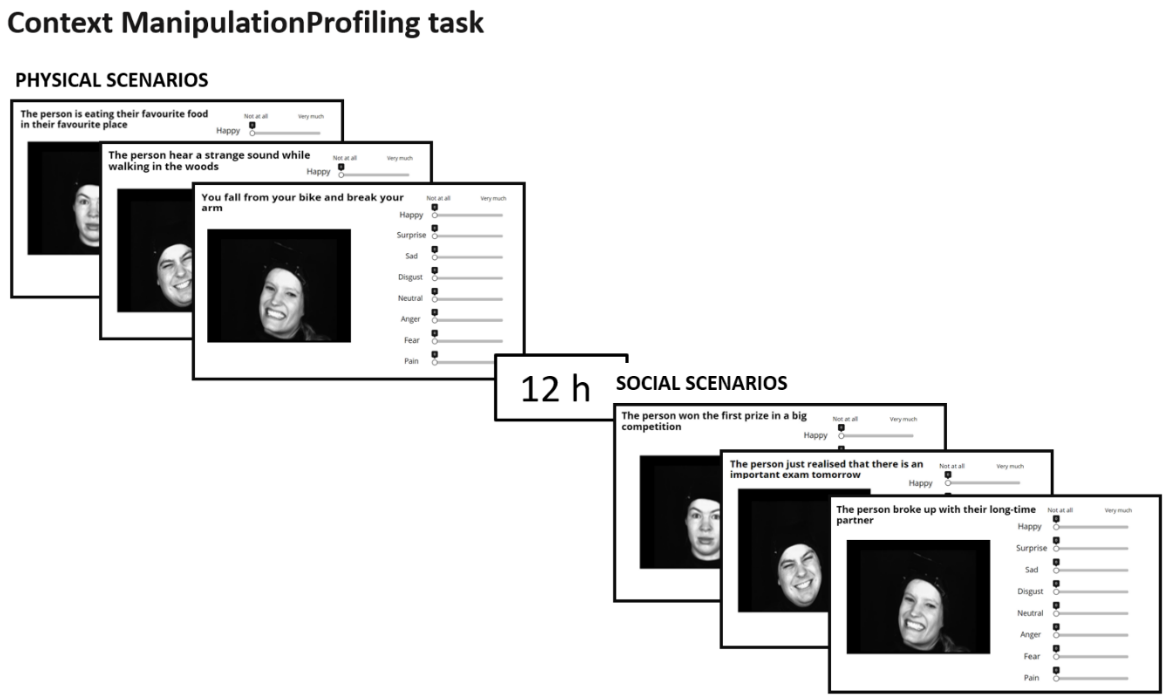

Figure 35. **Emotion profiling in Context task**. *The task was composed by two sessions about 12 hours apart. In both session participants were asked to rate the emotion profile of a facial expression displayed in a specific context. Each of the 54 stimuli was combined in a random order with 3 different scenarios evoking high intensity emotions of Happy, Fear or Pain, for a total of 162 trials per session. The first session used physical scenarios, that evoke the target emotion due to physical stimulation (e.g., The person is eating their favorite food in their favorite place), while the second session used social scenarios, that evoke the wanted emotion due to a socially meaningful event (e.g., The person won the first prize of a big competition).*

4.3.2 Results and Discussion

By applying the same exclusion criteria as used in previous studies, two British participants were excluded from the following data analysis, resulting in a final sample size of 34 British participants and 40 Chinese participants for analysis.

Congruent contexts enhance perception of the target dimension, more for physical than social scenarios

To investigate the impact of congruent contexts on the perceived emotion profiles, I compared participants' emotion profiling responses with emotionally consistent contexts (Study 6) to that generated in a baseline condition without any context (i.e., Study 3). The results are shown in Figure 36. First, emotion profiles in the baseline condition were overall very similar to those obtained with social and physical contexts. Second, congruent contextual information tended to enhance participants' response to the target emotion dimensions for physical scenarios but not for social scenarios. Third, responses to non-target emotion dimensions were generally higher with a congruent context compared to no-context, for both physical and social scenarios.

In particular, for facial expressions of Fear with/out physical contextual information (see Figure 36, top panel on the left), a 2 (presence/absence of physical contextual information) by 8 (emotion dimensions) mixed model ANOVA showed a significant effect of response dimensions, $F(7,1071) = 375$, $p < .001$, $\eta p^2 = .710$, indicating that participants provided diversified scores to the eight emotion dimensions, a significant main effect context presence, $F(7,153) = 4.36$, $p = .038$ $\eta p^2 = .028$, and a significant interaction between presence/absence of contextual information and response dimensions, $F(7,1071) = 8.69$, $p < .001$, $\eta p^2 = .054$. Similarly, analyzing the same response profiles but with/out social contextual information (see Figure 36, top panel on the right), the results were the same, with a significant effect of response dimensions, $F(7,1071) = 4.36$, $p = .038$, $\eta p^2 = .028$, a significant effect of context presence, $F(7,153) = 4.36$, $p = .038$, $\eta p^2 = .028$, and a significant interaction between the two factors, $F(7,1071) = 4.49$, $p < .001$, $\eta p^2 = .029$. To further explore these interactions, I performed independent t-tests for each emotion dimension. While the emotion profile generated with congruent physical contexts has a significantly higher score for the target emotion of Fear (36% (SD = 18.9), baseline, 24% (SD = 13.80); $t(153) = 4.87$, $p < .001$), it showed similar scores for all other dimensions observed in the baseline condition (all ts ≤ 1.85, all ps $\geq .067$). For social scenarios, when compared to the baseline condition, emotion profiles showed significantly higher score to the dimensions of Anger ($t(153) = 1.94$, $p = .053$), Pain ($t(153) = 2.00$, $p = .047$), Sad ($t(153) = 3.76$, $p < .001$) and Surprise ($t(153) = 2.97$, $p = .003$), but not to the target dimension of Fear ($t(153) = 1.61$, $p = .247$).

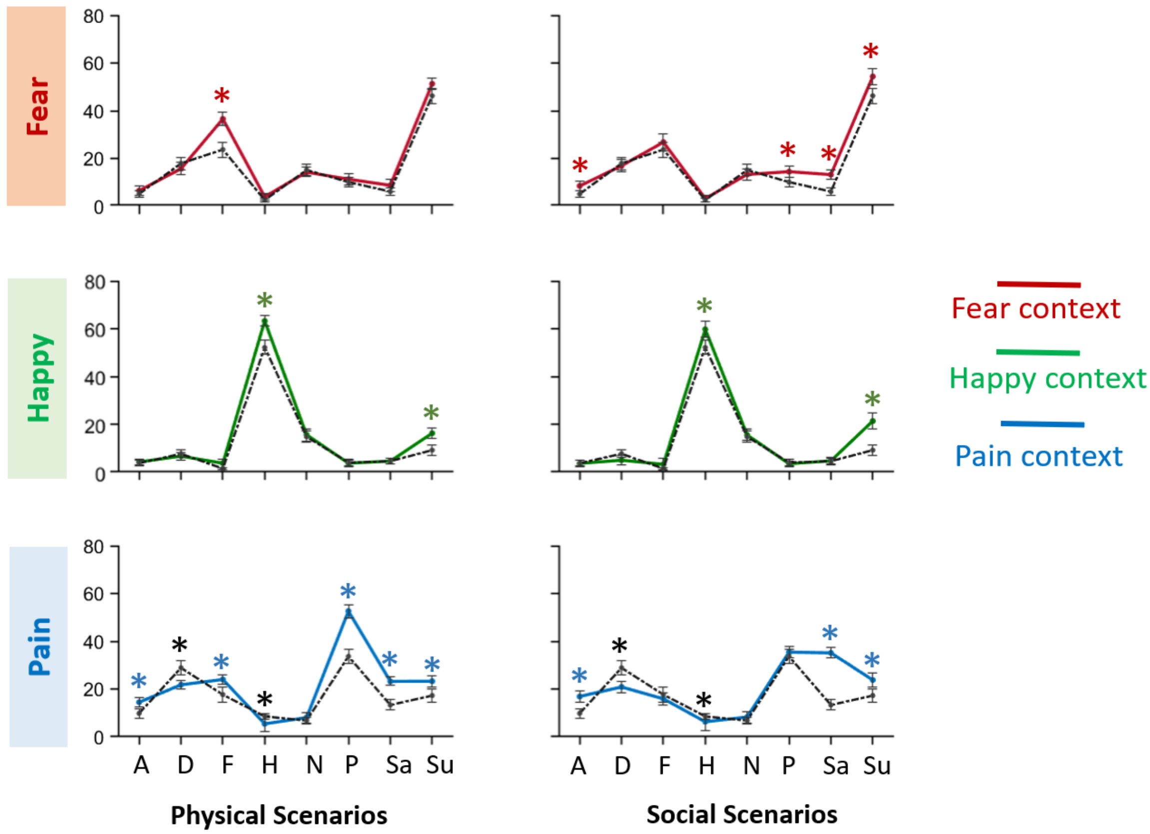

Figure 36. **Emotion profiles obtained with and without congruent social and physical contexts.** *Emotion profiles generated in response to facial expressions of Fear (top panels) Happy (middle panels) and Pain (bottom panels) within Physical or Social scenarios evoking congruent emotions of Fear (in red), Happy (in green) and Pain (in blue). Dashed line represents emotion profiles obtained without any contextual information, from Study 3. Asterisks indicate significant differences.*

For facial expressions of Happy with/out physical contextual information (see Figure 36, middle on the left), the same ANOVA showed a significant effect of response dimensions, $F(7,1071) = 375.15$, $p < .001$, $\eta p^2 = .710$, a significant effect of context presence, $F(7,153) = 4.36$, $p = .038$ $\eta p^2 = .028$, and a significant interaction between presence/absence of contextual information and response dimensions, $F(7, 1071) = 8.69$, $p < .001$, $\eta p^2 = .054$. The same results were found for social context condition (see Figure 36, middle panels on the right), showing a significant effect of response dimensions, $F(7,1071) = 562.5$, $p < .001$, $\eta p^2 = .786$, a significant effect of context presence, $F(7,153) = 6.47$, $p = .012$, $\eta p^2 = .041$, and a significant interaction between the two factors, $F(7,1071) = 10.6$, $p < .001$, $\eta p^2 = .065$. Comparisons showed that congruent context significantly increased the scores to the dimensions of Happy and Surprise, for both physical (for Happy 63%, SD = 16.4, $t(153) = 4.51$, $p = < .001$, for Surprise, 16%, SD = 16.1, $t(153) = 3.29$, $p = .001$) and social contextual information (for Happy, 59%, SD = 18, $t(153) = 2.86$, $p = .005$, for Surprise, 21%, SD = 18.2, $t(153) = 5.23$, $p < .001$). Responses to other emotion dimensions did not significantly differ with/out congruent contexts (for Physical scenarios, all ts(153) $\leq .657$, all ps $\geq .512$, for Social scenarios, all ts(153) ≤ 1.85, all ps $\geq .066$).

Finally, for facial expressions of Pain with/out physical contextual information (see Figure 36, bottom panels on the left), the same ANOVA showed a significant effect of response dimensions, $F(7,1071) = 137$, $p < .001$, $\eta p^2 = .472$, a significant main effect of context presence, $F(7,153) = 9.22$, $p = .003$, $\eta p^2 = .057$, and a significant interaction between presence/absence of contextual information and response dimensions, $F(7,1071) = 20.6$, $p < .001$, $\eta p^2 = .011$. Similar results were found for social contextual information, with a significant effect of response dimensions, $F(7,1071) = 110.6$, $p < .001$, $\eta p^2 = .420$, a significant difference in overall responses with or without congruent contextual information, $F(7,153) = 9.22$, $p = .030$, $\eta p^2 = .030$, and a significant interaction between the two factors, $F(7,1071) = 25.8$, $p < .001$, $\eta p^2 =.145$.

Independent t-test revealed that emotions profiles generated with physical contexts had significantly different scores compared to baseline at all response dimensions except for Neutral (all ts(153) ≥ 2.31, all ps $\leq .022$, except Neutral, $t(153) = .920$, $p = .359$). In particular, responses are higher compared to baseline for the target emotion Pain, with 52% (SD = 19.9) compared to the 34% (SD = 13.8), Anger, with 14% (SD = 16.3) compared to 10% (SD = 8.46), Fear, with 24% (SD = 19.6) compared to 18% (SD = 12.6), Sad, with 23% (SD = 18.6) compared to 13% (SD = 10.5), and Surprise, with 23% (SD = 18.5) compared to 18% (SD = 12.01). However, the scores were lower than baseline condition for Disgust, with 22% (SD = 17.5) compared 29% (SD = 13.0) and Happy, with 5% (SD = 7.86) compared to 8% (SD = 5.88). For social scenarios, comparisons showed significantly higher scores than baseline condition for Anger, with 17% (SD = 16.9) compared to 10% (SD = 8.46), Sad, with 35% (SD = 19.3) compared to 13% (SD = 10.5), and Surprise, with 24% (SD = 18.5) compared to 17% (SD = 12.01), all ts(153) ≥ 2.48, all ps ≤ .014. Significantly lower scores were also detected for Disgust, with 21% (SD = 16.6) compared to 29% (SD = 13.0), and Happy, with 6% (SD = 6.54) compared to 8% (SD = 5.88), both ts(153) ≥ 2.21, both ps ≤ .028, while no significant difference was found for the remaining dimensions including the target dimension Pain (all ts(153) ≤ 1.06, all ps ≥ .291).

Context information affects emotion profiles by reducing perception of the target emotion and enhancing perception of the emotion evoked by the context

To investigate how congruent and incongruent contextual information affect the perception of facial emotions, I compared participants' emotion profiles generated with congruent contexts to emotion profiles generated with incongruent contexts. As shown in Figure 37, the presence of context information strongly influences emotion profiles in most of its response dimensions. In particular, while scores to the target emotion (i.e., displayed by the stimulus) were decreased with incongruent contexts, scores were increased to the emotion dimensions that are in line with the context.

For facial expressions of Fear, the effect of physical contexts was tested with a 3 (physical contexts) by 8 (response dimensions) repeated measure ANOVA. The results showed a significant effect of response dimensions, $F(7,539) = 167$, $p < .001$, $\eta p^2 = .684$, a significant difference in overall responses across the three different contexts, $F(2,154) = 21.7$,

$p < .001$, $\eta p^2 = .220$, and a significant interaction between context and response dimensions, $F(14,1078) = 47$, $p < .001$, $\eta p^2 = .379$. Paired sample t-tests indicated that scores to the target response dimension Fear were significantly lower with a Happy context (14%, SD = 13.9; $t(77) = 11.13$, $p < .001$) or a Pain context (25%, SD = 17.7, $t(77) = 6.68$, $p < .001$) when compared to a Fear context (37%, SD = 18.9). In addition, scores attribute to the response dimensions congruent with the emotion evoked by the context (e.g., response to Happy dimension with a Happy context), were significantly higher compared to scores generated with context of Fear, for context of Happy (15%, SD =16.1, compared to 4%, SD = 7.50, $t(77) = 7.33$, $p < .001$) and context of Pain (26%, SD = 17.8, compared to 11%, SD = 7.50 , $t(77) = 4.18$, $p < .001$).

Similarly, for response profiles associated with social scenarios, the same ANOVA showed a significant effect of response dimensions, $F(7,539) = 186$, $p < .001$, $\eta p^2 = .707$, a significant difference across the different contexts, $F(2,154) = 21.0$, $p < .001$, $\eta p^2 = .215$, and a significant interaction between context and response dimensions, $F(14,1078) = 44.9$, $p < .001$, $\eta p^2 = .368$. Again, paired sample t-tests indicated that scores to the target Fear response dimension were significantly lower with both a Happy context (10%, SD = 13.1) , and a Pain context (17%, SD = 18.3), compared to that observed with a Fear context (27%, SD = 20.2, compared to Happy, $t(77) = 8.26$, $p < .001$, compared to Pain $t(77) = 7.22$, $p < .001$). Scores attributed to the response dimensions that are consistent with the context (e.g., Happy dimension with a Happy context), were significantly higher compared to the scores generated with a Fear context for Happy (17%, SD = 16.1, compared to 3%, SD = 7.50, $t(77) = 7.34$, $p < .001$) and Pain (17%, SD = 17.8, compared to 14%, SD = 13.50, $t(77) = 1.67$, $p = .005$).

In line with previous results, for response profiles to stimuli depicting facial expressions of Happy and Pain associated with social and physical scenarios, the ANOVA analysis showed a significant effect of emotions response dimensions, all Fs(7,539) ≥ 44.7, all ps $< .001$, a significant difference in overall responses (across emotion dimensions) among profiles generated across the different contexts, all Fs(2,154) ≥ 4.11, all ps $< .001$, and a significant interaction between context and response dimensions, Fs (14,1078) $\geq$ 47.82, all ps $< .001$. Paired t-tests indicated that, in response to facial expressions of Happy, scores to the target dimension Happy were significantly lower with a Fear and a Pain context for both physical (for Fear, 35%, SD = 18.9, for Pain 28%, SD = 20) and social scenarios (for Fear, 31%, SD = 18.9, for Pain 37%, SD = 19) when compared to that with a Happy context (for physical 63%, SD = 16.4, for social 60%, SD = 18.5; all ts(77) $\geq$ 8.53, all ps $\leq$ .001). Again, scores attribute to the emotional dimensions consistent with the context (e.g., Fear dimension with a Fear context) were significantly higher than that generated with a Happy context, both for physical and social contexts of Fear (for physical, 13%, SD = 15.2, compared to 4%, SD = 8.12 , $t(77) = 5.71$, $p < .001$; for social, 11%, SD =15.2 , compared to 3%, SD = 8.43, $t(77) = 5.62$, $p < .001$), and physical and social contexts of Pain (for physical, 20%, SD = 16.5 , compared to 4%, SD = 7.57, $t(77) = 8.76$, $p < .001$; for social, 12%, SD = 15, compared to 3%, SD = 8.67, $t(77) = 5.90$, $p < .001$).

For facial expressions of Pain, scores to the target dimension of Pain were significantly lower with a Fear or Happy context, for both physical (for Fear, 26%, SD =17.8, for Happy 19%, SD =15.7) and social scenarios (for Fear, 28%, SD = 20.9, for Happy 15%, SD =17.2), compared to that with a Pain context (for physical, 53%, SD =19.9, for social, 35%, SD = 20.6, ts(77) $\geq$ 4.58, all ps <.001). Also, scores attribute to the response dimensions consistent with the context (e.g., Fear within a Fear context), were significantly higher compared to corresponding scores generated within context of Pain, both for physical and social contexts of Fear (for physical, 37%, SD = 17.5, compared to 24%, SD = 19.6, $t(77) = 6.14$, $p < .001$; for social, 25%, SD = 17.9, compared to 16%, SD = 18.5, $t(77) = 6.54$, $p < .001$), and physical and social contexts of Happy (for physical, 20%, SD = 16.7 , compared to 5%, SD = 7.86, $t(77) =8.09$, $p < .001$; for social, 25%, SD = 17.5, compared to 6%, SD = 6.54, $t(77) = 9.58$, $p < .001$).

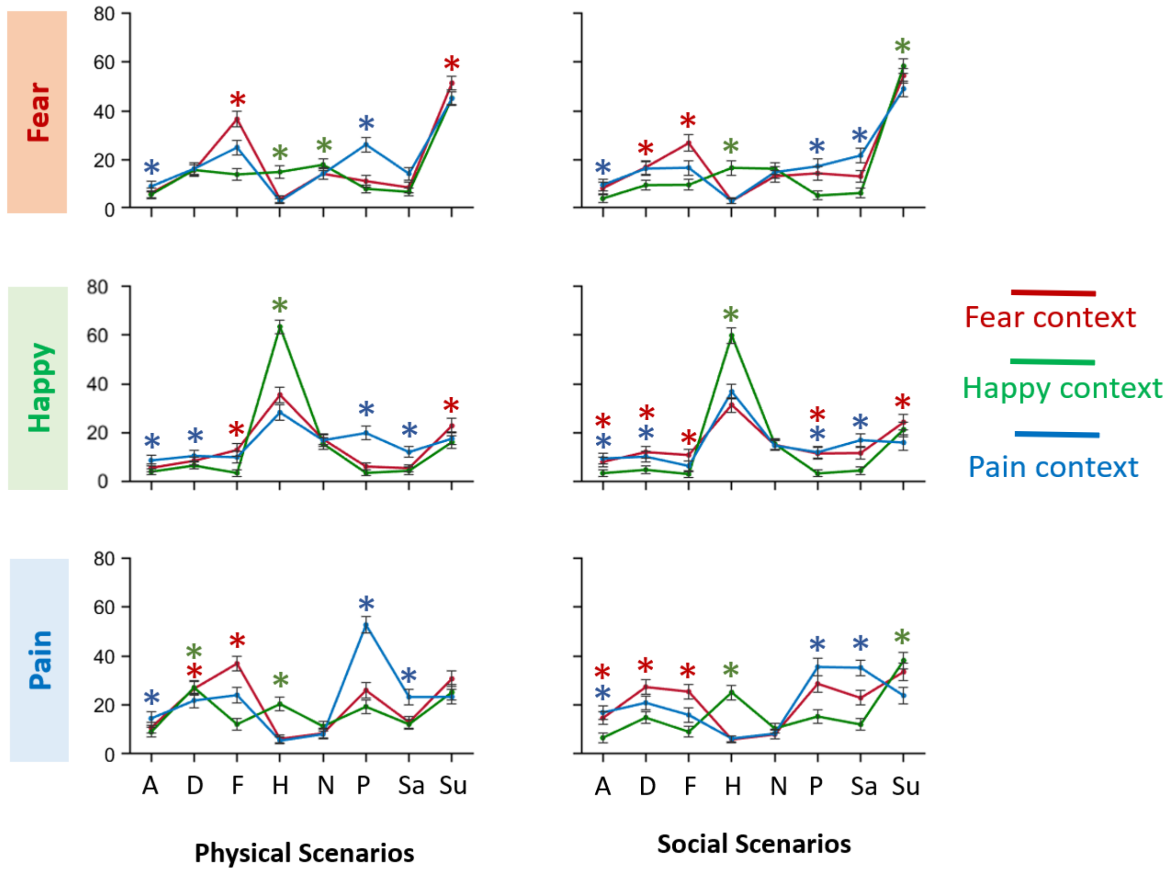

Figure 37. **Emotion profiles of Fear, Happy and Pain observed with different emotional contexts.** *Rows represent different facial emotions, and columns represent Physical and Social contexts respectively. Asterisks indicate significant differences between responses in congruent and incongruent contexts.*

Emotion Profiles in contexts are affected by Emotion Intensity and Culture

Next, I tested whether emotion profiles observed in different contexts were affected by facial motion (i.e., Dynamic vs Static stimuli), emotion intensity (i.e., High vs Low) and participants' cultural background (i.e., Chinese vs British participants). To do so, I obtained and analysed contrasting emotion profiles in response to the three facial emotions separately for physical and social contexts. I then compared them using a 2 (British vs Chinese / Dynamic vs Static stimuli / High vs Low emotion intensity) by 8 (response dimensions) ANOVA.

***British vs Chinese participants' responses: Physical Contexts*.** For Happy expressions (see Figure 38, left column), the main effect of response dimensions was significant with all three contexts of Happy, Fear and Pain (all Fs(7,532) ≥ 21.57, all ps < .001), indicating that participants provided significantly different scores to the eight emotion dimensions. There was also a significant difference in overall responses between British and Chinese participants with the Fear and Pain contexts (both Fs(1, 76) ≥ 4.81, both ps ≤ .031) but not with a Happy context (Fs(1, 76) = 3.4, p = .085). The interaction between culture and response dimensions was significant with both Happy and Pain contexts (both Fs(7,532) ≥ 2.92, both ps ≤ .005) but not with a Fear context (Fs(7,532) =1.62, p =.126). Independent t-tests showed significantly higher scores for British compared to Chinese participants to the target dimension of Happy when the stimulus is accompanied by a congruent context (for Happy context, t(76) = 2.52, p = .014; for Pain and Fear contexts, both ts(76) ≤ .781, ps ≥ .44), and higher scores for Chinese participants to some non-target dimensions across all contexts, including dimensions of Disgust, Pain, Sad and Surprise across Happy and Fear contexts (all ts(76) ≥ 2.15, all ps ≤ .035), and dimensions of Disgust, Fear and Sad with a context of Pain (all ts(76) ≥ 2.42, all ps ≤ .018).

For facial expression of Fear (see Figure 38, middle column), there was a significant effect of response dimensions for all three contexts (all Fs(7,532) ≥ 101.05, all ps < .001), a significant difference in overall responses between British and Chinese participants in the context of Fear (F(1,76) = 4.33, p = 01) but not in contexts of Happy and Pain (both Fs(1,76) ≤ 2.06, both ps ≥ .155), and a significant interaction between culture and response dimensions for all contexts (all Fs(7,532) ≥ 2.24, all ps ≤ .030). Follow-up independent t-tests showed significantly higher scores for British compared to Chinese participants to the target dimension of Fear when the stimulus is accompanied by a congruent context (for Fear context, $t(76) = 2.48$, $p = .016$; for Pain and Happy contexts, both ts(76) ≤ 1.23, ps $\geq .223$), and higher scores for Chinese participants to some non-target dimensions across contexts, including dimensions of Neutral and Pain (both ts(76) ≥ 1.97, both ps $\leq .050$) with a Happy context, all dimensions except Disgust with a Fear context (all ts(76) ≥ 1.96, all ps $\leq .050$), and the dimensions of Neutral and Sad with a Pain context (both ts(76) ≥ 2.06, both ps $\leq .042$).

Finally, for facial expression of Pain (see Figure 38, right column), it was found a significant effect of response dimensions in all contexts (all Fs(7,532) ≥ 22.23, all ps $< .001$), a significant difference British and Chinese participants' responses in the contexts of Pain and Fear (both Fs(7,532) ≥ 4.17, both ps $\leq .044$) but not in context of Happy ($F(7,532) = 3.643$, $p = .060$), and a significant interaction between culture and response dimensions for Pain and Fear contexts (both Fs(7,532) ≥ 3.56 , both ps $< .030$) but not for Happy context ($F(7,532) = 1.40$ p =201). Follow-up independent t-tests showed significantly higher scores for Chinese participants to some non-target dimensions, including the dimension of Neutral for the context of Happy ($t(76) = 2.50$, $p = .014$), dimensions of Disgust, Happy, Neutral, Pain and Surprise for context of Fear (all ts(76) ≥ 2.18, all ps $\leq .032$), and the dimensions of Fear, Neutral and Sad for contexts of Pain (all ts(76) ≥ 2.56, all ps $\leq .012$).

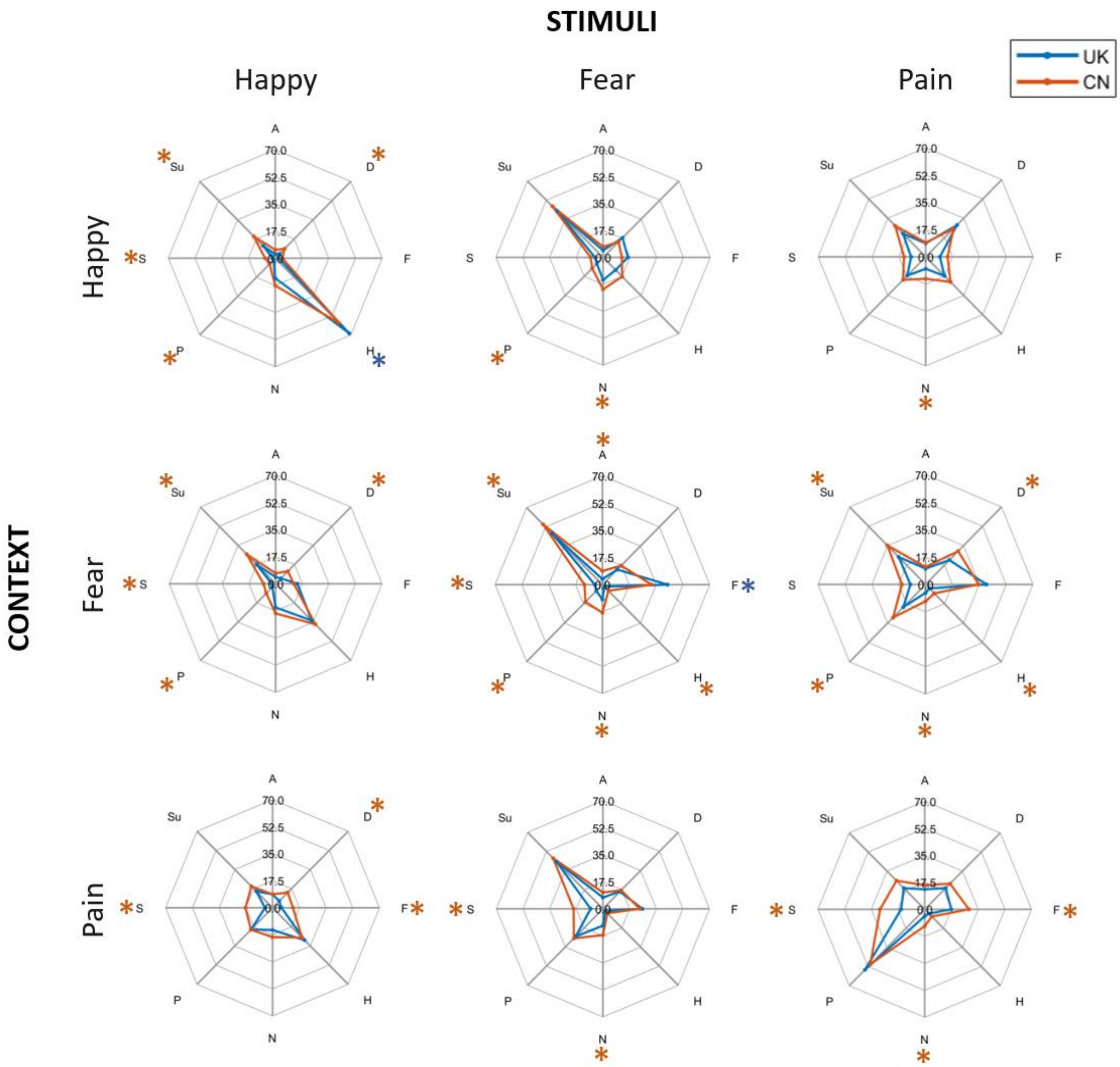

Figure 38. **Contrasts results of Emotion Profiles observed with different Physical contexts for British and Chinese participants.** *Responses of British (blue) and Chinese (orange) participants to facial expressions of Happy, Fear and Pain, columns, in Physical contexts intended to evoke emotions of Happy, Fear and Pain, rows. Asterisks indicate significant differences.*

***British vs Chinese participants' responses: Social Contexts*.** Similar results were found when contrasting emotion profiles associated with different social contexts. For Happy expression (see Figure 39, left column), it was found a significant effect of response dimensions in all contexts (all Fs(7,532) ≥ 30.62, all ps < .001), a significant difference between British and Chinese participants' responses in a Pain context (F(1,76) =1.35, p = .025) but not in contexts of Happy and Fear (both Fs(1, 76) ≤ 2.11, both ps ≥ .151), and a significant interaction between culture and response dimensions for the three contexts (all Fs(7,532) ≥ 2.01, all ps ≤ .052). Follow-up independent t-tests showed significantly higher scores for Chinese compared to British participants to the dimensions of Neutral and Sad for Happy and Fear contexts (both ts(76) ≥ 2.23, both ps ≤ .028) and to the dimensions of Disgust, Neutral and Sad for the context of Pain (ts(76) ≥ 1.96, all ps ≤ .050).

For Fear expression (see Figure 39, middle column), it was found a significant effect of response dimensions in all contexts (all Fs(7,532) ≥ 104.32, all ps < .001), a non-significant difference between British and Chinese participants' responses in the three contexts (F(1,76) ≤ 1.21, p ≥ .113), and a significant interaction between culture and response dimensions for Happy and Fear contexts (both Fs(7,532) ≥ 3.43, both ps ≤ .001), but not for Pain context (F(7,532) = 1.67, p = .113). Follow-up independent t-tests showed significantly higher scores for British compared to Chinese participants to the target dimension of Fear when the stimulus is accompanied by a congruent context (for Fear context, t(76) = 3.04 , p = .003; for Pain and Happy contexts, both ts(76) ≤ .861, ps ≥ .392), and higher scores for Chinese participants to some non-target dimensions across contexts, including dimensions of Neutral and Sad for context of Happy (both ts(76) ≥ 2.16, both ps ≤.034), dimensions of Neutral, Pain and Sad for the context of Fear (all ts(76) ≥ 2.64, all ps ≤ .008), and dimensions of Neutral for context of Pain (t(76) = 2.93, p = .004).

Finally, for Pain expressions (see Figure 39, right column), it was found a significant effect of response dimensions in all contexts (all Fs(7,532) ≥ 46.75, all ps < .001), a non-significant difference between British and Chinese participants' responses (all Fs(1,76) ≤ 2.56, all ps ≥ .114), and a significant interaction between culture and response dimensions for Fear context ($F(7,532) = 4.31$, $p < .001$) but not for Happy and Pain contexts (both Fs(7,532) < 2.16, both ps > 063). In particular, independent t-tests showed significantly higher scores for Chinese participants to some non-target dimensions across contexts, including the dimension of Neutral, Sad and Happy for the context of Happy (all ts(76) ≥ 2.03, all ps ≤ .046), dimensions of Disgust, Neutral, and Pain for the context of Fear (all ts(76) ≥ 1.97, all ps ≤ .050, and dimensions of Neutral for the context of Pain ($t(76) = 3.30$, $p = .001$).

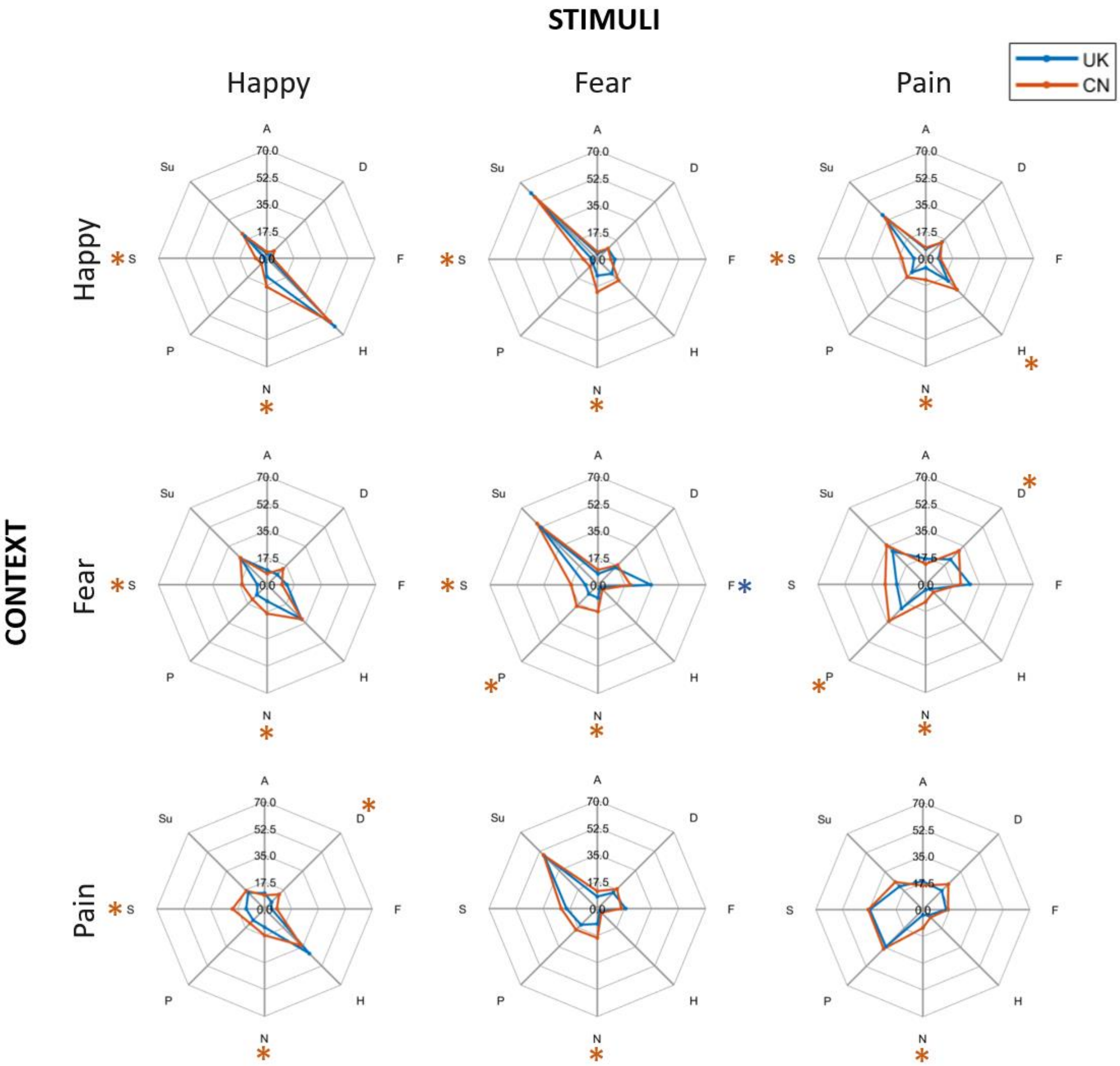

Figure 39. **Contrasts results of Emotion Profiles observed with different Social contexts for British and Chinese participants.** *Responses of British (blue) and Chinese (orange) participants to facial expressions of Happy, Fear and Pain, columns, in Social contexts intended to evoke emotions of Happy, Fear and Pain, rows. Asterisks indicate significant differences.*

***Static vs Dynamic Facial Expressions: Physical Contexts*.** The same analysis was applied to examine the influence of facial motion on emotion profiles observed with different emotional contexts. Overall, facial motion seemed to only slightly modulate emotion profiles with contextual information. For Happy expressions (see Figure 40, left column), there was a significant effect of response dimensions in all contexts (all Fs(7,532) ≥ 19.98, all ps < .001), a non-significant difference between responses to Dynamic and Static emotions for all contexts (all Fs(1, 76) ≥ .799, all ps ≤ .374), and a non-significant interaction between facial motion and response dimensions for all contexts (all Fs(7,532) ≥ .777, all ps ≤ .607). Independent t-tests showed no significant differences between the two profiles in neither contexts (all ts(76) ≤1.48, all ps ≥ .144).

For Fear expressions (see Figure 40, middle column), it was found a significant effect of response dimensions in all contexts (all Fs(7,532) ≥ 101.05, all ps < .001), a non-significant effect of facial motion in all three contexts (all Fs(1,76) ≤ 1.66, all ps ≥ .201), and a significant interaction between facial motion and response dimensions for all three contexts (all Fs(7,532) ≥ 2.17, all ps ≤ .035). Independent t-tests showed significantly higher scores for static than dynamic emotions to the dimension Neutral for context of Happy (t(76) = 2.36, p = .021), and higher scores to the dimension of Surprise for dynamic compared to static facial emotions in all contexts (all ts(76) ≥ 2.22, all ps ≤.029).

Finally, for Pain expression (see Figure 40, right column), it was found a significant effect of response dimensions in all contexts (all Fs(7,532) ≥ 22.33, all ps < .001), a non-significant difference between static or dynamic emotions for all contexts (all Fs(1,76) ≤ .609, all ps ≥ .45), and non-significant interaction for all three contexts (all Fs(7,532) ≤ 1.55, all ps ≥ .148). Independent t-tests showed no significant differences between the two profiles in neither context (all ts(76) ≤ 1.77, all ps ≥ .080).

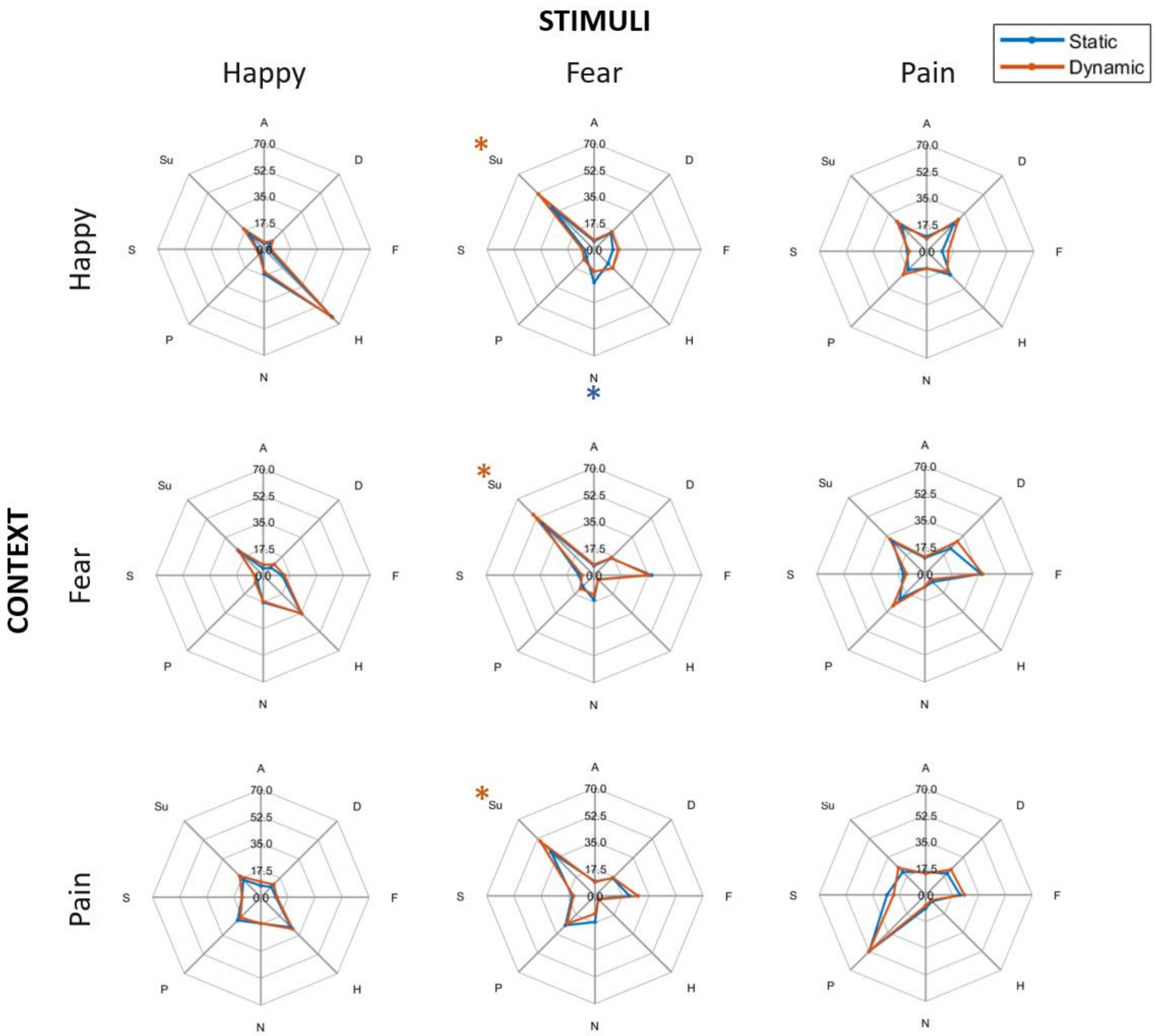

Figure 40. **Contrasts results of Emotion Profiles observed with different Physical contexts for dynamic and static facial emotion.** *Participants responses to static (blue) and dynamic (orange) facial expressions of Happy, Fear and Pain, columns, in Physical contexts intended to evoke emotions of Happy, Fear and Pain, rows. Asterisks indicate significant differences.*

Static vs Dynamic Facial Expressions: Social Contexts**.** Again, facial motion seemed not to affect the emotion profiles observed with different social contexts (Figure 41). For Happy expression (see Figure 41, left column), there was only a significant effect of response dimensions in all contexts (all Fs(7,532) ≥ 30.20, all ps < .001). There was no significant effect of facial motion (for all contexts, all Fs(1, 76) ≥ .105, all ps ≤ . 746), and no significant interaction between facial motion and response dimensions (for all contexts, all Fs(7,532) ≥ 1.67, all ps ≤ .113).

For Fear expression (see Figure 41, middle column), it was found a significant effect of response dimensions in all contexts (all Fs(7,532) ≥ 109.25, all ps < .001), a non-significant effect of facial motion in the three contexts (all Fs(1,76) ≤ .115, all ps ≥ .735), and a significant interaction between facial motion and response dimensions for all three contexts (all Fs(7,532) ≥ 2.42, all ps ≤ .019). Independent t-tests showed significantly higher scores for static compared to dynamic facial emotions to Neutral for the Happy context (t(76) = 2.02, p = .047), and higher scores to Surprise for dynamic compared to static facial emotions in the contexts of Fear and Pain (both ts(76) ≥ 2.17, both ps ≤.032).

Finally, for Pain expressions (see Figure 41, right column), there was a significant effect of response dimensions in all contexts (all Fs(7,532) ≥ 46.81, all ps < .001), a non-significant effect of facial motion for all contexts (all Fs(1,76) ≤ 2.55, all ps ≥ .115), and a significant interaction between motion and response dimensions for the context of Happy (F(7,532) = 2.16, p = .036) but not for the contexts of Fear and Pain (both Fs(7,532) ≤ 1.69 , both ps ≥ .108). Independent t-tests showed no significant differences between the two profiles in neither context (all ts(76) ≤ 1.54, all ps ≥ .127).

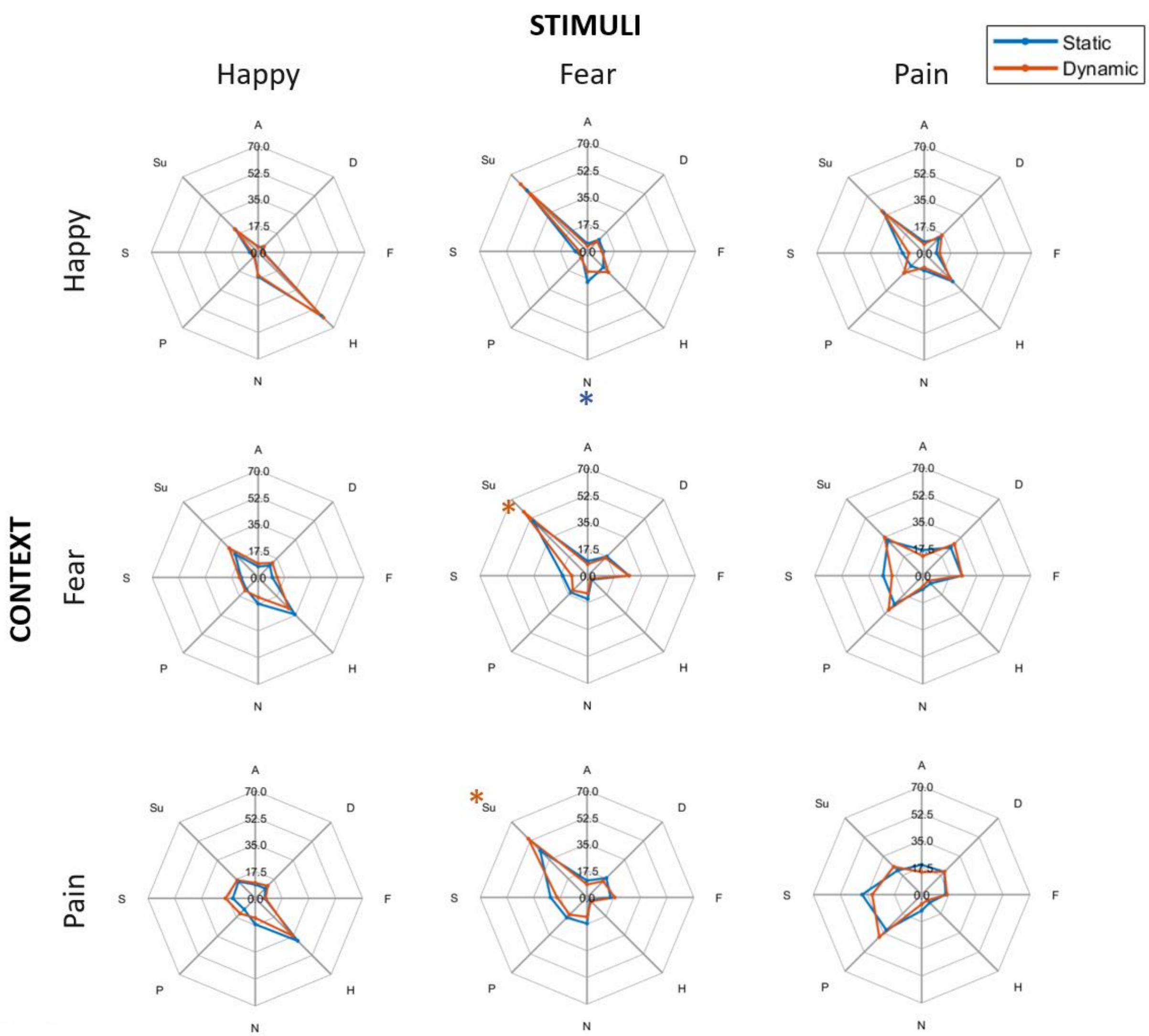

Figure 41. **Contrasts results of Emotion Profiles observed with different Social contexts for dynamic and static facial emotion.** *Participants responses to static (blue) and dynamic (orange) facial expressions of Happy, Fear and Pain, columns, in Social contexts intended to evoke emotions of Happy, Fear and Pain, rows. Asterisks indicate significant differences.*

<u>High vs Low Intensity Emotions: Physical Contexts</u>. Different from facial motion, emotion intensity affected the observed emotions substantially (Figure 42 and 43). For Happy expression (see Figure 42, left column), there was a significant effect of response dimensions in all contexts (all Fs(7,539) ≥ 20.18, all ps < .001), a significant effect of emotion intensity in contexts of Happy and Fear (both Fs(1, 77) ≥ 9.81, both ps ≤ .002) but not in context of Pain (Fs(1, 77) = 1.13, p = .292), and a significant interaction between emotion intensity and response dimensions for all contexts (all Fs(7,539) ≥ 11.36, all ps < .001). In particular, follow-up paired t-tests showed significantly higher scores to high- than low-intensity Happy for the dimensions of Happy, Pain and Surprise in contexts of Happy and Fear (for Happy context, all ts(77) ≥ 5.14, all ps ≤.001; for Fear context, all ts(77) ≥ 5.44, all ps ≤.001), and for the dimensions of Surprise and Happy in context of Pain (both ts(77) ≥ 4.36, both ps ≤.001). Also, higher scores to low intensity happy were found for Neutral in contexts of Happy and Fear (both ts(77) ≥ 6.57, both ps ≤ .001) and for Neutral and Sad in context of Pain (both ts(77) ≥ 2.41, both ps ≤.045).

For Fear expression (see Figure 42, middle column), it was found a significant effect of response dimensions in all contexts (all Fs(7,539) ≥ 100.6, all ps < .001), a significant effect of emotion intensity for all contexts (all Fs(1, 77) ≥ 11.3, all ps ≤ .001), and a significant interaction between intensity and response dimensions for the three contexts (all Fs(7,539) ≥ 18.6, all ps < .001). In particular, follow-up paired t-tests showed significantly higher scores for high- than low- intensity Fear for the dimensions of Disgust, Fear, Pain and Surprise in all contexts (for Happy context, all ts(77) ≥ 2.87, all ps ≤ .005; for Fear context, all ts(77) ≥ 2.38, all ps ≤ .020; for Pain context, all ts(77) ≥ 2.88, all ps ≤ .005). Also, higher scores for low intensity Fear were found for the dimension Happy in the context of Happy (t(77) = 4.17, p < .001) and for Surprise in all contexts (all ts(77) ≥ 3.62, all ps <.001).

For Pain expression (see Figure 42, right column), it was found a significant effect of response dimensions in all contexts (all Fs(7,539) ≥ 22.16, all ps < .001), a significant effect of emotion intensity for contexts of Happy and Pain (both Fs(1, 77) ≥ 5.50, both ps ≤ .022), but not for Fear context (F(77) = .378, p = .540), and a significant interaction between intensity and response dimensions for the three contexts (all Fs(7,532) ≥ 9.10, all ps < .001).

Follow-up paired t-tests showed significantly higher scores for high- than low- intensity Pain to Neutral dimensions for context of Happy and Pain (both *t*s(77) $\geq$ 4.17, both *p*s <.001), and for Neutral, Surprise and Anger for context of Fear (all *t*s(77) $\geq$ 2.56 , all *p*s $\leq$ 012). Also, higher scores for low intensity Pain were found for the dimensions of Disgust, Fear, and Sad in the context of Happy (all *t*s(77) $\geq$ 2.47, all *p*s $\leq$ .015), for Disgust in the context of Fear ($t(77) = 6.43$, $p < .001$), and for Disgust and Happy in the context of Pain (both *t*s(76) $\geq$ 2.52, both *p*s $\leq$ .014).

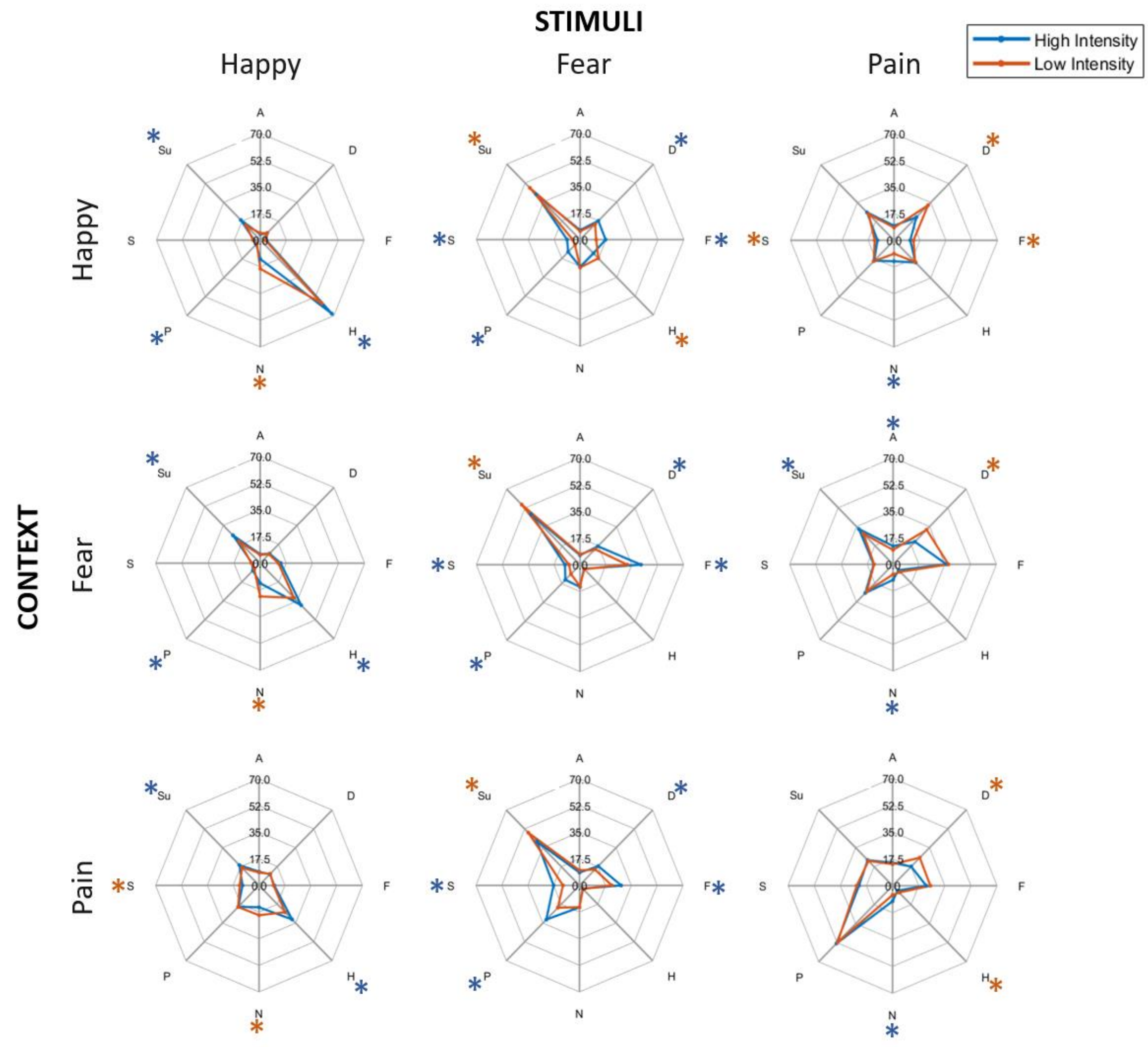

Figure 42. **Contrasts results of Emotion Profiles observed with different physical contexts for high- vs low intensity facial emotions.** *Participants responses to high (blue) and low (orange) intensity facial expressions of Happy, Fear and Pain, columns, in physical contexts intended to evoke emotions of Happy, Fear and Pain, rows. Asterisks indicate significant differences.*

<u>High vs Low Intensity Emotions: Social Contexts</u>. Emotion profiles associated with social contexts also affected by emotion intensity. For Happy expression (see Figure 43, left column) , there was a significant effect of response dimensions in all contexts (all Fs(7,539) $\geq$ 30.10, all ps < .001), a non-significant effect of emotion intensity in all contexts (all Fs(1, 77) $\leq$ 2.84, all ps $\geq$ .096), and a significant interaction between emotion intensity and response dimensions for all contexts (all Fs(7,539) $\geq$ 29.09, all ps < .001). In particular, paired t-tests showed significantly higher scores for high- than low-intensity Happy for the dimensions of Happy and Surprise in contexts of Happy and Fear (for Happy context, all ts(77) $\geq$ 3.26, all ps $\leq$.002; for Fear context, all ts(77) $\geq$ 2.12, all ps $\leq$.036), and for the dimensions of Happy in context of Pain (t(77) = 9.64, p <.001). Also, higher scores for low intensity happy were found for Neutral in context of Happy (t(77) = 8.69, p <.001) and for Neutral and Sad in context of Fear and Pain (for Fear, both ts(77) $\geq$ 5.14, both ps < .001; for Pain, both ts(77) $\geq$ 3.93, both ps $\leq$.045).

For Fear expressions (see Figure 43, middle column), it was found a significant effect of response dimensions in all contexts (all Fs(7,539) $\geq$ 104.5, all ps < .001), a significant effect of emotion intensity for all contexts (all Fs(7,539) $\geq$ 11.3, all ps $\leq$ .003), and a significant interaction between intensity and response dimensions for the three contexts (all Fs(7,539) $\geq$ 15.20, all ps < .001). In particular, paired t-tests showed significantly higher scores for high- than low-intensity Fear for the dimensions of Fear, Pain and Sad in contexts of Happy and Fear (for Happy context, all ts(77) $\geq$ 3.91, all ps < .001; for Fear context, all ts(77) $\geq$ 4.89, all ps < .001), and for the dimensions of Fear, Pain, Sad and Disgust in context of Pian (all ts(77) $\geq$ 2.86, all ps $\leq$ .005). Also, higher scores for low intensity Fear were found for Surprise in all contexts (all ts(77) $\geq$ 4.98, all ps <.001).

Finally, for facial expression of Pain (see Figure 43, right column), it was found a significant effect of response dimensions in all contexts (all Fs(7,539) $\geq$ 46.15, all ps < .001), a non-significant effect of emotion intensity for all contexts (all Fs(7,539) $\leq$ 2.83, all ps $\geq$ .096), and a significant interaction between intensity and response dimensions for the three contexts (all Fs(7,532) $\geq$ 10.21, all ps < .001). Follow-up paired t-tests showed significantly higher scores for high- than low-intensity Pain for Surprise and Neutral for context of Happy (both *t*s(77) $\geq$ 2.99, both *p*s <.004), for Surprise, Neutral and Anger for context of Fear and Pain (for Fear, all *t*s(77) = 2.64, all *p*s $\leq$ .010 ; for Pain all *t*s(77) = 2.65, all *p*s $\leq$ .010). Also, higher scores for low-intensity Pain were found for Disgust in the context of Happy ($t(77) = 5.89$, $p < .001$), for Disgust and Fear in the context of Fear and Pain (for Fear, both *t*s(77) $\geq$ 2.97, both *p*s $\leq$ .004; for Pain, both *t*s(77) $\geq$ 2.25, both *p*s $\leq$ .027).

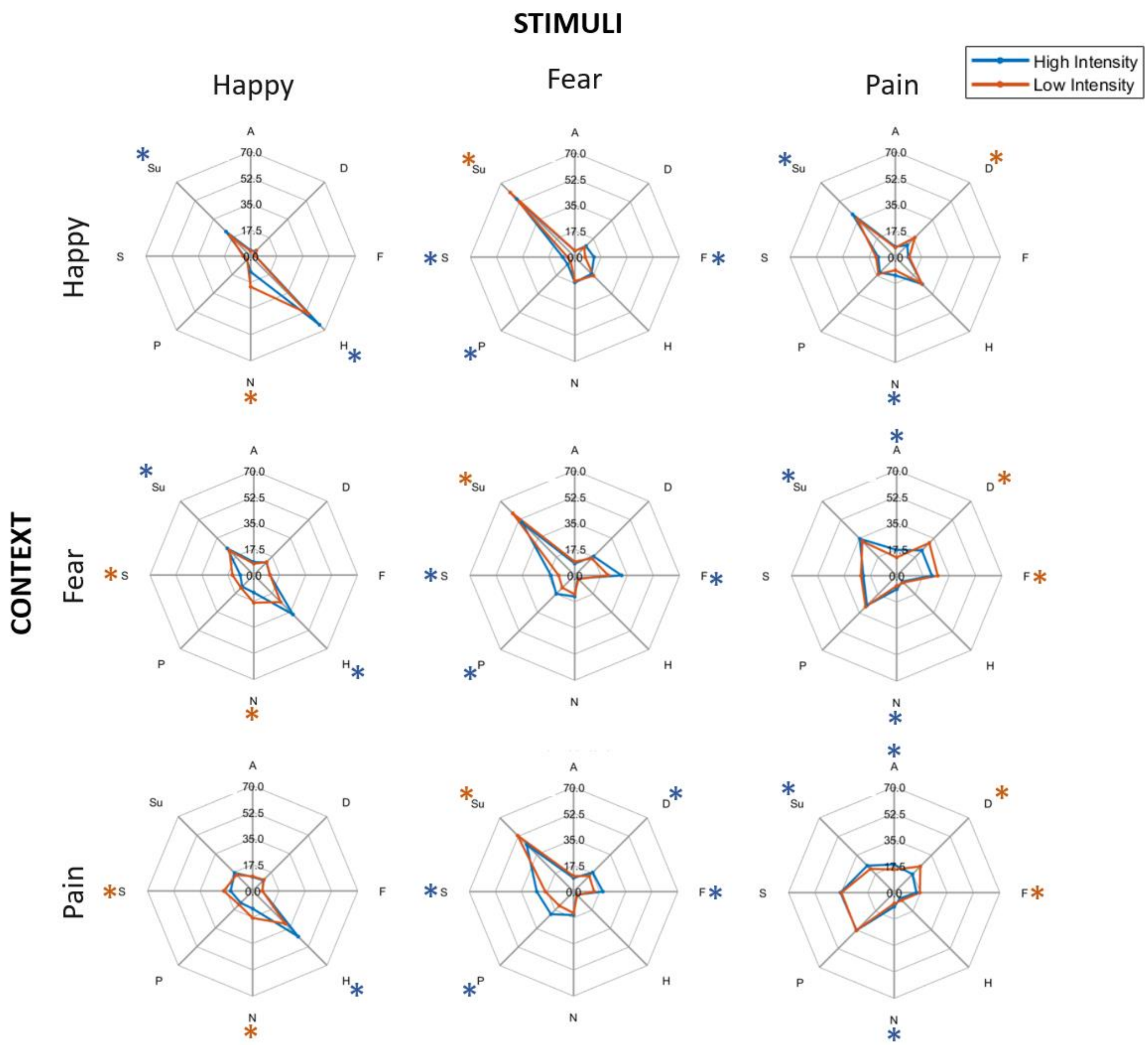

Figure 43. **Contrasts results of Emotion Profiles observed with different Social contexts for high- vs low intensity facial emotions.** *Participants responses to high (blue) and low (orange) intensity facial expressions of Happy, Fear and Pain, columns, in Social contexts intended to evoke emotions of Happy, Fear and Pain, rows. Asterisks indicate significant differences.*

4.4 General discussion

In this chapter, I present two cross-cultural behavioural studies that illustrate how participants associate multiple contextual information to the same facial expression of Happy, Fear, or Pain, creating a rich context profile consisting of emotionally congruent and incongruent scenarios. These context profiles exhibited different levels of variability depending on whether the scenarios evoked the emotion through physical or social stimulation. Responses to social contexts were notably more diverse and richer compared to responses to physical contexts, reflecting a greater engagement with emotionally incongruent scenarios. For instance, when participants assessed the likelihood of observing facial expressions of pain in a set of different scenarios elicited by physical contexts, they assigned higher scores to congruent scenarios involving physical pain, followed by scenarios eliciting Fear (incongruent) and, finally, much lower scores to physical scenarios eliciting Happy (incongruent). In contrast, when participants evaluated the same facial expressions of pain within various social contexts, they assigned higher scores to scenarios evoking Fear (incongruent), closely followed by scenarios eliciting Pain (congruent), and, finally, scenarios eliciting Happy (incongruent), showing a stronger endorsement of scenarios that elicit incongruent emotions.

The finding that participants bind the same facial emotion with multiple social/physical contexts (i.e., context profiles), gained further support from a Random Forest classification model and feature importance analysis. When these context profiles were used to train a model for distinguishing between the three facial emotions, the model's predictive performance often relied on a range of crucial contextual dimensions present in participants' responses. Notably, different key context dimensions emerged for the same facial emotion between social or physical context profiles, highlighting the contrast between the two. However, as observed in previous studies of this thesis (e.g., Study 3), participants' responses to facial expressions of Happy appeared to exhibit a more categorical approach compared to other emotions like Fear and Pain, especially for Physical scenarios where scores were particularly higher for congruent contexts of happiness exclusively.

While it might be surprising that the same facial expression can be linked to both emotionally congruent and incongruent contexts, these findings are consistent with prior research in the literature. Previous studies have demonstrated that identical facial configurations can convey strikingly different emotion categories and different levels of dimensional values (i.e., valence and arousal) depending on the affective context in which they are embedded (Aviezer et al., 2008). Similarly, individuals experiencing intense positive or negative emotions (e.g., pain and orgasm) may produce facial expressions that appear very similar, highlighting the importance of context information in the processing and interpretation of these facial expressions (Barrett et al., 2011; Hughes & Nicholson, 2008). In contrast to the prevailing notion that facial emotions have a one-to-one mapping to corresponding contexts, the present study indicates that the same facial expression can be associated, more or less strongly, with different emotional contexts, resulting in a specific context profile.

Facial emotions seem to be more strongly linked to physical scenarios compared to social scenarios. That is, the same facial expression is associated more strongly with congruent scenarios when judged along a set of physical contexts, while it seems to embrace a richer context profile, involving more incongruent scenarios, when judged along a set of social contexts. This finding generally aligns with the principles of functionalist theories of emotion. These theories argue that contemporary facial expressions may mostly be vestigial reflection of adaptive reactions developed in response to environmental challenges (Barrett & Campos, 1987; Ekman, 1992; Johnson-Laird & Oatley, 1998; Lazarus, 1991; Levenson, 1994; Oatley & Jenkins, 1992). If this is indeed the case, we might anticipate that emotional expressions would exhibit a stronger association with situational contexts that more closely align with the functions they originated from (e.g., avoidance of a physical threat, regulation of physical pain), i.e., physical scenarios. Moreover, while facial expressions elicited by physical stimuli (e.g., encountering a dangerous snake, experiencing a painful wound) are more likely to directly reflect or closely align with the emotions experienced, facial expressions elicited by social situations are the result of a more complex interplay between social rules, expectations, and individual interpretations.

I investigated the influence of culture on the context profiles. Results indicate that the way people link facial emotions to social and physical contexts is affected by their culture background. For facial emotion of Happy and Pain (but not for Fear), Chinese participants tend to assign higher scores to incongruent scenarios, both in social and physical contexts. This tendency of Chinese participants to adopt a wider and richer approach to interpreting facial expressions of emotions is consistent with previous findings of this project (e.g., Chapter 3). In Study 3, compared to British participants, Chinese participants showed a tendency to attribute higher scores to non-target emotions and lower scores to target emotions, providing a more diverse representation of the same facial expressions. Similarly, in this study, Chinese participants were found to associate the same facial expression with a more diverse profile of contexts compared to British participants. These results seem to provide converging evidence to the view that individuals from Asian cultures are more inclined to perceive multiple and blended emotions compared to participants from Western cultures (Fang et al., 2018; Grossmann et al., 2016; Jack et al., 2012; Miyamoto et al., 2010).

The association between facial emotions and related contexts did not exhibit relevant differences between static and dynamic facial expressions. Across both physical and social contexts, participants profiles were very similar, with the exception of static Fear and Happy expressions which received overall higher scores compared to their dynamic counterparts. This trend of assigning higher scores to static rather than dynamic facial emotions was also found in Study 4 investigating semantic dimensions (i.e., semantic profiles). However, it's important to note that the same tendency was not observed in emotional profiles where, instead, responses modulated depending on motion cues with some dimensions perceived as higher and others as lower for dynamic compared to static facial emotions (Study 3). These results may suggest that while dynamic cues seem to contain specifically relevant information about the emotional content of facial expressions, they do not seem to convey additional information regarding the semantic and context aspects of facial expressions. Moreover, the absence of facial motion cues seems to enhance the intensity at which information is extracted from facial expressions.

In line with my previous studies (e.g., Study 3), the intensity of emotion conveyed by facial expressions induced subtle variations in context profiles. When facial expressions conveyed high-intensity emotions, participants tended to assign higher scores to congruent contexts. This finding highlights the value of adopting a profiling approach, as it allows for the reveal of subtle differences in the perception of facial emotions including both emotion contents and the intensity of the emotional content conveyed.

When context profiles were used to derive an indirect measure of similarity between facial expressions (i.e., context profile similarity), I found that context profile similarity significantly correlated with participants' perceptual similarity across cultural backgrounds, static and dynamic facial emotions, and social and physical emotional scenarios. Moreover, a participant-level analysis revealed that this correlation was overall stronger in response to dynamic compared to static facial emotions and stronger for physical compared to social context profiles. These results are consistent with my previous studies (e.g., emotion profiles in Study 3), where the emotion profile similarity obtained from dynamic stimuli also showed a stronger correlation with perceived similarity compared to that obtained with static stimuli. Dynamic cues seem to generate a stimuli representation that better explains how similarities between facial expressions of emotions are perceived. In accordance with the above findings that responses generated from physical contexts were more strongly associated with facial emotions compared to responses generated to social contexts, context profile similarity derived from physical contexts also demonstrated a stronger connection to perceived similarity than that derived from social contexts.

It is important to consider that, in line with previous studies reported in this work, participants' similarity scores to identical stimuli were not excluded from our analyses. Also in this case, despite being recognized as highly similar, identical stimuli were often not rated as identical. While this decision may have slightly inflated the scores resulting from our correlation analysis, we also conducted analyses excluding trials containing identical faces, and the results were not significantly different.

In Study 6, I investigated how participants profile the emotional content of facial expressions presented with either congruent or incongruent contexts. Results demonstrated that the presence of a congruent context enhanced the perception of the target emotion dimensions compared to the condition without any context. This effect was more evident for physical scenarios, while for social scenarios, the enhancement effect was found only for facial emotion of Happy. Responses to other emotion dimensions were also higher with a congruent context compared to no-contexts, for both physical and social scenarios. In previous literature, studies investigating the effect of context on facial expression processing often rely on such congruency effects (i.e., facilitation of emotion perception when context information is congruent) (Todorov, 2010). Our study, however, revealed that congruent contexts not only enhanced participants' perception of the target emotion, but also intensified the perception of non-target emotions, regardless of whether the context provide social or physical scenarios. These findings suggest that while the overall shape of the emotion profile remains relatively constant, the presence of a congruent context enhances the emotion profile (without altering its overall shape), rather than simply facilitating the perception of the target emotion conveyed by that context.

On the other hand, when facial emotions were presented alongside an incongruent context, the influence of the context mostly reduced the perception of the target emotion dimensions while simultaneously enhancing the perception of the emotions evoked by the context. These results, once again, align with previous research showing that information conveyed by a facial expression and its surrounding context is combined during face perception and processing (Aviezer et al., 2008; Carroll & Russell, 1996; Kim et al., 2004; Mobbs et al., 2006; Righart & de Gelder, 2008). Our results illuminate the complex influence of context on the way we extract rich emotional content from facial expressions. While facial emotions are strongly linked to specific physical contexts, they are often associated with a more diversified profile of social contexts.

Emotion profiles obtained with context are also subject to the influence of culture and the intensity of facial emotions. In particular, in line with our previous results, Chinese participants demonstrated a tendency to assign higher scores to incongruent contexts compared to British participants, while mostly no differences were found regarding responses to congruent contexts. It seems that at the same way as Chinese participants tend to extract more complex emotion content to the same facial expressions compared to British participants, as shown in Study 3, they similarly associate a more complex profile of contexts to the same facial expression. Regarding emotion intensity, participants attributed higher scores to congruent contexts when stimuli were conveying high intensity emotions, except for Pain, where responses were similar across levels of emotion intensity. Interestingly, irrespective of whether contexts were congruent or incongruent, participants tended to perceive higher-intensity facial emotions by attributing higher scores to the target congruent contexts.

Finally, there were no significant differences between responses to static or dynamic facial emotions (i.e., just two dimensions among all facial emotions for physical and other two for social contexts), suggesting that facial motion cues do not significantly influence how participants extract emotional content when perceiving faces in context. This is quite interesting, especially when considering our previous results. In our investigation of participants' emotion profiles without any context (i.e., Study 3), facial motion did influence participants' emotion profiles. However, when both context and facial expression information are provided, it appears that motion cues are not integrated and do not significantly impact participants' responses. These results suggest that context information may overshadow the role of facial motion in emotion perception.

Chapter 5

General Discussion

5.1 Summary of relevant findings

In this thesis, I introduced a novel and sensitive approach to the study of facial emotion processing that transcends classical perspectives, which often involve either categorizing our emotional experiences into single, discrete, and universally defined emotion categories or reducing them into a limited set of elementary emotion dimensions. My goal was to unveil the rich and complex information perceived from facial expressions of emotions and to examine how these various forms of information play a role in evaluating the differences and similarities between different facial emotions. I conducted six studies that specifically examined the emotional, semantic, and contextual profiles of emotion perception. These studies sought to understand how this profile information determines our emotion perception, and how our emotion perception, when represented as high-dimension profiles, is modulated by both observer- and stimulus-based factors (e.g., culture background, facial motion, emotion intensity, and physical properties of facial expressions).

The results of this thesis suggest that (1) emotion perception from facial expression is complex, blended, and high-dimensional, consisting of rich information about the emotional content, fine-grained semantics, and related contextual information. This information can be effectively represented as emotion profiles, semantic profiles, and context profiles; (2) emotional profiles, which capture the high-dimensional nature of emotion perception, outperform classical categorical representations of facial emotion in predicting how people make judgments about the differences and similarities between facial expressions of emotion; (3) perceptual similarity between facial emotions is determined by both stimulus-based physical cues and high-level emotional contents, and can be better predicted when both factors are combined; (4) emotion profiles derived from facial expressions are sensitive to participants' cultural background, the intensity of emotions conveyed, the presence of facial motion, and the contextual scenario in which facial expressions are embedded; finally, (5) while deep neural network models trained for emotion perception can achieve human-level emotion categorization, they fall short in fully capturing the richness of our emotion experience as measured by emotion profiles.

When considering the implications of our results, it is important to note that these interesting findings are specific to facial expressions of Happy, Fear, and Pain and we cannot confidently generalize them to all facial emotions. Moreover, when presented with multiple facial expressions, participants often base their judgments on facial expressions by comparing them to previously observed ones (Russell, 1991). Therefore, we cannot exclude the possibility that the presentation of images depicting three emotions throughout the task may have influenced participants decision making process.

In the following sections, I discuss the implications of these findings on the nature of emotion perception; the paradigm used to investigate emotion perception; the way we perceive the similarity and differences between facial emotions; and finally, how our emotion experience, especially that investigated in a laboratory setting, may be joined modulated by both observer- and stimuli-based factors.

5.2 Using emotion profiling to approach the core of natural emotional experience.

Through six behavioural studies I explored the complexity of human emotion experience during the perception of scenario-elicited non-stereotypical facial emotions. The results showed that the high-dimensional emotion profiling approach offers a more comprehensive and fine-grained understanding of emotion perception. This approach not only accounts for the blending and variability across basic emotions, but it also remains sensitive to the influences of both emotional and non-emotional factors on emotion perception.

Previous studies have often relied on posed, stereotypical and sometime exaggerated facial emotions, which have been shown to be significantly different in their dynamics and morphology from real-world spontaneous facial expressions (Delannoy & McDonald, 2009;

Namba et al., 2017; Park et al., 2020; Valstar et al., 2007). To address this gap, I specifically selected static and dynamic facial emotions from the MPI Facial Expression Database (Kaulard et al., 2012), where facial expressions were recorded by asking non-professional actors to mimic the expression elicited by every-day scenarios. Similarly, I defined high and low levels of emotion intensity based on validated inducing scenarios, in contrast to previous studies that defined emotion intensity based on the extent of facial muscle movements (e.g., more or less contracted facial expressions).

When adopting classical tasks commonly used in emotion perception, such as emotion categorization (Study 1) and intensity rating (Study 2), these more naturally elicited facial emotions could be effectively categorized into their target emotion categories and perceived as conveying high or low intensity emotions. More interestingly, when using these more natural stimuli, novel results emerged, which are not entirely consistent with the findings in previous literature. One such finding is a dynamic advantage in emotion recognition, even for high-intensity facial emotions (Study 1). While the beneficial effect of dynamic cues on emotion categorization has been reported in previous works, this is often detected for emotion processing under suboptimal situations, where information about facial emotion is somehow limited such as with point-light or blurred stimuli (Krumhuber et al., 2013). This dynamic advantage often disappears when facial expression information is fully available, or when facial emotions displayed are of high intensity (Bould et al., 2008; Fiorentini & Viviani, 2011; Gold et al., 2013). Here, results unrevealed that the dynamic advantage also applies to high intensity facial emotions where full facial expression information is available.

A second interesting finding is that our participants perceived static facial expressions as being more emotionally intense compared to dynamic facial expressions (Study 2). This is in contract to prior research, which often shows that dynamic expressions lead to higher judgments in terms of emotion intensity, arousal and authenticity (Krumhuber et al. 2013). Biele & Grabowska (2006) found that animations of angry and happy faces are perceived as conveying higher intensity compared to photographs. This phenomenon, referred to as "representational momentum" (Yoshikawa & Sato, 2008), is attributed to the perception that dynamic changes entail a progressive shift in the direction of the observed motion, and is also enhanced by the increased in velocity of these changes in facial expression. However, while posed prototypical expressions typically exhibit a crescendo in muscle contractions leading to a peak, adhering to the onset-apex-offset model, this pattern does not necessarily hold true for spontaneously elicited facial expressions. Posed emotions differ significantly in their dynamics from natural real-world expressions. For instance, assuming that an expression attains its utmost intensity during the apex is not always the case as natural expressions can go through multiple apexes or even achieve a low intensity apex, following a more intricate progression, such as onset-apex-onset-apex-offset, as suggested by Delannoy & McDonald (2009). Similarly, research has shown that the sequence and velocity of Action Units (AUs) may frequently deviate in spontaneous facial expressions compared to posed ones, as evidenced by Namba et al. (2017). These discrepancies in the temporal dynamics between posed and spontaneous facial expressions may explain why the dynamic intensity effect observed in prior studies could not be replicated here.

Another interesting finding emerged with our more natural elicited stimuli, is that while intensity similarity (measured as the difference between rated intensity for two facial expressions) correlated with participants' similarity on a trial level, the same was not found on a group- level (Study 2). In the task each trial displayed two facial expressions from the same identity, which were first rated based on their intensity and then on their similarity, with a total of 9 different identities displayed throughout the task. The significant correlation on a trial-, where the same identity was judged, but not on a group- level, considering all 9 identities, may be explained by the high variability across the 9 actors in conveying emotion intensity. These results highlight one of the core properties of naturally induced facial emotions: they vary across individuals. Such individual differences are often minimized or eliminated in posed stereotypical basic emotions.

The adoption of these more naturally elicited facial emotions in conjunction with more sensitive high-dimensional profiling tasks (Study 3 to 6), revealed that people extract a more rich and diversified emotional, semantic and contextual information than a single emotion category would suggest. In particular, the same facial expression can be represented through a rich profile of different emotions (Study 3), attributed with multiple semantics

(Study 4), and linked to diverse situational contexts (Study 5). These findings not only present a different perspective on how facial emotions are perceived compared to classical theories of emotion processing, but also offer a better account on how people perceive the differences and similarities between facial emotions. Compared to viewing facial emotions as discrete emotion categories, defining facial expressions as high-dimensional emotion profiles better explained participants' perception of facial emotions and their perceptual similarity (Study 3). Moreover, a profile representation of facial emotions proves to be more sensitive to the fine-grained impact that facial motion, emotion intensity, culture and context may have on the way we represent, perceive and process facial emotions. For instance, participants' profiling responses on the emotion content, semantics, and context of facial expressions not only indicate an overall similarity across cultures, facial motion and emotion intensity, suggesting universality in emotion processing, but also reveal consistent cross-culture differences, aligning with a culture dialect view of emotion perception.

The meanings, intensions, and mood expressed in a human face can be very complex and rich, and slightly different expressions may convey dramatically different information. Every-day emotions are fluid and evolving phenomena that do not neatly fit into predefined boxes or labels; they blend and transition, making it challenging to capture their entirety through a single emotion label or a limited number of dimensions. Recently, research transcending low-dimensional models has yielded significant progress in understanding facial emotion processing, creating a new common ground that moves beyond classical dimensional and categorical approaches to emotion perception (Cowen et al., 2019; Cowen & Keltner, 2020; Keltner et al., 2023; Keltner, Sauter, et al., 2019). New models of emotion perception highlight that both emotional dimensions and categories play a significant role in explaining emotional experiences, and that culture (Cordaro et al., 2018; Gendron et al., 2020), prior beliefs (Brooks et al., 2019; Brooks & Freeman, 2018) and contextual information (Greenaway et al., 2018; Wieser & Brosch, 2012) contribute to explain variance in emotion perception across individuals and cultures (Snoek et al., 2023). It is becoming widely accepted that to understand natural facial emotions we need to acknowledge their complexity, adopt research paradigms that use more ecological facial emotions as stimuli, collect rich behavioural responses, account for multiple dimensions of emotions, and capture emotional blends. The results of this thesis suggest that adopting a profiling approach helps us better capture the richness and complexity of emotion perception in daily life.

5.3 The complex nature of facial expressions of emotions

The results of the 6 studies reported here highlight the high-dimensional nature of facial emotion perception, offering fresh insights into the rich and diverse information embedded in a facial expression. Participants' responses to my profiling tasks exhibited a diverse profile of multiple prominent emotional, semantic, and contextual dimensions, accompanied by significant, albeit less prominent, secondary dimensions. In support of the view that participants adopt a profiling approach to emotion perception instead of merely perceiving the target emotion, results from a Random Forest classification model and feature importance analysis showed that differentiating facial emotional experience is supported by diverse set of key emotional dimensions rather than a single prominent one. More specifically, I first demonstrated that participants can often easily perceive several different categories of emotions from the same facial expression, forming a unique Emotion Profile (Study 3). Subsequently, I showed that facial expressions of emotions are linked to more diverse semantic concepts than just the prevailing label of a target emotion category, resulting in coexisting Semantic Profiles (Study 4). Finally, I showed that facial expressions of emotions are associated with multiple context dimensions of emotional scenarios, creating multi-dimensional Context Profiles (Study 5).

The finding that facial expressions of emotions are not solely represented categorically as one basic emotion is consistent with Scherer's Component Process Model (CPM) of emotion perception (Scherer, 1992). According to this theory, facial emotion perception is the result of sequential and cumulative stimulus evaluations that take into account emotional and non-emotional processes. As a result, facial expressions simultaneously and dynamically express both cognition and emotion components (Scherer, 1992). In line with this perspective, facial emotion perception may be more flexible than what classical theories implied, depending on the dynamic interplay between the information extracted from the stimulus and what considered relevant in a given moment (Brosch et al., 2010). In our studies, various types of response profiles provide us with a fine-grained understanding of the rich and diverse emotional experience that participants associate with the same facial expression of emotion, suggesting some degree of flexibility in the way a facial expression may be interpreted in daily life.

Another finding in support of emotion perception flexibility, is that participants' emotion profiles are shaped by various factors, such as the situational context where the facial expression is embedded, the perceived intensity conveyed by the facial expression and the cultural background of the perceiver. For instance, when profiling facial expressions of Pain presented as isolated stimuli, participants perceived a strong component of pain, followed by disgust, fear and surprise, establishing a multiple-dimension representation of the emotional content of the expression. At the same time, when participants are further asked to semantically characterize what constitutes their perception of the emotion pain (i.e., semantic profiling task), they again extract a complex profile of semantics that define and refine their perception. For instance, they link the perception of Pain more strongly to hurt, followed by ache, grief, and anxiety (i.e., a semantic profile). When the same Pain expression is embedded in emotional scenarios eliciting physical pain (e.g., "You got a paper cut"), the emotion profile slightly changes, Pain receives even higher scores, and is now followed by Fear, Sad and Surprise, instead of Disgust. While, when the emotional scenario is eliciting social pain (e.g., "You broke up with your long-time partner"), the score attributed to Pain remains the same, as the non-context condition, but responses to Sad are now as high as Pain. Responses to Fear increased with a physical pain context but remain the same as in the contextless condition with a social pain context (Figure 36).

Results also showed higher variability in semantic profiles compared to emotion profiles within participants of the same cultural background, suggesting that while the emotion content conveyed by a facial expression is more consistent across participants, the way such emotion content is defined and extracted from faces is more prone to individual variability. This finding aligns with the view that our emotion experience is based on culturally learned concepts and social rules (Barrett, 2017; Barrett et al., 2007; Matsumoto, Keltner, et al., 2008), and, as a result, emotions and their processing may be a variable and individual phenomenon. Variability in the frequency and quality of the emotional experience has been observed within individuals over time, across individuals from the same culture, and across cultures (Barrett, 2009). Therefore, whereas classical theories contend that basic emotions are more universally shared experiences, their emotional content and semantics vary across cultures and individuals. For instance, while describing emotions through the same set of words, some people feel the heat of anger, the despair of sadness, the dread of fear, whereas others, instead, experience pleasant or unpleasant feelings with little specificity (Barrett, 1998, 2004; Feldman, 1995). In line with this literature our results showed that, across participants from the same culture, while the shape of emotion profiles tended to be stable, the semantic profiles beyond the target emotional concept varied dramatically, suggesting that, beyond the basic emotion label, the emotional experience is more diversified and complex than that proposed by classical emotion theories.

Participants also extracted a profile of diverse contextual information, recognizing how the same facial muscle configurations may be expressed in a range of emotional contexts. Such contextual profile is even more diverse when elicited by social stimulation compared to physical stimulation. When judged over physically elicited emotional contexts, facial emotions are in fact associated mostly to congruent scenarios (i.e., situations validated to convey the same emotion expressed by the face), while a more diverse profile, involving both congruent and incongruent scenarios, is generated when stimuli are judged over socially elicited scenarios. This finding is in line with functionalist theories of emotion perception, which consider nowadays facial expressions as vestigial reflection of adaptive responses developed in response to environmental challenge (Barrett & Campos, 1987; Ekman, 1992; Johnson-Laird & Oatley, 1998; Lazarus, 1991; Levenson, 1994; Oatley & Jenkins, 1992). That is, expressions of emotions (e.g., Fear) would be strongly associated with the situational contexts that are more in line with the function they developed from (e.g., “A dangerous animal appears”). For instance, adaptive responses such as widening the eyes in expressions of fear help create a larger visual field, allowing for an efficient scanning of the environment for threats (Susskind et al., 2008). Even if the same facial expression may be assumed in social contexts, the relevant adaptive feature loses its function when the threat becomes no-physical (e.g., “You just realised that there is an important exam tomorrow”), which may make the way we associate facial expressions to social contexts more flexible and diverse.

Together, our results suggest that from the same facial expression we can extract a rich profile of emotional content, mostly shared within participants of the same culture, and additional information, linked to the emotional content (i.e., semantic and contextual dimensions), which is more related to the cognitive components of the emotional experience (Scherer, 1992). Profiles of semantic concepts vary within participants of the same culture, depicting a slightly more personalized experience in the processing of facial emotions. This personalized experience also shows some level of flexibility as the way specific emotional profiles are formed is influenced by different factors such as facial motion, intensity of the emotion conveyed, perceiver’s culture background, and the context where the stimulus is embedded.

Interestingly, the influence of culture and contextual information suggests that participants’ adoption of a profiling approach in defining facial expressions, may not be completely driven by low-level, image-based, cues. Additionally, when examining the factors that best explain how participants judge the similarity of different facial expressions, while stimuli physical similarity and emotion (profile) similarity independently contribute to the model performance in explaining participants scores, their contribution do not overlap. This indicates that integrating information about the emotion profile of the stimuli significantly enhance the model’s performance in predicting participants’ scores. Nonetheless, it is important to consider that previous literature shown how the disruption of high-level processes, normally characterizing the way we identify and recognize faces compared to other stimuli, does not influence the way we perceive the emotional content expressed by a face (Lipp et al., 2009).

In literature is well documented how faces differ from other stimuli in that they are processed holistically rather than on the basis of features (Farah et al., 1998). This is supported by the inversion effect (Yin, 1969). The inversion effect It's characterized by the impairment of face recognition and the disruption of holistic processing when faces are inverted but not nonface object. Faces have a specific configuration of features, and our brains are sensitive to this configuration. In the upright orientation, we automatically process these configurations, but when faces are inverted, our ability to perceive and process these configurations is impaired. When a face is upside down, it disrupts the way low-level features are processed, leading to impaired recognition. This implies that while low-level features play a role in face processing, they are not sufficient on their own; they interact with higher-level processes for efficient recognition. In a study of Lipp et al. (2009) it has been shown how face inversion did not affect explicit or implicit evaluation of face stimuli as assessed with verbal ratings and affective priming, suggesting how holistic or configural processing is not necessary to decode facial emotions. However, it is important to consider how in Lipp study to obtain explicit evaluations participants were asked to judge the emotional content of upright and inverted faces by rating how pleasant/unpleasant faces were. Implicit evaluations were also assessed by using an affective priming task that provides a bias-free measure of participants' evaluation of the stimuli on the dimension of pleasantness. It would be interesting to investigate whether the perception of a fine-grained profile of emotion may be influenced when faces are presented inverted.

The view that we adopt a high-dimensional emotion profile representation during the perception of facial expressions is further supported by the finding that perceptual similarities between facial expressions of emotions are better explained by a rich profiling representation of facial emotions compared to a single discrete categorical representation. The role of categorical similarity between facial emotions became largely redundant once emotional profile information is combined with stimulus-based physical information to predict perceptual similarity in a multilinear regression model.

These findings are consistent with emerging view that in our daily life facial expressions of emotion do not usually convey a pure, single affective state (Cowen & Keltner, 2020;

Moore & Martin, 2022), and that emotions are often represented within a semantic space that include plentiful of terms that refer to a rich variety of emotional states (Barrett, 2009; Sabini & Silver, 2005; Shaver et al., 1987). Our results also align with the proposal that, extracting emotion information from a facial expression carries additional meanings and implications that are closely linked to the target emotional content, resulting in a network of coexisting concepts and semantics (Jackson et al., 2019; Liu et al., 2022). In conclusion, when we approach emotion perception by using single discrete emotion categories or a few affective dimensions (i.e., valence, arousal), we may end up over-simplifying the complexity and richness of daily emotional experiences, losing precious information that is crucial for a comprehensive understanding of emotion perception. Recent studies have consistently highlighted the limited power of classical models in explaining the richness of real-world emotion experiences (Cowen et al., 2019; Snoek et al., 2023) and showed that emotion behavior, even conceptualized in terms of basic emotion categories, is high-dimensional and much more complex and nuanced than previously thought (Cowen & Keltner, 2020). Identifying and investigating this complex and nuanced emotional experience may help develop advance our understanding of facial emotion processing (Greenaway et al., 2018).

5.4 Perceptual similarity between facial expressions of emotions

Previous studies on facial emotion perception have often focused on the categorization and recognition of emotions, little research has directly examined how we perceive differences and similarities between facial emotions and what determines their perceptual similarity. Across 6 studies I explored the mechanism underlying participants' perceptual similarity of static and dynamic facial expressions of emotions. In particular, I investigated the role played by stimulus-based physical similarity, and perceiver-based emotion (i.e., categorical and profiling) and Intensity similarity. In line with previous works (Dobs et al., 2014; Xu &

Biederman, 2010) and consistently observed in our studies, stimulus-based physical similarity strongly correlated with participants' perceived similarity for both static and dynamic facial emotions. Physical similarity was also the individual predictor that better explained participants' judgments of stimuli similarity. However, prediction performance is significantly improved when further perceiver-based information was integrated, particularly regarding the emotional content perceived by participants (i.e., categorical and profiling emotion similarity). Categorical and profiling emotion similarity both correlated with perceptual similarity on a group-, participant-, and trial- level, across cultures and stimulus type (i.e., dynamic, static). Moreover, a multiple linear regression analysis showed that fine-grained emotion profiling similarity contributes significantly more to explaining variations in participants' perceived similarity compared to categorical similarity, especially when dynamic cues were available. Overall, when facial motion was integrated, all behavioural models (i.e., Profiling, Categorical and Intensity Similarity) better explained perceptual similarity compared to their performance in response to static stimuli. Regarding emotional Intensity similarity, while this slightly enhance the model performance for static facial emotions, it did not contribute to improve the performance for dynamic facial emotions, suggesting that dynamic cues made explicit intensity information redundant.

These results suggest that when judging the similarity between facial expressions, people integrate both stimulus-based and perceiver-based information, mostly regarding facial expressions' physical and emotional similarity. These findings are in line with recent research showing that both stimulus-based cues and perceiver's conceptual knowledge about emotional concepts influence performance on tasks requiring the perceptual matching and categorisation of facial expressions of emotion (Brooks & Freeman, 2018). In particular, idiosyncratic differences in emotion concept knowledge can predict subtle differences in how those emotions are perceived (Brooks & Freeman, 2018). Stimulus-based information such as facial shape and surface textures have also been shown to be important cues contributing to the perception of facial emotion. For instance, Sormaz, Watson, et al. (2016) employed Procrustes analysis to compute the similarities between face shapes and correlated images' pixel intensities to compute the similarities of surface textures. They found that these stimulus properties predicted subjective ratings of perceptual similarity. Recently, Murray et al. (2021) investigated the relative roles of shape, surface, and conceptual information in the perception and categorisation of facial expressions. Using multiple linear regression analysis, they found that the similarity of emotion concepts was related to emotion categorization and perception, while the similarity of face shapes was more related to emotion perception than categorization and the similarity of surface textures was more related to emotion categorization than perception.

Importantly, compared to previous works, the present study not only highlights the importance of both stimulus- and perceiver-based information in judging the similarity of facial emotions, it also shows that a fine-grained profile representation of facial emotion can better predict perceptual similarity compared to that based on categorical emotion information. Interestingly, a participant-level analysis also revealed that dynamic stimuli seem to enhance the use of emotion profile information in participants' similarity judgments. While this effect should be investigated further, this result highlights the relevance of facial motion in conveying key cues for the extraction of emotional content from faces, and suggest that people rely more on rich and multi-dimensional emotional information than single emotion label for judging the similarity and differences between facial emotions.

Finally, investigating perceptual similarity, I found some interesting differences across cultures. A participant-level analysis revealed that British participants seem to integrate emotion profile information in their perceptual similarity judgments more compared to Chinese participants, while Chinese participants seem to integrate physical information in their perceptual similarity judgments more compared to British participants. While this result is quite interesting, it is important to take into account that stimuli used in these studies all depicted western actors. In the literature, it has been systematically reported an in-group advantage in processing facial expressions; people seem to be more accurate in judging emotional expressions from their own cultural group (Elfenbein, 2015; Elfenbein & Ambady, 2002). We may speculate that Chinese participants rely more on physical similarity in their similarity judgements may be the result of a disadvantage in promptly extracting or making use of the emotional content compared to British participants.

To sum up, our results show that a high-dimensional representation of facial emotion better explains, compared to a categorical approach to emotion perception, the way we perceive differences and similarity between facial expressions of emotions. I also showed that facial motion seems to transmit key information regarding the emotional content of a facial expression and when asked to judge the similarity of dynamic stimuli we are more likely to form our responses based on stimuli's emotional content. Finally, while previous research demonstrates cultural differences in emotion recognition and categorization, my results show that cultural differences in emotion processing also exists in the way we perceive differences and similarity between facial expressions. However, caution should be taken when generalising this specific finding, as I only used facial expressions of western people in the study and further research is needed to clarify this point.

5.5 The role of facial motion, emotion intensity, culture and context in profiling facial emotions

In our daily life facial information is rarely presented to us in isolated stimuli. When extracting meanings from facial expressions we are often part of a complex and dynamic environment where different factors jointly determine the way emotion is expressed, perceived, and regulated. To establish theoretical models that are able to better explain how facial emotions are perceived in the real world, it is important to consider both emotional and non-emotional and both face- and observer-based factors that may affect how facial emotions are expressed, perceived, and regulated. In the 6 studies of this thesis, I specifically investigated the role of facial motion, emotion intensity, participants' cultural background and the context where faces are embedded, in perceiving and profiling facial emotions.

5.4.1 Facial motion

The role of facial motion in emotion processing has often been investigated in relation to our ability to correctly recognize the emotion of posed stereotypical facial expressions. In particular, a beneficial effect of dynamic cues has been reported for faces shown in suboptimal situations (e.g., point-light, blurred stimuli), and this beneficial effect seems to disappear when facial information is fully available or when the emotion displayed is of high intensity (Bould et al., 2008; Fiorentini & Viviani, 2011; Gold et al., 2013)

. Previous studies have also shown that facial motion information, besides enhancing the coherence in the identification of affects, also leads to stronger emotion judgments and facilitates the differentiation between posed and spontaneous expressions (Krumhuber et al., 2013). Our findings, while generally aligning with previous research, offer new insights into how facial motion influences the fine-grained perception of multi-dimensional emotional content from facial expressions. Consistent with previous studies, I found a dynamic advantage in the categorization of non-stereotypical facial expressions of Happy, Fear and Pain, even though face stimuli were not presented in a suboptimal situation and the emotions depicted were of both high and low intensity. This result suggests that, for less prototypical and naturally elicited facial expressions of emotions, dynamic cues convey crucial information that facilitate people interpretation and categorization of facial expressions.

Facial motion also affects the perception of emotion intensity of facial emotions. However, the results are inconsistent with previous works. Our participants tend to rate static facial emotions as more intense than dynamic ones. In previous works, dynamic facial expressions are often perceived as having higher level of intensity, arousal and authenticity than static facial expressions (Krumhuber et al., 2013). This intensity effect is often attributed to the “representational momentum” of dynamic change that implies a forward shift in the direction of the observed motion. As the velocity of change in facial expression increases, the more intense facial expression will be perceived. However, while posed prototypical facial expressions often have a crescendo in the muscle contractions toward the peak, following the onset-apex-offset model, this is not necessarily true for elicited non-stereotypical facial expressions. Temporal analysis has shown that more natural expressions could have multiple apexes, or may apex a low intensity, following a more complex development (e.g., onset-apex-onset-apex-offset) (Delannoy & McDonald, 2009). Similarly, the order of appearance and the velocity of Action units also often differ in spontaneous compared to posed facial expressions (Namba et al., 2017). In conclusion, these discrepancies in the temporal dynamics between posed and spontaneous facial expressions may explain why the dynamic intensity effect observed in prior studies could not be replicated here. Together these results highlight, once more, that findings generated from the investigation of posed prototypical emotional expressions cannot always be extended to more spontaneous facial emotions, supporting the importance of adopting stimuli that more closely resemble our daily experience.

When looking at fine-grained multi-dimensional emotion perception, the results showed that facial motion does not simply enhance or weaken the perception of different emotional contents. Consistent with the view that facial motion conveys relevant information regarding the emotional content of a facial expression, I found that dynamic cues do modulate the perceived emotion profiles, with certain emotion dimensions perceived more strongly with dynamic stimuli while other with static stimuli. Different from its role in the emotion profiles, facial motion does not strongly affect how participants extract high-dimensional semantic or contextual information from facial expressions. Participants' semantic and context profiling responses were generally very similar. Therefore, while dynamic cues seem to specifically contain information about the emotional content, which modulates emotion profiling responses, they may not convey crucial information regarding the semantic or contextual information of facial expressions.

Results of perceptual and profile similarity provide further evidence that facial motion is mostly related to the extraction and interpretation of the emotional content of facial expressions. The presence of dynamic cues affects similarity judgments. When facial motion cues are available, participants tend to make their judgments on stimuli similarity more based on the emotional content (i.e., emotion profiles) compared to when static stimuli are used. However, interestingly, the effect of facial motion on emotion processing is also modulated by the context where facial expressions are embedded. The presence of motion cues does not significantly modulate the extraction of the emotional content when faces are embedded into a specific context. Emotion profiles formed in the absence of emotional context are modulated by facial motion, however, when both context and facial expressions are presented simultaneously, facial motion cues no longer significantly influence participants' emotion profiling responses.

In sum, our findings suggest that while facial motion convey relevant information about the emotional content extracted from facial expressions, enhancing our ability to identify the intended emotions, and affecting our extraction of complex emotional content. The effect of facial motion on emotion profiles can be overshadowed by contextual information, which seems to be more influential on emotion perception than dynamic cues. When contextual cues are presented, emotion profiles obtained with static and dynamic facial expressions become similar to each other.

5.4.2 Culture

Previous studies have shown that Asians and Western participants differ in how emotions are experienced, recognized and expressed. Asian participants show a tendency to experience multiple different emotions concurrently compared to Western who are more likely to feel specific emotions (Grossmann et al., 2016; Miyamoto et al., 2010). Chinese observers endorse non-target emotions more than Dutch observers, specifically when it comes to morphologically similar emotions (Fang et al., 2018). Jack et al. (2012) used a FACS-based random facial expression generator and reverse correlation to reconstruct 3D dynamic models of the six basic emotions in Western Caucasian and East Asian cultures, founding that while Western Caucasians represent each of the six basic emotions with a distinct set of facial muscles, East Asians show a considerable overlap between emotion categories. To sum up, Asians seem to be more inclined to experience and perceive blends of emotions than Western participants.

In line with the literature, comparing participants' performance on the extraction of multi-dimensional emotional content from facial expressions of emotions, I found that

Chinese participants tend to attribute lower scores to target emotion dimensions while detecting higher levels of other co-occurring dimensions, providing a more diverse representation of the same facial expression compared to British participants. Similarly, for context profiles, the results show that Chinese participants tend to give higher scores to non-target contexts compared to British participants, resulting in a more diverse profile of contexts associated with the same facial expressions. Therefore, it seems that Chinese participants tend to extract more complex emotional content from faces and similarly associate a more complex profile of contexts to the same facial expression compared to British participants.

Similar cross-culture differences are observed in response to facial expressions presented within specific contexts. Participants' responses are influenced by both culture and context, with Chinese participants attributing to non-target emotion dimensions higher scores compared to British participants. However, it is important to consider that Nisbett (2003) suggests that Western Europeans tend to employ a more "analytic" pattern of attention, where they categorize reality into discrete elements with defining attributes. In contrast, East Asians adopt a more "holistic" pattern, perceiving people, objects, and events in relation to one another rather than focusing solely on their individual properties. This difference is reflected in Eastern philosophical traditions like Confucianism, Buddhism, and Taoism, which emphasize holistic perspectives, while Western traditions such as Platonism, Aristotelianism, and monotheism highlight the distinct characteristics of entities and individuals. Western individualistic cultures like those in America tend to endorse open emotion expression, while Asian, collectivistic cultures encourage controlling expressions of affect to maintain group harmony (Markus, 1991; Matsumoto, Seung Hee Yoo, et al., 2008). Americans typically interpret emotional expressions as spontaneous reflections of inner feelings, perceiving individuals as autonomous entities with personal goals (Markus et al. 1991; Markus et al.1997). In contrast, East Asians see people as interconnected, where relationships are fundamental to consciousness. Emotions of an individual in Asian contexts are seen as inseparable from the feelings and responses of the larger group (Markus et al., 1997). In a study of Masuda et al. (Masuda et al., 2008), compared to Americans, Japanese participants incorporated the emotions of background figures when evaluating central person's facial expressions. This was also reflected in their patterns of attention, with Japanese participants looking at the surrounding people more than Westerns.

Thus, the difference in reading face in context between Chinese and British participants, may also being explained as Chinese participants integrating more emotional information from the contextual cues compared to British participants. Similarly when participants are asked to profile facial expressions without a context, differences may be explained by the fact that Chinese participants may have more difficulties in defining the emotional content conveyed without other relevant contextual cues.

Interestingly, when extracting semantic content from facial expressions, response profiles of Chinese and British participants differ only on a few dimensions, suggesting that they tend to agree on the semantic meanings linked to the intended emotion extracted from facial expression. Together, our results suggest that Chinese and British participants differ in the emotional content and related contexts perceived from facial expressions, whereas they agree on the fine-grained semantics extracted from the same expression, which are more closely associated to the target emotion and the congruent contextual scenarios. To sum up, cross-cultural differences are primarily seen in Chinese participants' higher tendency to form a more diverse representation of the same facial expression than British participants.

I also investigated cross-cultural differences in the way participants perceived differences and similarities between facial expressions of emotions. Participants’ cultural background impacts their emotion profile similarity more than their perceptual similarity. Moreover, differences in perceptual similarity suggest that British participants integrate emotional profile information in their similarity judgments more than Chinese participants, who, conversely, tend to integrate more physical information in their similarity judgment than British participants.

Nonetheless, it is important to note that stimuli used in these studies depicted facial expressions of western non-professional actors, and the observed difference might be caused by an in-group advantage in processing facial expressions of emotions (Elfenbein, 2015). Our Chinese participants may be driven to judge the similarity of facial expressions based more on physical similarity because of a disadvantage in accurately extract or make use of the emotional content compared to British participants. Similarly, Chinese participants' tendency of attributing higher scores to non-target emotion and context dimensions may be due to their difficulty in perception of out-group facial emotions. Further investigations are needed to better understand the nature of the cultural differences observed in our studies.

5.4.3 Emotion Intensity

Different levels of emotion intensity are usually obtained by directly controlling the physical muscle involved in facial emotion, such as exaggerating muscle movements, or using a morphing between two facial expressions. A few assumptions are at the basis of these methods which are generally in line with categorical and dimensional theories of emotion. One assumption is that emotional intensities differ primarily in a quantitative rather than qualitative manner. In other words, experiencing intense happiness is not fundamentally distinct from experiencing mild happiness in terms of their intrinsic nature. Instead, they share the same underlying experience that is heightened in intense emotions. As a result, different levels of intensity are expressed through the same patterns of facial muscles. Another assumption is that, as for the underlying emotional experience, the emotion intensity expression differs more quantitatively than qualitatively. This means that the more facial muscles are contracted or displayed, the more intense the expressed emotion is.

Different from previous studies, the stimuli used in this project have been selected to convey a more natural and semantic definition of emotion intensity. Facial expressions conveying different intensities of emotion were elicited by specific emotional scenarios, which have been validated to evoke a high or low intensity emotion. Our results show that participants could effectively detect different levels of emotion intensity, both when they were asked to directly rate the intensity of the stimuli (Study 2), and when they were asked to extract multi-dimensional emotional content (Study 3). In this case, responses to the target emotion dimensions were higher for stimuli depicting high intensity emotions compared to low intensity emotions.

Interestingly, similar results were obtained for the profiling of semantic and contextual information (Study 4 and 5). Importantly, this influence of emotion intensity was also observed when facial emotions were embedded in contextual scenarios. Regardless of whether the context was congruent or non-congruent to the facial emotion, participants perceived higher intensity from the expressions conveying high intensity emotions, as demonstrated by the attributing of higher scores to the target emotion dimension (Study 6). Together, these results suggest that natural variation of emotion intensity, as defined by its eliciting emotional scenarios, can be effectively detected by participants. Moreover, the profiling approach to emotion perception can not only detect the diverse emotional content, but it is also sensitive to the different levels of perceived emotion intensity, even when participants are engaged in indirect and unrelated tasks (e.g., semantic and context profiling tasks).

Importantly, our results show that emotion intensity not only influences the way we perceive the target emotion dimension, but it also slightly modulates the rest of the emotion profile. In response profiles, the intensity of the facial emotion slightly modulated the perception of non-target emotions, with some of these non-target emotions perceived more strongly with high intensity stimuli and other with low intensity stimuli. This result suggests that perceived emotion intensity carries additional information about the emotional content of the expression, not only intensity modulate *quantitatively* the “amount” of emotion perceived but also shape *qualitatively* the profile of the emotional content extracted from the facial expression. However, since the different levels of emotion intensity were generated by different emotional scenarios, the slight changes in responses may be the result of the different shades of emotions conveyed by the different expressions.

5.4.4 Context

Functionalist theories of emotion emphasize the adaptive utility of emotional expressions within specific contexts. These theories posit that emotions have evolved as adaptive responses to environmental challenges, optimizing human responses to the demands of the physical and social environment (Barrett & Campos, 1987; Ekman, 1992; Johnson-Laird &

Oatley, 1998; Lazarus, 1991; Levenson, 1994; Oatley & Jenkins, 1992). As a result, nowadays facial expressions of emotions have been adapted for their communicative role. Consistent with their origins, expressions of emotions are expected to be more strongly associated with situational contexts that are in line with the functions they originally developed to fulfill. At the same way, we can expect that similar expressions, maybe originated from the same evolutive function, may have evolved over time to suit various social and communicative needs, and that even the same expression could be carrying out different social or communicative meanings depending on the specific social context where this is embedded (Aviezer et al., 2008; Barrett et al., 2011; Hughes & Nicholson, 2008).

In Study 5, I showed that participants form a profile of the contextual scenarios that they associate with a facial expression of emotion, particularly when the emotional scenarios considered are socially related. This finding suggests that the same facial configurations can in fact be linked to strikingly different contexts, highlighting that context is an important dimension of emotion processing and interpretation. Furthermore, the same facial expression is more strongly associated with congruent scenarios when judged over physically elicited emotional contexts, while it generates a more mixed context profile, involving both congruent and incongruent scenarios, when judged over socially elicited scenarios. This result aligns with functionalist theories of emotion. When facial expressions are triggered by physical stimuli (e.g., encountering a dangerous snake or experiencing a painful wound), they are more likely to reflect the original emotional experience from which these emotions evolved. In contrast, when facial emotions are elicited by social contexts, they are interpreted as a result of more complex interaction between social rules, expectations, and personal interpretations, thereby linking to a more diverse range of possible social contexts.

Previous research has shown that contextual information plays a key role in shaping the way we express and perceive emotions (Barrett et al., 2011; Greenaway et al., 2018; Wieser & Brosch, 2012). A key evidence supporting this claim is the “congruency effect”, a facilitation in the perception of the target emotion when the information conveyed by the context is congruent with it (Todorov, 2010). In line with the literature, I found that the presence of congruent contexts can enhance the perception of target emotion dimensions compared to non-context situations. Again, this facilitation effect is more evident for physical contexts, highlighting, once more, their stronger association with the facial emotions compared to social contexts. Interestingly, responses to other emotion dimensions are also heightened in the presence of a congruent context compared to conditions without any contexts, which is observed for both physical and social contexts. This result suggests that the presence of a congruent context enhances perception of more diverse emotions, probably due to their relevance to that specific context.

Regarding the role of incongruent contexts in emotion perception, previous studies have shown that the information conveyed by facial expressions and contexts are often combined and influence each other (Aviezer et al., 2008; Carroll & Russell, 1996; Kim et al., 2004; Mobbs et al., 2006; Righart & de Gelder, 2008). The present studies also showed results in line with the literature. Incongruent contexts influence the way our participants form their multi-dimensional emotion profiles extracted from facial expressions. When facial emotions are presented together with incongruent social or physical contexts, perception of the target emotion (related to facial expression) is reduced, whereas perception of context-related emotions is enhanced. Future research is needed to further quantify the relative contribution of facial expressions and emotional context to our emotional experience (e.g., measured as emotion profiles).

Finally, while the contexts presented in our study are a small subset of the numerous possible contexts that could be relevant to the three facial emotions we investigated, participants' responses to these contexts (i.e., context profiles) do have a functional role in the way they perceive the similarity and differences between facial expressions of emotions. Our findings indicate that the perceived similarity computed solely based on the associated contexts, can effectively predict the perceptual similarity between facial emotions as observed in a direct rating task. Again, the prediction power is stronger for profile similarity based on responses to physical contexts compared to social contexts. These results suggest that participants perceived rich and multi-dimension information does contribute to their judgement on how two facial expressions differ from each other.

5.6 Conclusions and future directions

The way we perceive the world around us is strongly shaped by top-down categorization processes that continuously guide and constrain how we process incoming information (Barett, 2006; Davidoff, 2001; Palmeri & Gauthier, 2004; Rosch, 1975; Rosch et al., 1976). Through this categorization process, we attribute meaning to the stimuli perceived by classifying them according to multiple principles such as perceptual similarity (Rosch, 1978), semantic rules (Murphy & Medin, 1985), and implications for goal states (Barsalou, 1983). This process applies to colours, shapes, orientations, objects, faces, as well as facial expressions of emotion (Barrett, 2006; Cohen & Lefebvre, 2005; Davidoff, 2001). In this thesis, I showed that simply classifying facial expressions of emotion into a single discrete emotion category cannot fully capture the rich and diverse information conveyed by faces and perceived by people. Instead, our experience with facial expression of emotion can be better explained by multi-dimensional emotional, semantic and context profiles reconstructed with a profiling approach.

Facial expressions are a powerful and complex form of nonverbal communication. However, its richness and complexity, in expression and perception, has been dramatically reduced or simplified in both classical theoretical frameworks (e.g., BET or intensity-valence models) and prevailing research methodologies (e.g., stimuli depicting posed prototypical facial expressions, force-choice categorization tasks, etc.). Given the vast amount of possible facial movements and the near infinity number of emotionally related social and physical environments we may encounter in daily life, a reductionist approach (by reducing emotion contents to limited number of emotion category or dimensions) may be not sensitive enough to reveal the rich and gradient representation that we have of facial, bodily, and vocal emotions in real life (Cowen et al., 2019; Jack et al., 2018).

In this project, I explored a more sensitive approach to the study of facial emotion processing, showing that the adoption of a multi-dimensional profiling paradigm may establish a more comprehensive framework that is able to better account for the blends and variability inherent emotion perception, better explain our perceptual similarity between facial emotions, and uncover the fine-grained influences that both emotional and non-emotional, face- and observer-based factors, may have on emotion perception.

Our results consistently demonstrated that, from the same facial expression, we can extract a rich profile of emotional content and related cognitive components (i.e., semantic and contextual information). Our findings support the idea that classic basic emotion categories are more a matter of degree than all-or-none, and that there are no sharp clear-cut boundaries separating different emotion categories. The observed *emotion*, *semantic* and *contextual profiles* indicate that multiple dimensions of information are simultaneously activated and available to our conscious experience, resulting in something similar to the co-occurrent meanings for categorization (Rosch, 1978; Rosch et al., 1976). This rich and complex representation also affords some degree of flexibility in the way facial expressions are interpreted. As shown in our study, these diverse dimensions of emotion perception could be influenced by various factors such as facial motion, intensity of the emotion conveyed, perceiver's culture, and context where the stimulus is embedded, all playing a role in shaping the profiles of the emotion we experience.

Due to the key role played by facial expressions in our social life, models and theories of facial expression processing not only help us understand the underlying mechanisms of the emotion processing, but also have a wider influence on our daily life. For instance, theories formulated to interpret and make sense of facial emotions become part of our mental structure, influencing the way we make assumptions about the others, the way we communicate, the way we construct our educational practice, and guiding our diagnostic and therapeutic approaches (Barrett et al., 2019). Similarly, machine learning techniques for face processing have been implemented into wide settings of our life, with the intent to extract relevant information from faces and to ultimately interact with humans naturally (e.g., by identifying and classifying combinations of facial movements into a defined set of emotion categories) (Altameem & Altameem, 2020; Fei et al., 2020; Kaushik et al., 2022). Tools designed to improve our ability to read faces or facilitate children development are often developed in accordance with influential emotion perception theories (Holmes, 2011; Margoudi et al., 2016; Payton et al., 2000).

Given the fundamental influence of The Basic Emotion Theory, machine-based face processing systems are still mostly trained based on the basic emotion assumptions, involving a few categories of posed stereotypical expressions of the seven basic emotions. Not surprisingly, such systems tend to perform worse when tested with non-stereotypical fine-grained facial expressions (Pantic & Bartlett, 2007; Stöckli et al., 2018; Yitzhak et al., 2017) . Efforts to develop algorithms that can process more natural facial expressions are emerging, with some models' performance being comparable to that of humans (Krumhuber et al., 2021). However, in the present project, when comparing the performance of a machine-learning based algorithm to our participants' responses, results indicate that, even though such computational model can achieve human-level performance in a standard forced choice emotion categorization task, significant differences are observed between human and model performance when extracting rich fine-grained emotional content (i.e., emotion profiles) from the same faces. If we approach facial emotion perception through a multi-dimensional profiling approach, future research may similarly train machine learning algorithms with emotion profiles and examine whether algorithm-based responses can capture and explain the fine-grained characteristics of human emotional experience.

While the present project focuses on how perceptual similarity between facial expressions may be predicted by their physical, categorical, intensity and profiling similarity, the neural mechanisms supporting these associations remain to be elucidated. Results from behavioural studies reported in this dissertation suggest that both stimulus-based and perceiver-based information influence perceptual similarity between facial emotions. It could be hypothesized that these two types of information contributing to emotion processing may be supported by neural processes involving separated brain regions. Specifically, the Occipital Face Area (OFA), sensitive to physical facial information, and the posterior Superior Temporal Sulcus (pSTS), sensitive to social emotional information of faces. These two areas may support different functions underlying similarity judgments: low-to-middle-level physical/appearance-based face similarity (in OFA) and high-level emotional/semantic/intensity-based face similarity (in pSTS). Even if it is beyond the scope of the present dissertation to test this hypothesis, I conducted an EEG study to obtain measures of neural similarity between perceptions of facial emotions, with the aim to further investigate the mechanism underlying our perceptual similarity of facial emotions and explore the neural mechanisms supporting it. Data analysis is still in progress; however, I expect the similarity of neural responses in these two brain regions to be associated with corresponding stimulus-based (i.e., physical similarity) or perceptual similarity between facial expressions.

Research using functional magnetic resonance imaging (fMRI) with Multivariate Pattern Analysis (MVPA) has demonstrated that basic emotion categories of facial expressions are decodable from patterns of neural activation (Harry et al., 2013; Said et al., 2010; Wegrzyn et al., 2015) in three core face processing regions (the Fusiform Face Area (FFA), the OFA, and the Superior Temporal Sulcus (STS); Haxby et al., 2000). Similarly, Machine learning based classifiers can be trained to successfully discriminate between patterns of activation in response to different emotions in EEG studies (Farran et al., 2020; Mares et al., 2020). While previous research showed that perceptual similarity of facial expressions can be explained by representational similarity of stimulus-properties (i.e., surface and feature shape information in the image) (Sormaz et al., 2016), and that the perceptual similarity of expressions could be predicted from the patterns of neural response in the face-selective STS (Said et al., 2010; Sormaz et al., 2016), the relative roles of STS and OFA in emotional profiling and their similarities remains unknown.

Further investigations are also needed to better understand the nature of the cross-cultural differences observed in our studies. An in-group advantage in processing facial expressions has often been reported, with people being more accurate at judging emotional expressions from their own cultural group. Different theories have been proposed to explain this phenomenon, arguing that in-group faces convey culturally specific elements of emotional expressions, that emotion is a universal language characterized by subtly different dialects, and that individuals may be less motivated to recognize the emotions of other individuals of foreign cultures (Elfenbein, 2015). Since the stimuli used in the present studies only depict Western non-professional actors, further investigations should be conducted to determine whether the cultural differences detected in our studies are the results of a cultural disadvantage in perceiving out-group facial expressions of emotions or are cultural-specific characteristics to the extraction and processing of facial emotions.

Furthermore, our results also show that, within the same culture, while the extraction of emotional content is relatively consistent across participants, semantic profiles vary more, depicting a slightly more personalized experience in the semantic processing of facial emotions. This may be further investigated to determine whether such difference between emotional and semantic profiles persists for facial emotions of different cultures, thereby reflecting a high variability and flexibility in the way we specifically experience and interpret facial emotion beyond their basic emotional labels.

In conclusion, to understand our daily emotional experience we need to acknowledge its complexity. Adopting a more holistic and ecological approach, such as the profiling tasks used in the present thesis, will not only help us reveal the rich and diverse contents of our emotion perception, but it will also have wide implications in the way theories of emotion processing are used in developing products and technologies capable of face processing, and in informing educational and therapeutic interventions.

References

Abelson, R. P., & Sermat, V. (1962). Multidimensional scaling of facial expressions. *Journal of Experimental Psychology*, *63*(6), 546–554. https://doi.org/10.1037/h0042280

Addabbo, M., Longhi, E., Marchis, I. C., Tagliabue, P., & Turati, C. (2018). Dynamic facial expressions of emotions are discriminated at birth. *PLoS ONE*, *13*(3). https://doi.org/10.1371/journal.pone.0193868

Altameem, T., & Altameem, A. (2020). Facial expression recognition using human machine interaction and multi-modal visualization analysis for healthcare applications. *Image and Vision Computing*, *103*, 104044. https://doi.org/10.1016/j.imavis.2020.104044

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). *Deciphering the Enigmatic Face The Importance of Facial Dynamics in Interpreting Subtle Facial Expressions*. http://www.pitt.edu/$ambadar/stimuli.htm.

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., & Bentin, S. (2008). *Angry, Disgusted, or Afraid? Studies on the Malleability of Emotion Perception: Vol. Darwin*. Ekman.

Aviezer, H., Trope, Y., Todorov, A., & Wherry, E. J. (2012). Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. *Science*, *338*(6111), 1225–1228. https://doi.org/10.1126/science.1229620

Barrett, K. C., & Campos, J. J. (1987). *Perspectives on emotional development II: A functionalist approach to emotions.* (J. D. Osofsky, Ed.). John Wiley & Sons.

Barrett, L. F. (1998). Discrete Emotions or Dimensions? The Role of Valence Focus and Arousal Focus. *Cognition and Emotion*, *12*(4), 579–599. https://doi.org/10.1080/026999398379574

Barrett, L. F. (2004). Feelings or words? Understanding the content in self-report ratings of experienced emotion. *Journal of Personality and Social Psychology*, *87*(2), 266–281. https://doi.org/10.1037/0022-3514.87.2.266

Barrett, L. F. (2006). Solving the Emotion Paradox: Categorization and the Experience of Emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition & Emotion*, *23*(7), 1284–1306. https://doi.org/10.1080/02699930902985894

Barrett, L. F. (2012). Emotions are real. *Emotion*, *12*(3), 413–429. https://doi.org/10.1037/a0027555

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(11), 1833–1833. https://doi.org/10.1093/scan/nsx060

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., & Bliss-Moreau, E. (2009). *Chapter 4 Affect as a Psychological Primitive* (pp. 167–218). https://doi.org/10.1016/S0065-2601(08)00404-8

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, *11*(8), 327–332. https://doi.org/10.1016/j.tics.2007.06.003

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290. https://doi.org/10.1177/0963721411422522

Barrett, L. F., & Satpute, A. B. (2019). Historical pitfalls and new directions in the neuroscience of emotion. *Neuroscience Letters*, *693*, 9–18. https://doi.org/10.1016/j.neulet.2017.07.045

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227. https://doi.org/10.3758/BF03196968

Bastiaansen, J. A. C. J., Thioux, M., & Keysers, C. (2009). Evidence for mirror systems in emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2391–2404. https://doi.org/10.1098/rstb.2009.0058

Biele, C., & Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Experimental Brain Research*, *171*(1), 1–6. https://doi.org/10.1007/s00221-005-0254-0

Bould, E., Morris, N., & Wink, B. (2008). Recognising subtle emotional expressions: The role of facial movements. *Cognition and Emotion*, *22*(8), 1569–1587. https://doi.org/10.1080/02699930801921156

Braddick, O., & Atkinson, J. (2011). Development of human visual function. In *Vision Research* (Vol. 51, Issue 13, pp. 1588–1609). https://doi.org/10.1016/j.visres.2011.02.018

Brooks, J. A., Chikazoe, J., Sadato, N., & Freeman, J. B. (2019). The neural representation of facial-emotion categories reflects conceptual structure. *Proceedings of the National Academy of*

Sciences of the United States of America, *116*(32), 15861–15870. https://doi.org/10.1073/pnas.1816408116

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. In *Nature Human Behaviour* (Vol. 2, Issue 8, pp. 581–591). Nature Publishing Group. https://doi.org/10.1038/s41562-018-0376-6

Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. *Cognition & Emotion*, *24*(3), 377–400. https://doi.org/10.1080/02699930902975754

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception.* Oxford University Press.

Buck, R., & Arthur Vanlear, C. (2002). Verbal and Nonverbal Communication: Distinguishing Symbolic, Spontaneous, and Pseudo-Spontaneous Nonverbal Behavior. In *Journal of Communication*.

Butler, A., Oruc, I., Fox, C. J., & Barton, J. J. S. (2008). Factors contributing to the adaptation aftereffects of facial expression. *Brain Research*, *1191*, 116–126. https://doi.org/10.1016/j.brainres.2007.10.101

Calvo, M. G., & Beltrán, D. (2013). Recognition advantage of happy faces: Tracing the neurocognitive processes. *Neuropsychologia*, *51*(11), 2051–2061. https://doi.org/10.1016/j.neuropsychologia.2013.07.010

Calvo, M. G., Fernández-Martín, A., & Nummenmaa, L. (2012). Perceptual, categorical, and affective processing of ambiguous smiling facial expressions. *Cognition*, *125*(3), 373–393. https://doi.org/10.1016/j.cognition.2012.07.021

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, *40*(1), 109–115. https://doi.org/10.3758/BRM.40.1.109

Carnevali, L., Gui, A., Jones, E. J. H., & Farroni, T. (2022). Face Processing in Early Development: A Systematic Review of Behavioral Studies and Considerations in Times of COVID-19 Pandemic. In *Frontiers in Psychology* (Vol. 13). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2022.778247

Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, *70*(2), 205–218. https://doi.org/10.1037/0022-3514.70.2.205

Carroll, J. M., & Russell, J. A. (1997). Facial Expressions in Hollywood’s Portrayal of Emotion. In *Journal of Personality and Social Psychology* (Vol. 72, Issue I).

Clore, G. L., & Ortony, A. (2013). Psychological construction in the OCC model of emotion. *Emotion Review*, *5*(4), 335–343. https://doi.org/10.1177/1754073913489751

Cohen, H., & Lefebvre, C. (2005). *Handbook of Categorization in Cognitive Science*. Elsevier. https://doi.org/10.1016/B978-0-08-044612-7.X5053-7

Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion*, *16*(1), 117–128. https://doi.org/10.1037/emo0000100

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion*, *18*(1), 75–93. https://doi.org/10.1037/emo0000302

Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 Emotions Conveyed by Brief Human Vocalization. *American Psychologist*. https://doi.org/10.1037/amp0000399.supp

Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). *What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures*. *117*, 1924–1934. https://doi.org/10.1073/pnas.1910704117/-/DCSupplemental

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(38), E7900–E7909. https://doi.org/10.1073/pnas.1702247114

Cowen, A. S., & Keltner, D. (2018). Clarifying the Conceptualization, Dimensionality, and Structure of Emotion: Response to Barrett and Colleagues. In *Trends in Cognitive Sciences* (Vol. 22, Issue 4, pp. 274–276). Elsevier Ltd. https://doi.org/10.1016/j.tics.2018.02.003

Cowen, A. S., & Keltner, D. (2020). Semantic Space Theory: A Computational Approach to Emotion. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2020.11.004

Cowen, A., Sauter, D., Tracy, J. L., & Keltner, D. (2019). Mapping the Passions: Toward a High-Dimensional Taxonomy of Emotional Experience and Expression. *Psychological Science in the Public Interest*, *20*(1), 69–90. https://doi.org/10.1177/1529100619850176

Crivelli, C., & Fridlund, A. J. (2019). Inside-Out: From Basic Emotions Theory to the Behavioral Ecology View. *Journal of Nonverbal Behavior*, *43*(2), 161–194. https://doi.org/10.1007/s10919-019-00294-2

Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J. M. (2017). Recognizing spontaneous facial expressions of emotion in a small-scale society of papua New Guinea. *Emotion*, *17*(2), 337–347. https://doi.org/10.1037/emo0000236

Cutzu, F., & Edelman, S. (1998). Representation of object similarity in human vision: psychophysics and a computational model. In *Vision Research* (Vol. 38).

Cutzu, F., & Edelmant, S. (1996). *Faithful representation of similarities among three-dimensional shapes in human vision (object recognition/multidimensional scaling/computational model)* (Vol. 93). https://www.pnas.org

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion*, *12*(5), 1085–1101. https://doi.org/10.1037/a0025737

Darwin, C. (1897). *The expression of the emotions in man and animals* (D. Appleton and Company, Ed.).

Davidoff, J. (2001). Language and perceptual categorisation. *Trends in Cognitive Sciences*, *5*(9), 382–387. https://doi.org/10.1016/S1364-6613(00)01726-5

Del Líbano, M., Calvo, M. G., Fernández-Martín, A., & Recio, G. (2018). Discrimination between smiling faces: Human observers vs. automated face analysis. *Acta Psychologica*, *187*, 19–29. https://doi.org/10.1016/j.actpsy.2018.04.019

Delannoy, J. R., & McDonald, J. B. (2009). Estimation of the temporal dynamics of posed and spontaneous facial expression formation using LLE. *IMVIP 2009 - 2009 International Machine Vision and Image Processing Conference*, 139–144. https://doi.org/10.1109/IMVIP.2009.32

DiGirolamo, M. A., & Russell, J. A. (2017). The emotion seen in a face can be a methodological artifact: The process of elimination hypothesis. *Emotion*, *17*(3), 538–546. https://doi.org/10.1037/emo0000247

Dobs, K., Bülthoff, I., Breidt, M., Vuong, Q. C., Curio, C., & Schultz, J. (2014). Quantifying human sensitivity to spatio-temporal information in dynamic faces. *Vision Research*, *100*, 78–87. https://doi.org/10.1016/j.visres.2014.04.009

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(15). https://doi.org/10.1073/pnas.1322355111

Ekman, P. (1972). *Universal and Cultural Differences in Facial Expression of Emotions.* (J. Cole). University of Nebraska Press.

Ekman, P. (1973). *Cross-Cultural Studies of Facial Expression.* (P. Ekman). Academic Press.

Ekman, P. (1992). An argument for basic emotions. In *Cognition and Emotion* (pp. 169–200).

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Ekman, P. E., & Davidson, R. J. (1994). *The Nature of Emotion: Fundamental Questions.* (Oxford University Press, Ed.).

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. In *Journal ol Personality and Social Psychology* (Vol. 71, Issue 2).

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face.* NJ: Prentice-Hall.

Ekman, P., & Friesen, W. V. (1976). Measuring Facial Movement. *Environmental Psychology and Nonverbal Behavior*.

Ekman, P., & Friesen, W. V. (1978). *Manual for the Facial Action Code.* (Consulting Psychologist Press, Ed.).

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System. Manual and Investigator's Guide.* Research Nexus.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86–88. https://doi.org/10.1126/science.164.3875.86

Elfenbein, H. A. (2013). Nonverbal dialects and accents in facial expressions of emotion. In *Emotion Review* (Vol. 5, Issue 1, pp. 90–96). SAGE Publications Ltd. https://doi.org/10.1177/1754073912451332

Elfenbein, H. A. (2015). In-Group Advantage and Other-Group Bias in Facial Emotion Recognition. In *Understanding Facial Expressions in Communication* (pp. 57–71). Springer India. https://doi.org/10.1007/978-81-322-1934-7_4

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Elfenbein, H. A., & Ambady, N. (2003). Universals and Cultural Differences in Recognizing Emotions. *Current Directions in Psychological Science*, *12*(5), 159–164. https://doi.org/10.1111/1467-8721.01252

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion*, *7*(1), 131–146. https://doi.org/10.1037/1528-3542.7.1.131

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*(3), 227–240. https://doi.org/10.1016/0010-0277(92)90002-Y

Fang, X., Sauter, D. A., & Van Kleef, G. A. (2018). Seeing Mixed Emotions: The Specificity of Emotion Perception From Static and Dynamic Facial Expressions Across Cultures. *Journal of Cross-Cultural Psychology*, *49*(1), 130–148. https://doi.org/10.1177/0022022117736270

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review*, *105*(3), 482–498. https://doi.org/10.1037/0033-295X.105.3.482

Farran, E. K., Mares, I., Papasavva, M., Smith, F. W., Ewing, L., & Smith, M. L. (2020). Characterizing the neural signature of face processing in Williams syndrome via multivariate pattern analysis and event related potentials. *Neuropsychologia*, *142*, 107440. https://doi.org/10.1016/j.neuropsychologia.2020.107440

Farroni, T., Menon, E., Rigato, S., & Johnson, M. H. (2007). The perception of facial expressions in newborns. *European Journal of Developmental Psychology*, *4*(1), 2–13. https://doi.org/10.1080/17405620601046832

Fei, Z., Yang, E., Li, D. D. U., Butler, S., Ijomah, W., Li, X., & Zhou, H. (2020). Deep convolution network based emotion analysis towards mental health care. *Neurocomputing*, *388*, 212–227. https://doi.org/10.1016/j.neucom.2020.01.034

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*(4), 967–984. https://doi.org/10.1037/0022-3514.74.4.967

Feldman, L. A. (1995). Valence Focus and Arousal Focus: Individual Differences in the Structure of Affective Experience. In *Journal of Personality and Social Psychology* (Vol. 69, Issue 1).

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *Journal of Vision*, *11*(3), 17–17. https://doi.org/10.1167/11.3.17

Fox, R., & McDaniel, C. (1982). The perception of biological motion by human infants. *Science*, *218*(4571), 486–487. https://doi.org/10.1126/science.7123249

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants’ attention to faces during the first year. *Cognition*, *110*(2), 160–170. https://doi.org/10.1016/j.cognition.2008.11.010

Friedman, H. S. (1978). The Relative Strength of Verbal versus Nonverbal Cues. *Personality and Social Psychology Bulletin*, *4*(1).

Frijda, N. H., & Philipszoon, Els. (1963). Dimensions of recognition of expression. *The Journal of Abnormal and Social Psychology*, *66*(1), 45–51. https://doi.org/10.1037/h0042578

Gendron, M., Crivelli, C., & Barrett, L. F. (2018). Universality Reconsidered: Diversity in Making Meaning of Facial Expressions. *Current Directions in Psychological Science*, *27*(4), 211–219. https://doi.org/10.1177/0963721417746794

Gendron, M., Hoemann, K., Crittenden, A. N., Mangola, S. M., Ruark, G. A., & Barrett, L. F. (2020). Emotion Perception in Hadza Hunter-Gatherers. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-60257-2

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska directed emotional faces: A validation study. *Cognition and Emotion*, *22*(6), 1094–1118. https://doi.org/10.1080/02699930701626582

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., Goode, R. A., Lee, D., Simmons, M., & Srinath, A. (2013). The efficiency of dynamic and static facial expression recognition. *Journal of Vision*, *13*(5), 23–23. https://doi.org/10.1167/13.5.23

Graham, R., & LaBar, K. S. (2007). Garner interference reveals dependencies between emotional expression and gaze in face perception. *Emotion*, *7*(2), 296–313. https://doi.org/10.1037/1528-3542.7.2.296

Greenaway, K. H., Kalokerinos, E. K., & Williams, L. A. (2018). Context is Everything (in Emotion Research). *Social and Personality Psychology Compass*, *12*(6). https://doi.org/10.1111/spc3.12393

Grewe, O., Kopiez, R., & Altenmüller, E. (2009). The Chill Parameter: Goose Bumps and Shivers as Promising Measures in Emotion Research. *Music Perception*, *27*(1), 61–74. https://doi.org/10.1525/mp.2009.27.1.61

Grossmann, I., Huynh, A. C., & Ellsworth, P. C. (2016). Emotional complexity: Clarifying definitions and cultural correlates. *Journal of Personality and Social Psychology*, *111*(6), 895–916. https://doi.org/10.1037/pspp0000084

Hager, J. C., & Ekman, P. (1979). *Long-Distance Transmission of Facial Affect Signals*.

Haidt, J., & Keltner, D. (1999). Culture and Facial Expression: Open-ended Methods Find More Expressions and a Gradient of Recognition. *Cognition & Emotion*, *13*(3), 225–266. https://doi.org/10.1080/026999399379267

Harmon-Jones, E., & Sigelman, J. (2001). State anger and prefrontal brain activity: Evidence that insult-related relative left-prefrontal activation is associated with experienced anger and aggression. *Journal of Personality and Social Psychology*, *80*(5), 797–803. https://doi.org/10.1037/0022-3514.80.5.797

Harris, R. J., Young, A. W., & Andrews, T. J. (2012). Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain. *Proceedings of the National Academy of Sciences*, *109*(51), 21164–21169. https://doi.org/10.1073/pnas.1212207110

Harry, B., Williams, M. A., Davis, C., & Kim, J. (2013). Emotional expressions evoke a differential response in the fusiform face area. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00692

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

Herrando, C., & Constantinides, E. (2021). Emotional Contagion: A Brief Overview and Future Directions. In *Frontiers in Psychology* (Vol. 12). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2021.712606

Hertenstein, M. J., Holmes, R., McCullough, M., & Keltner, D. (2009). The communication of emotion via touch. *Emotion*, *9*(4), 566–573. https://doi.org/10.1037/a0016108

Holmes, M. (2011). National security behavioral detection: A typography of strategies, costs, and benefits. In *Journal of Transportation Security* (Vol. 4, Issue 4, pp. 361–374). Springer New York LLC. https://doi.org/10.1007/s12198-011-0076-1

Hughes, S. M., & Nicholson, S. E. (2008). Sex differences in the assessment of pain versus sexual pleasure facial expressions. . *Journal of Social, Evolutionary, and Cultural Psychology*.

Izard, C. E. (2011). Forms and functions of emotions: Matters of emotion-cognition interactions. *Emotion Review*, *3*(4), 371–378. https://doi.org/10.1177/1754073911410737

Jack, R. E., Crivelli, C., & Wheatley, T. (2018). Data-Driven Methods to Diversify Knowledge of Human Psychology. *Trends in Cognitive Sciences*.

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, *24*(2), 187–192. https://doi.org/10.1016/j.cub.2013.11.064

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7241–7244. https://doi.org/10.1073/pnas.1200155109

Jack, R. E., & Schyns, P. G. (2017). Toward a Social Psychophysics of Face Communication. In *Annual Review of Psychology* (Vol. 68, pp. 269–297). Annual Reviews Inc. https://doi.org/10.1146/annurev-psych-010416-044242

Jack, R. E., Sun, W., Delis, I., Garrod, O. G. B., & Schyns, P. G. (2016). Four Not Six: Revealing Culturally Common Facial Expressions of Emotion. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000162.supp

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). *Emotion semantics show both cultural variation and universal structure*.

Johnson, M. H., Dziurawiec, S., Ellis, H., Morton, J., Green, A., Hopson Tara Keenan Julie Phillips, P., & Bartrip, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline*. In *Cognition* (Vol. 40).

Johnson-Laird, P. N., & Oatley, K. (1998). Basic emotions, rationality, and folk theory. In *Consciousness and Emotion in Cognitive Science* (pp. 289–311). Routledge.

Johnston, L., Miles, L., & Macrae, C. N. (2010). Why are you smiling at me? Social functions of enjoyment and non-enjoyment smiles. *British Journal of Social Psychology*, *49*(1), 107–127. https://doi.org/10.1348/014466609X412476

Kaulard, K., Cunningham, D. W., Bülthoff, H. H., & Wallraven, C. (2012). The MPI facial expression database - a validated database of emotional and conversational facial expressions. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0032321

Kaushik, H., Kumar, T., & Bhalla, K. (2022). iSecureHome: A deep fusion framework for surveillance of smart homes using real-time emotion recognition. *Applied Soft Computing*, *122*, 108788. https://doi.org/10.1016/j.asoc.2022.108788

Keltner, D., Brooks, J. A., & Cowen, A. (2023). Semantic Space Theory: Data-Driven Insights Into Basic Emotions. *Current Directions in Psychological Science*. https://doi.org/10.1177/09637214221150511

Keltner, D., & Ekman, P. (2000). Emotion: An overview. In *Encyclopedia of Psychology* (Vol. 3, pp. 162–167). Oxford University Press.

Keltner, D., Ellsworth, P. C., & Edwards, K. (1993). Beyond simple pessimism: Effects of sadness and anger on social perception. *Journal of Personality and Social Psychology*, *64*(5), 740–752. https://doi.org/10.1037/0022-3514.64.5.740

Keltner, D., & Gross, J. J. (1999). Functional Accounts of Emotions. *Cognition and Emotion*, *13*(5), 467–480.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, *13*(5), 505–521. https://doi.org/10.1080/026999399379168

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional Expression: Advances in Basic Emotion Theory. In *Journal of Nonverbal Behavior* (Vol. 43, Issue 2, pp. 133–160). Springer New York LLC. https://doi.org/10.1007/s10919-019-00293-3

Keltner, D., Tracy, J. L., Sauter, D., & Cowen, A. (2019). What Basic Emotion Theory Really Says for the Twenty-First Century Study of Emotion. In *Journal of Nonverbal Behavior* (Vol. 43, Issue 2, pp. 195–201). Springer New York LLC. https://doi.org/10.1007/s10919-019-00298-y

Keltner, D., Tracy, J., Sauter, D. A., Cordaro, D. C., & McNeil, G. (2016). *Expression of Emotion*.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309. https://doi.org/10.1152/jn.00024.2007

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., & Whalen, P. J. (2004). Contextual Modulation of Amygdala Responsivity to Surprised Faces. *Journal of Cognitive Neuroscience*, *16*(10), 1730–1745. https://doi.org/10.1162/0898929042947865

Kitayama, S., Mesquita, B., & Karasawa, M. (2006). Cultural affordances and emotional experience: Socially engaging and disengaging emotions in Japan and the United States. *Journal of Personality and Social Psychology*, *91*(5), 890–903. https://doi.org/10.1037/0022-3514.91.5.890

Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors (Switzerland)*, *18*(2). https://doi.org/10.3390/s18020401

Krautheim, J. T., Steines, M., Dannlowski, U., Neziroğlu, G., Acosta, H., Sommer, J., Straube, B., & Kircher, T. (2020). Emotion specific neural activation for the production and perception of facial expressions. *Cortex*, *127*, 17–28. https://doi.org/10.1016/j.cortex.2020.01.026

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. In *Emotion Review* (Vol. 5, Issue 1, pp. 41–46). https://doi.org/10.1177/1754073912451349

Krumhuber, E. G., Küster, D., Namba, S., Shah, D., & Calvo, M. G. (2021). *Emotion recognition from posed and spontaneous dynamic expressions: Human observers vs. machine analysis*.

Krumhuber, E. G., Skora, L., Küster, D., & Fou, L. (2017). A Review of Dynamic Datasets for Facial Expression Research. In *Emotion Review* (Vol. 9, Issue 3, pp. 280–292). SAGE Publications Ltd. https://doi.org/10.1177/1754073916670022

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, *42*(3), 300–311. https://doi.org/10.1109/12.210173

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: brain mechanisms and psychophysiology. *Biological Psychiatry*, *44*(12), 1248–1263. https://doi.org/10.1016/S0006-3223(98)00275-3

Latané, B., & Darley, J. M. (1968). Group Inhibition of bystander intervention in emergencies. *Journal at Personality and Social Psycholoty*, *10*(3), 215–221.

Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W., Iraki, F. K., Rockstuhl, T., & Thingujam, N. S. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00353

Laukka, P., Neiberg, D., & Elfenbein, H. A. (2014). Evidence for cultural dialects in vocal emotion expression: Acoustic classification within and across five nations. *Emotion*, *14*(3), 445–449. https://doi.org/10.1037/a0036048

Lazarus, R., & Folkman, S. (1984). *Stress, Appraisal, and Coping.* Springer.

Lazarus, R. S. (1991). *Cognition and motivation in emotion* (46th ed.). American psychologist.

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological Research*, *69*(1–2), 22–29. https://doi.org/10.1007/s00426-003-0157-2

Leppänen, J. M., & Hietanen, J. K. (2007). Is there more in a happy face than just a big smile? *Visual Cognition*, *15*(4), 468–490. https://doi.org/10.1080/13506280600765333

Levenson, R. W. (1994). Human Emotions: A Functional View. In P. Ekman & R. . J. Davidson (Eds.), *The Nature of Emotion: Fundamental Questions* (pp. 123–126). Oxford University.

Levenson, R. W. (2011). Basic emotion questions. *Emotion Review*, *3*(4), 379–386. https://doi.org/10.1177/1754073911410743

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, *7*(4), 227–236. https://doi.org/10.1037/npe0000028

Lindquist, K. A., & Barrett, L. F. (2008). Constructing Emotion. *Psychological Science*, *19*(9), 898–903. https://doi.org/10.1111/j.1467-9280.2008.02174.x

Lipp, O. V., Price, S. M., & Tellegen, C. L. (2009). No Effect of Inversion on Attentional and Affective Processing of Facial Expressions. *Emotion*, *9*(2), 248–259. https://doi.org/10.1037/a0014715

Liu, M., Duan, Y., Ince, R. A. A., Chen, C., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2022). Facial expressions elicit multiplexed perceptions of emotion categories and dimensions. *Current Biology*, *32*(1), 200-209.e6. https://doi.org/10.1016/j.cub.2021.10.035

Lobmaier, J. S., Tiddeman, B. P., & Perrett, D. I. (2008). Emotional expression modulates perceived gaze direction. *Emotion*, *8*(4), 573–577. https://doi.org/10.1037/1528-3542.8.4.573

Lundqvist, D., Flykt, A., & Öhman, A. (1998). Karolinska Directed Emotional Faces (KDEF). In *APA PsycTests.*

Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). *Coding Facial Expressions with Gabor Wavelets*.

Magnussen, S., Sunde, B., & Dyrnes, S. (1994). Patterns of Perceptual Asymmetry in Processing Facial Expression. *Cortex*, *30*(2), 215–229. https://doi.org/10.1016/S0010-9452(13)80194-3

Mares, I., Ewing, L., Farran, E. K., Smith, F. W., & Smith, M. L. (2020). Developmental changes in the processing of faces as revealed by EEG decoding. *NeuroImage*, *211*, 116660. https://doi.org/10.1016/j.neuroimage.2020.116660

Margalit, E., Biederman, I., Herald, S. B., Yue, X., & von der Malsburg, C. (2016). An applet for the Gabor similarity scaling of the differences between complex stimuli. *Attention, Perception, and Psychophysics*, *78*(8), 2298–2306. https://doi.org/10.3758/s13414-016-1191-7

Margoudi, M., Hart, J., Adams, A., & Oliveira, M. (2016). Exploring emotion representation to support dialogue in police training on child interviewing. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9894 LNCS*, 73–86. https://doi.org/10.1007/978-3-319-45841-0_7

Markus, H. R. , K. S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224–253.

Markus, H. R. , M. P. R. , & K. S. (1997). *Selfways: Diversity in modes of cultural participation. The conceptual self in context: Culture, experience, self-understanding.* . Cambridge University Press.

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, *94*(3), 365–381. https://doi.org/10.1037/0022-3514.94.3.365

Matsumoto, D. (1993). Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motivation and Emotion*, *17*(2), 107–123. https://doi.org/10.1007/BF00995188

Matsumoto, D. (2006). Are Cultural Differences in Emotion Regulation Mediated by Personality Traits? *Journal of Cross-Cultural Psychology*, *37*(4), 421–437. https://doi.org/10.1177/0022022106288478

Matsumoto, D., Frank, M., & Hwang, H. (2013). *Nonverbal Communication: Science and Applications*. SAGE Publications, Inc. https://doi.org/10.4135/9781452244037

Matsumoto, D., & Hwang, H. C. (2013). Assessing Cross-Cultural Competence. *Journal of Cross-Cultural Psychology*, *44*(6), 849–873. https://doi.org/10.1177/0022022113492891

Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. In *Handbook of emotions* (pp. 211–234). The Guilford Press.

Matsumoto, D., Seung Hee Yoo, & Fontaine, J. (2008). Mapping Expressive Differences Around the World. *Journal of Cross-Cultural Psychology*, *39*(1), 55–74. https://doi.org/10.1177/0022022107311854

Matsumoto, D., Yoo, S. H., Fontaine, J., Anguas-Wong, A. M., Arriola, M., Ataca, B., Bond, M. H., Boratav, H. B., Breugelmans, S. M., Cabecinhas, R., Chae, J., Chin, W. H., Comunian, A. L., Degere, D. N., Djunaidi, A., Fok, H. K., Friedlmeier, W., Ghosh, A., Glamcevski, M., … Grossi, E. (2008). Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology*, *39*(1), 55–74. https://doi.org/10.1177/0022022107311854

Mehrabian, A. (1971). Silent Messages: Implicit Communication of Emotions and Attitudes. *Wadsworth Publishing.*

Mende-Siedlecki, P., Qu-Lee, J., Lin, J., Drain, A., & Goharzad, A. (2020). The Delaware Pain Database: A set of painful expressions and corresponding norming data. *Pain Reports*, *5*(6). https://doi.org/10.1097/PR9.0000000000000853

Miyamoto, Y., Uchida, Y., & Ellsworth, P. C. (2010). Culture and mixed emotions: Co-occurrence of positive and negative emotions in Japan and the United States. *Emotion*, *10*(3), 404–415. https://doi.org/10.1037/a0018430

Mobbs, D., Weiskopf, N., Lau, H. C., Featherstone, E., Dolan, R. J., & Frith, C. D. (2006). The Kuleshov Effect: the influence of contextual framing on emotional attributions. *Social Cognitive and Affective Neuroscience*, *1*(2), 95–106. https://doi.org/10.1093/scan/nsl014

Moore, M. M., & Martin, E. A. (2022). Taking Stock and Moving Forward: A Personalized Perspective on Mixed Emotions. *Perspectives on Psychological Science*, *17*(5), 1258–1275. https://doi.org/10.1177/17456916211054785

Mortillaro, M., Mehu, M., & Scherer, K. R. (2011). Subtly different positive emotions can be distinguished by their facial expressions. *Social Psychological and Personality Science*, *2*(3), 262–271. https://doi.org/10.1177/1948550610389080

Mumenthaler, C., & Sander, D. (2012). Social appraisal influences recognition of emotions. *Journal of Personality and Social Psychology*, *102*(6), 1118–1135. https://doi.org/10.1037/a0026885

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316. https://doi.org/10.1037/0033-295X.92.3.289

Murray, T., O'Brien, J., Sagiv, N., & Garrido, L. (2021). The role of stimulus-based cues and conceptual information in processing facial expressions of emotion. *Cortex*, *144*, 109–132. https://doi.org/10.1016/j.cortex.2021.08.007

Muukkonen, I., & Salmela, V. R. (2022). Representational structure of fMRI/EEG responses to dynamic facial expressions. *NeuroImage*, *263*, 119631. https://doi.org/10.1016/j.neuroimage.2022.119631

Nakajima, K., Minami, T., & Nakauchi, S. (2017). Interaction between facial expression and color. *Scientific Reports*, *7*. https://doi.org/10.1038/srep41019

Namba, S., Makihara, S., Kabir, R. S., Miyatani, M., & Nakao, T. (2017). Spontaneous Facial Expressions Are Different from Posed Facial Expressions: Morphological Properties and Dynamic Sequences. *Current Psychology*, *36*(3), 593–605. https://doi.org/10.1007/s12144-016-9448-9

Nelson, C. A. (2001). The Development and Neural Bases of Face Recognition. *Infant and Child Development*, *10*(1–2), 3–18. https://doi.org/10.1002/icd.239

Nelson, N. L., & Russell, J. A. (2013). Universality revisited. In *Emotion Review* (Vol. 5, Issue 1, pp. 8–15). SAGE Publications Ltd. https://doi.org/10.1177/1754073912457227

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently and why* (New York: Free Press).

Oatley, K., & Jenkins, J. M. (1992). Human Emotions: Function and Dysfunction. *Annual Review of Psychology*, *43*(1), 55–85. https://doi.org/10.1146/annurev.ps.43.020192.000415

Oosterhof, N. N., & Todorov, A. (2008). *The functional basis of face evaluation*. www.pnas.orgcgidoi10.1073pnas.0805664105

Osgood, C. E. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology*, *7*(1), 1–30. https://doi.org/10.1111/j.1467-9450.1966.tb01334.x

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. In *Behavior Research Methods, Instruments, & Computers* (Vol. 36, Issue 4). www.psychonomic.org/archive/.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, *5*(4), 291–303. https://doi.org/10.1038/nrn1364

Panksepp, J., & Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review*, *3*(4), 387–396. https://doi.org/10.1177/1754073911410741

Pantic, M., & Bartlett, M. (2007). *Machine Analysis of Facial Expressions* (pp. 377–416). I-Tech Education and Publishing.

Pantic, M., & Stewart, M. (2007). Machine Analysis of Facial Expressions. In *Face Recognition*. I-Tech Education and Publishing. https://doi.org/10.5772/4847

Park, S., Lee, K., Lim, J. A., Ko, H., Kim, T., Lee, J. I., Kim, H., Han, S. J., Kim, J. S., Park, S., Lee, J. Y., & Lee, E. C. (2020). Differences in facial expressions between spontaneous and posed smiles: Automated method by action units and three-dimensional facial landmarks. *Sensors (Switzerland)*, *20*(4). https://doi.org/10.3390/s20041199

Parkinson, B., & Simons, G. (2009). Affecting others: Social appraisal and emotion contagion in everyday decision making. *Personality and Social Psychology Bulletin*, *35*(8), 1071–1084. https://doi.org/10.1177/0146167209336611

Parr, L. A., Waller, B. M., & Vick, S. J. (2007). New Developments in Understanding Emotional Facial Signals in Chimpanzees. *Current Directions in Psychological Science*, *16*(3), 117–122. https://doi.org/10.1111/j.1467-8721.2007.00487.x

Payton, J. W., Wardlaw, D. M., Graczyk, P. A., Bloodworth, M. R., Tompsett, C. J., & Weissberg, R. P. (2000). *Social and Emotional Learning: A Framework for Promoting Mental Health and Reducing Risk Behavior in Children and Youth*. www.CASEL.org

Prochazkova, E., & Kret, M. E. (2017). Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neuroscience & Biobehavioral Reviews*, *80*, 99–114. https://doi.org/10.1016/j.neubiorev.2017.05.013

Righart, R., & de Gelder, B. (2006). Context Influences Early Perceptual Analysis of Faces—An Electrophysiological Study. *Cerebral Cortex*, *16*(9), 1249–1257. https://doi.org/10.1093/cercor/bhj066

Righart, R., & de Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. *Social Cognitive and Affective Neuroscience*, *3*(3), 270–278. https://doi.org/10.1093/scan/nsn021

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233. https://doi.org/10.1037/0096-3445.104.3.192

Rosch, E. (1978). Principles of Categorisation. In *Cognition and Categorisation* (pp. 27–48). Lawrence Erlbaum.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rosenberg, E. L., & Ekman, P. (2020). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).* Oxford University Press.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition

Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (1991). Culture and the Categorization of Emotions. In *Psychological Bulletin* (Vol. 110, Issue 3).

Russell, J. A. (1994). Is There Universal Recognition of Emotion From Facial Expression? A Review of the Cross-Cultural Studies. In *Psychological Bulletin* (Vol. 115, Issue 1).

Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Russell, J. A., Bachorowski, J. A., & Fernández-Dols, J. M. (2003). Facial and Vocal Expressions of Emotion. In *Annual Review of Psychology* (Vol. 54, pp. 329–349). https://doi.org/10.1146/annurev.psych.54.101601.145102

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Russell, J. A., & Bullock, M. (1985). Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults. *Journal of Personality and Social Psychology*, *48*(5), 1290–1298. https://doi.org/10.1037/0022-3514.48.5.1290

Russell, J. A., & Carroll, J. M. (1999). On the Bipolarity of Positive and Negative Affect. In *Psychological Bulletin* (Vol. 125, Issue 1).

Russell, J. A., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. *Journal of Personality and Social Psychology*, *67*(2), 186–205. https://doi.org/10.1037/0022-3514.67.2.186

Russell, J. A., & Sato, K. (1995). Comparing Emotion Words between Languages. *Journal of Cross-Cultural Psychology*, *26*(4), 384–391. https://doi.org/10.1177/0022022195264004

Rzymski, C., Tresoldi, T., Greenhill, S. J., Wu, M.-S., Schweikhard, N. E., Koptjevskaja-Tamm, M., Gast, V., Bodt, T. A., Hantgan, A., Kaiping, G. A., Chang, S., Lai, Y., Morozova, N., Arjava, H., Hübler, N., Koile, E., Pepper, S., Proos, M., Van Epps, B., … List, J.-M. (2020). The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific Data*, *7*(1), 13. https://doi.org/10.1038/s41597-019-0341-x

Saarni, C. (1999). *DOCUMENT RESUME*.

Sabini, J., & Silver, M. (2005). Why emotion names and experiences don't neatly pair. In *Psychological Inquiry* (Vol. 16, Issue 1, pp. 1–10). https://doi.org/10.1207/s15327965pli1601_01

Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision*, *10*(5). https://doi.org/10.1167/10.5.11

Sander, D., Grandjean, D., Kaiser, S., Wehrle, T., & Scherer, K. R. (2007). Interaction effects of perceived gaze direction and dynamic facial expression: Evidence for appraisal theories of emotion. *European Journal of Cognitive Psychology*, *19*(3), 470–480. https://doi.org/10.1080/09541440600757426

Sauter, D. (2010). More than happy: The need for disentangling positive emotions. *Current Directions in Psychological Science*, *19*(1), 36–40. https://doi.org/10.1177/0963721409359290

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual Cues in Nonverbal Vocal Expressions of Emotion. *Quarterly Journal of Experimental Psychology*, *63*(11), 2251–2272. https://doi.org/10.1080/17470211003721642

Sauter, D. A., & Scott, S. K. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion*, *31*(3), 192–199. https://doi.org/10.1007/s11031-007-9065-x

Scarantino, A. (2017). How to Do Things with Emotional Expressions: The Theory of Affective Pragmatics. *Psychological Inquiry*, *28*(2–3), 165–185. https://doi.org/10.1080/1047840X.2017.1328951

Scherer, K. R. (1992). What does facial expression express? . In *International review of studies on emotion* (K. T. Strongman, Vol. 2, pp. 139–165). John Wiley & Sons.

Schlosberg, H. (1941). A comparison of five shaving creams by the method of constant stimuli. *Journal of Applied Psychology*, *25*(4), 401–407. https://doi.org/10.1037/h0056840

Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology*, *44*(4), 229–237. https://doi.org/10.1037/h0055778

Shao, J., & Qian, Y. (2019). Three convolutional neural network models for facial expression recognition in the wild. *Neurocomputing*, *355*, 82–92. https://doi.org/10.1016/j.neucom.2019.05.005

Shariff, A. F., & Tracy, J. L. (2009). Knowing Who's Boss: Implicit Perceptions of Status From the Nonverbal Expression of Pride. *Emotion*, *9*(5), 631–639. https://doi.org/10.1037/a0017089

Shaver, P., Schwartz, J., Kirson, D., & O'connor, G. (1987). *Emotion Knowledge: Further Exploration of a Prototype Approach*.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, *27*(3), 219–246. https://doi.org/10.1007/BF02289621

Shiota, M. N., Campos, B., Oveis, C., Hertenstein, M. J., Simon-Thomas, E., & Keltner, D. (2017). Beyond happiness: Building a science of discrete positive emotions. *American Psychologist*, *72*(7), 617–643. https://doi.org/10.1037/a0040456

Shuman, V., Clark-Polner, E., Meuleman, B., Sander, D., & Scherer, K. R. (2015). Emotion perception from a componential perspective. *Cognition and Emotion*, *31*(1), 47–56. https://doi.org/10.1080/02699931.2015.1075964

Sifre, R., Olson, L., Gillespie, S., Klin, A., Jones, W., & Shultz, S. (2018). A Longitudinal Investigation of Preferential Attention to Biological Motion in 2- to 24-Month-Old Infants. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-20808-0

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. In *PNAS* (Vol. 105, Issue 2).

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion*, *9*(6), 838–846. https://doi.org/10.1037/a0017810

Skerry, A. E., & Saxe, R. (2014). A common neural code for perceived and inferred emotion. *Journal of Neuroscience*, *34*(48), 15997–16008. https://doi.org/10.1523/JNEUROSCI.1676-14.2014

Snoek, L., Jack, R. E., Schyns, P. G., Garrod, O. G. B., Mittenbühler, M., Chen, C., Oosterwijk, S., & Scholte, H. S. (2023). *Testing, explaining, and exploring models of facial expressions of emotions*. https://www.science.org

Sormaz, M., Watson, D. M., Smith, W. A. P., Young, A. W., & Andrews, T. J. (2016). Modelling the perceptual similarity of facial expressions from image statistics and neural responses. *NeuroImage*, *129*, 64–71. https://doi.org/10.1016/j.neuroimage.2016.01.041

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, *50*(4), 1446–1460. https://doi.org/10.3758/s13428-017-0996-1

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience*, *11*(7), 843–850. https://doi.org/10.1038/nn.2138

Susskind, J. M., Littlewort, G., Bartlett, M. S., Movellan, J., & Anderson, A. K. (2007). Human and computer recognition of facial expressions of emotion. *Neuropsychologia*, *45*(1), 152–162. https://doi.org/10.1016/j.neuropsychologia.2006.05.001

Szameitat, D. P., Alter, K., Szameitat, A. J., Darwin, C. J., Wildgruber, D., Dietrich, S., & Sterr, A. (2009). Differentiation of Emotions in Laughter at the Behavioral Level. *Emotion*, *9*(3), 397–405. https://doi.org/10.1037/a0015692

Thayer, R. E. (1989). *The biopsychology of mood and arousal.* Oxford University Press.

Todorov, A. (2010). *Neural responses during the perception of face traits and expressions. A dissertation presented to the faculty of princeton university in candidacy for the degree of doctor of philosophy recommended for acceptance by the department of psychology*.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Tracy, J. L. (2014). An evolutionary approach to understanding distinct emotions. *Emotion Review*, *6*(4), 308–312. https://doi.org/10.1177/1754073914534478

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. In *Emotion Review* (Vol. 3, Issue 4, pp. 397–405). SAGE Publications Ltd. https://doi.org/10.1177/1754073911410747

Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in everyday life. *PLoS ONE*, *10*(12). https://doi.org/10.1371/journal.pone.0145450

Valstar, M., Gunes, H., & Pantic, M. (2007). *How to Distinguish Posed from Spontaneous Smiles using Geometric Features*. 387.

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science*, *18*(3), 184–188. https://doi.org/10.1111/j.1467-8721.2009.01633.x

Walle, E. A., Reschke, P. J., Camras, L. A., & Campos, J. J. (2017). Infant differential behavioral responding to discrete emotions. *Emotion*, *17*(7), 1078–1091. https://doi.org/10.1037/emo0000307

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, *76*(5), 820–838. https://doi.org/10.1037/0022-3514.76.5.820

Watson, M., Haviland, J., Greer, S., Davidson, J., & Bliss, J. (1999). Influence of psychological response on survival in breast cancer: a population-based cohort study. *The Lancet*, *354*(9187), 1331–1336. https://doi.org/10.1016/S0140-6736(98)11392-2

Wegrzyn, M., Bruckhaus, I., & Kissler, J. (2015). Categorical Perception of Fear and Anger Expressions in Whole, Masked and Composite Faces. *PLOS ONE*, *10*(8), e0134790. https://doi.org/10.1371/journal.pone.0134790

Whalen, P. J., Raila, H., Bennett, R., Mattek, A., Brown, A., Taylor, J., van Tieghem, M., Tanner, A., Miner, M., & Palmer, A. (2013). Neuroscience and Facial Expressions of Emotion: The Role of Amygdala–Prefrontal Interactions. *Emotion Review*, *5*(1), 78–83. https://doi.org/10.1177/1754073912457231

Wierzbicka, A. (1999). *Emotions across Languages and Cultures*. Cambridge University Press. https://doi.org/10.1017/CBO9780511521256

Wierzbicka, A. (2009). Language and Metalanguage: Key Issues in Emotion Research. *Emotion Review*, *1*(1), 3–14. https://doi.org/10.1177/1754073908097175

Wieser, M. J., & Brosch, T. (2012). Faces in context: A review and systematization of contextual influences on affective face processing. *Frontiers in Psychology*, *3*(NOV). https://doi.org/10.3389/fpsyg.2012.00471

Woodworth, R. (1938). *Experimental psychology*. Holt.

Xiao, N. G., Quinn, P. C., Liu, S., Ge, L., Pascalis, O., & Lee, K. (2015). Eye tracking reveals a crucial role for facial motion in recognition of faces by infants. *Developmental Psychology*, *51*(6), 744–757. https://doi.org/10.1037/dev0000019

Xu, X., & Biederman, I. (2010). Loci of the release from fMRI adaptation for changes in facial expression, identity, and viewpoint. *Journal of Vision*, *10*(14), 36–36. https://doi.org/10.1167/10.14.36

Yik, M., Widen, S. C., & Russell, J. A. (2013). The within-subjects design in the study of facial expressions. *Cognition and Emotion*, *27*(6), 1062–1072. https://doi.org/10.1080/02699931.2013.763769

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. https://doi.org/10.1037/h0027474

Yitzhak, N., Giladi, N., Gurevich, T., Messinger, D. S., Prince, E. B., Martin, K., & Aviezer, H. (2017). Gently does it: Humans outperform a software classifier in recognizing subtle, nonstereotypical facial expressions. *Emotion*, *17*(8), 1187–1198. https://doi.org/10.1037/emo0000287

Yoshikawa, S., & Sato, W. (2008). Dynamic facial expressions of emotion induce representational momentum. *Cognitive, Affective and Behavioral Neuroscience*, *8*(1), 25–31. https://doi.org/10.3758/CABN.8.1.25

Yrizarry, N., Matsumoto, D., & Wilson-Cohn, C. (1998). American–Japanese differences in multiscalar intensity ratings of universal facial expressions of emotion. *Motivation and Emotion*, *22*(4), 315–327. https://doi.org/10.1023/A:1021304407227

Yue, X., Biederman, I., Mangini, M. C., Malsburg, C. von der, & Amir, O. (2012). Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. *Vision Research*, *55*, 41–46. https://doi.org/10.1016/j.visres.2011.12.012

Appendix A

Chapter 2

Category	Intensity	Everyday scenario
Happy	Low	“You are lying on your couch after a delicious dinner”
Happy	High	“You have reached a goal and you are happy to have accomplished it”
Fear	Low	“After leaving your flat you realize you forgot to switch off the cooker”
Fear	High	“A monster appears suddenly”
Pain	Low	“You watch a TV transmission of your favourite sport event. One player has a serious accident. You can see bones sticking out of the player’s body”
Pain	High	“While doing sports you suddenly have an accident in which you twist one ankle and graze your knee.”

Figure A.1. **Emotional scenarios.** *Every-day emotional scenarios used in Kaulard et al. (2012) to prompt the facial emotions selected for our studies.*

Possible conditions for each actor

		Fear		Pain		Happy	
		I	II	I	II	I	II
Fear	I	S	W	B	B	B	B
	II	W	S	B	B	B	B
Pain	I	B	B	S	W	B	B
	II	B	B	W	S	B	B
Happy	I	B	B	B	B	S	W
	II	B	B	B	B	W	S

Total conditions

Same	Within	BetweenW	BetweenB	
6	3	6	6	
6	6	12	12	Inverted order
54	54	108	108	* Actors (9)
				216 total

40 **40** **40** **40**

160 trials
each for Dynamic and Static Task

Figure A.2. **Selection of pairs of stimuli for the similarity rating task.** *I selected 40 pairs of stimuli for each of 4 possible conditions out of 216 possible stimuli combinations. In the "Same" condition, faces with the same emotion and same intensity; in the "Within" condition, faces with the same emotion but with different intensity; in the "Between-Within" condition, faces with different emotions at the same level of intensity; finally, in the "Between-Between" condition, faces of different emotions at different level of intensity.*

Appendix B

Chapter 3

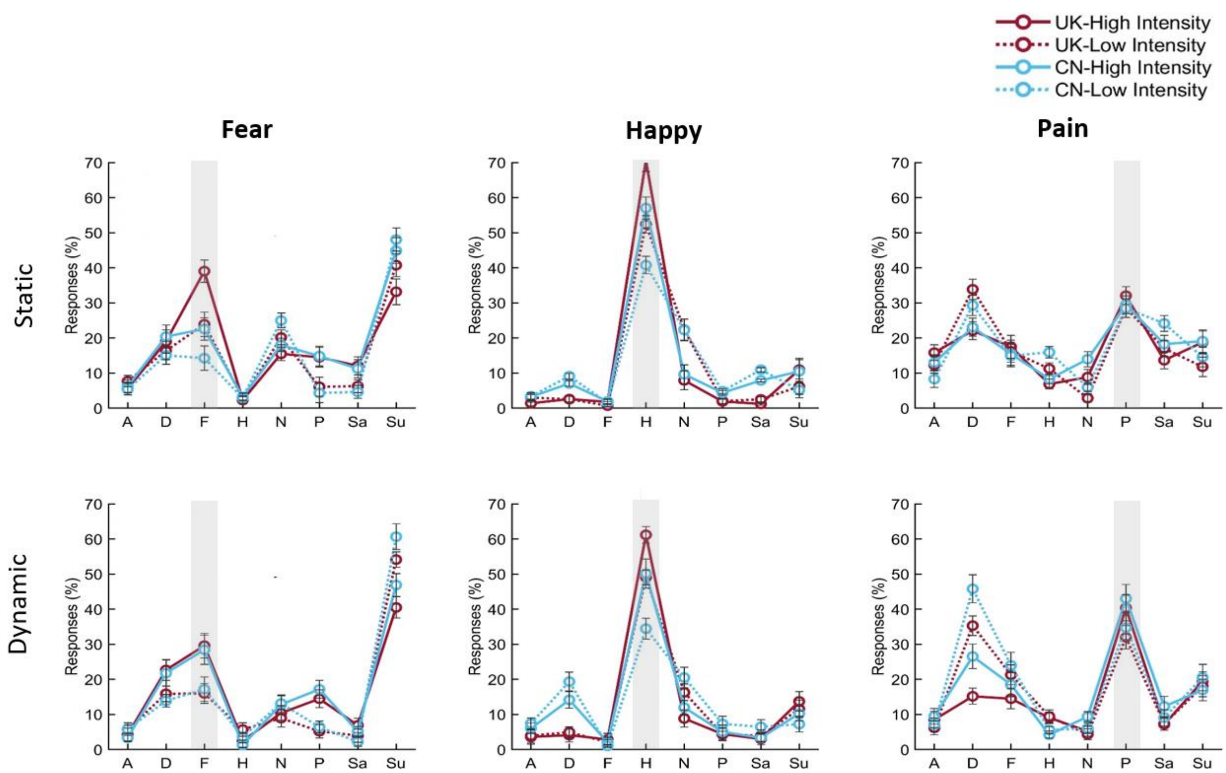

Figure B.1. **Emotion Profiles.** *Responses averaged across Chinese (in blue) and British (in red) participants to the Emotion profiling task for Static (top row) and Dynamic (bottom row) facial emotions of High-intensity (solid line) and Low intensity (dashed lines) Fear, Happy and Pain (columns).*

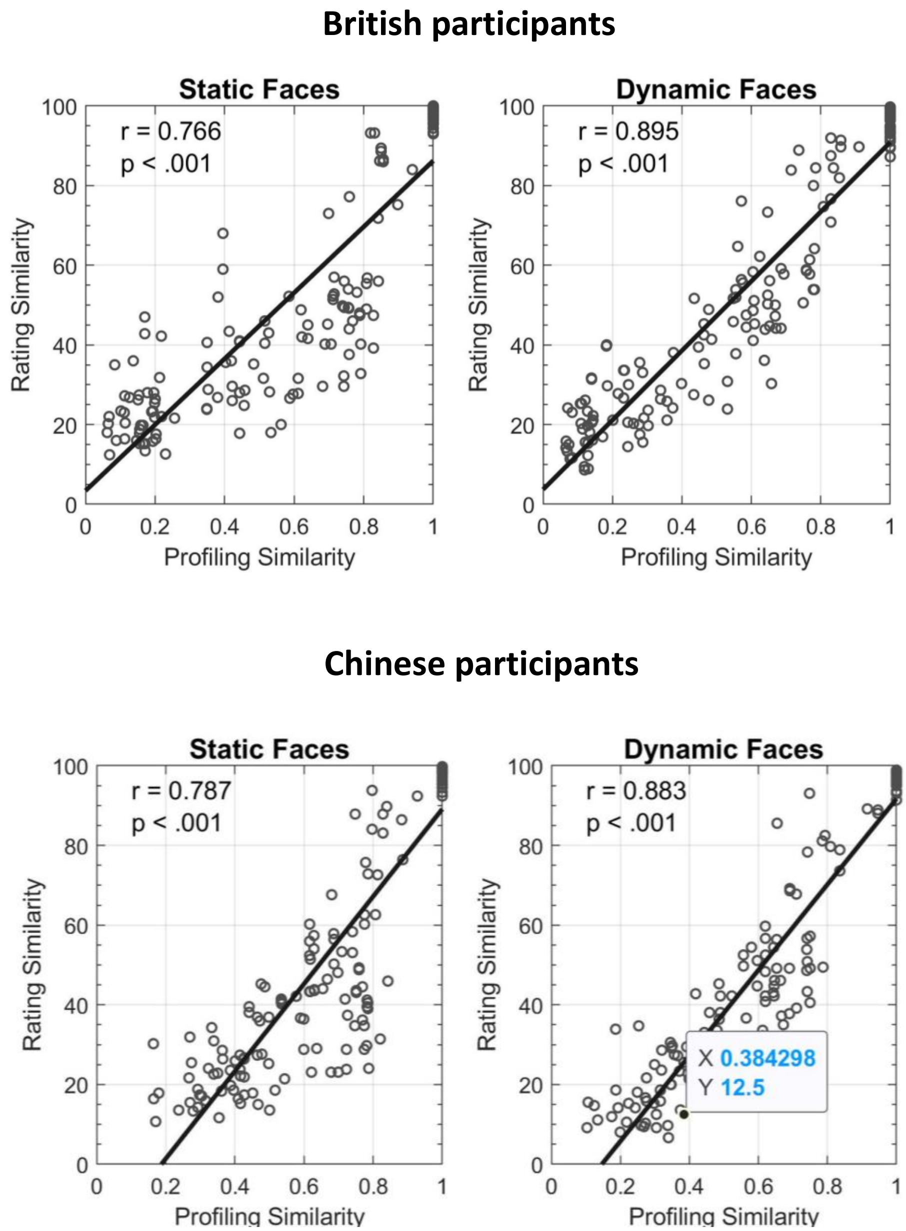

Figure B.2. **Correlation between Profiling similarity and participants similarity ratings.** *Each dot represents one of the 160 pairs of videos or images shown in the Similarity Rating task.*

<

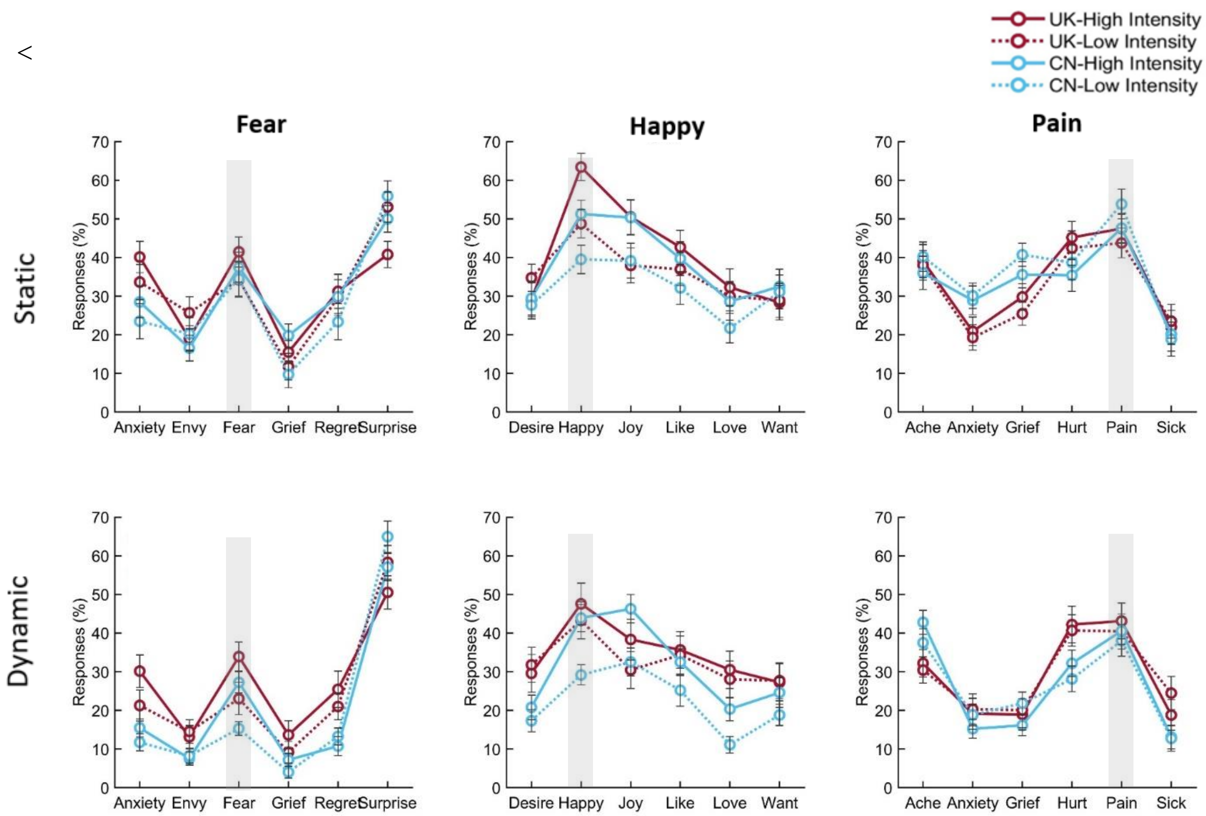

Figure B.3. **Semantic Profiles.** *Responses averaged across Chinese (in blue) and British (in red) participants to the Semantic profiling task for Static (top row) and Dynamic (bottom row) facial emotions of High-intensity (solid line) and Low intensity (dashed lines) Fear, Happy and Pain (columns).*

Appendix C

Chapter 4

Category	Intensity	Social scenario	Physical scenario
Happy	**High**	You won the first prize in a big competition (M=6.37; SD=.68)	You're eating your favourite food in your favourite place (M=5.5; SD=1.19)
	Low	You're spending time with your friends (M=5.85; SD=.98)	You are sunbathing on a beach (M=4.53; SD=1.47)
Fear	**High**	You just realised that there is an important exam tomorrow (M=5.35; SD=0.988)	You find a snake slithering into your sleeping bag (M=6.60; SD=0.94)
	Low	You're about to meet your partner's parents for the first time (M=3.50; SD=1.91)	You hear a strange sound while walking in the woods (M=4.80; SD=1.67)
Pain	**High**	You broke up with your long-time partner (M=6.05; SD=1.47)	You fall from your bike and break your arm (M=5; SD=1.50)
	Low	You are not invited to your best friend's party (M=5; SD=1.30)	You got a paper cut (M=2.95; SD=1.36)

Figure C.1. **Physical and Social scenarios.** *List of validated scenarios evoking emotions of happiness, pain, and fear. Each scenario would induce the target emotion either due to physical stimulation, or a socially meaningful event, at a high- lever or low- level of intensity. Values reported are the average intensity ratings (on a 7-points Likert scale) obtained from the pilot study N=20.*

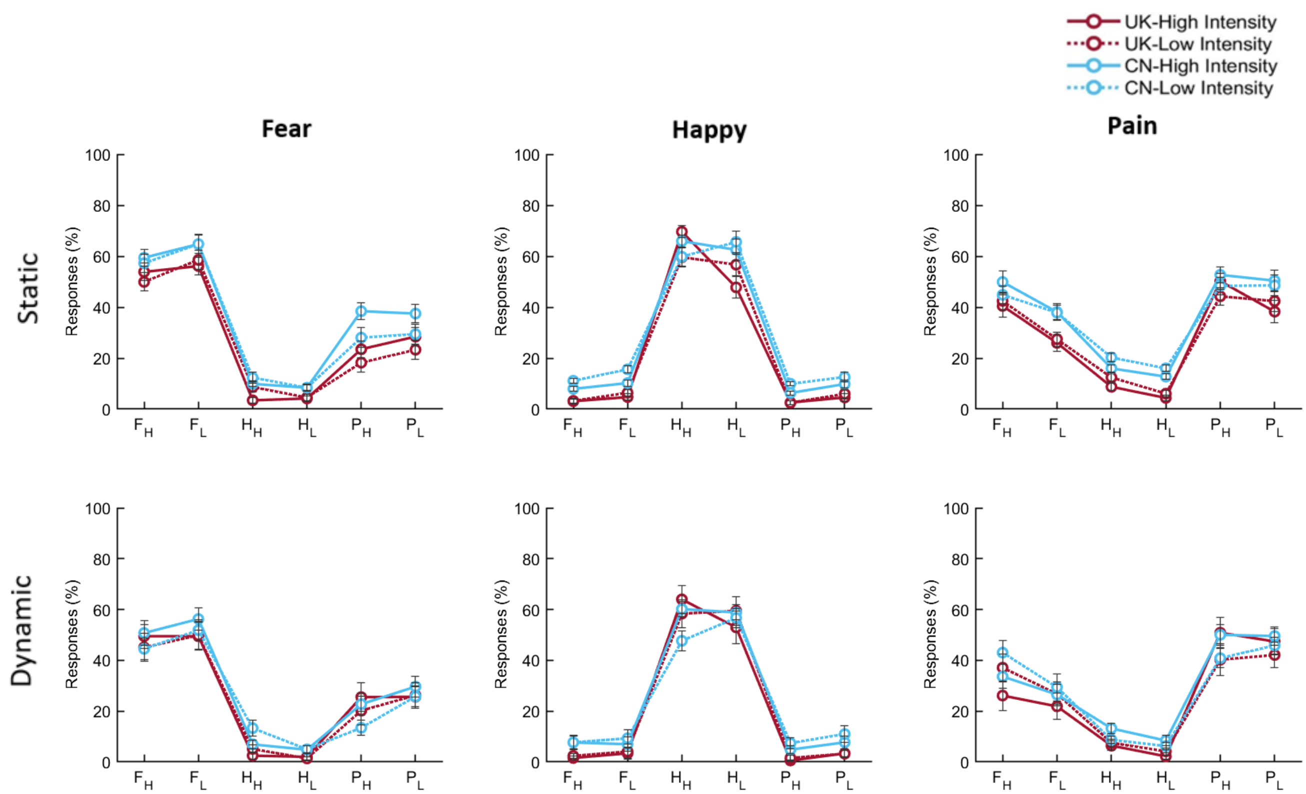

Figure C.2. **Context Profiles in response to Physical scenarios.** *Responses averaged across Chinese (in blue) and British (in red) participants to the Context profiling task for Physical scenarios. Responses to Static (top row) and Dynamic (bottom row) facial emotions of High-intensity (solid line) and Low intensity (dashed lines) Fear, Happy and Pain (columns).*

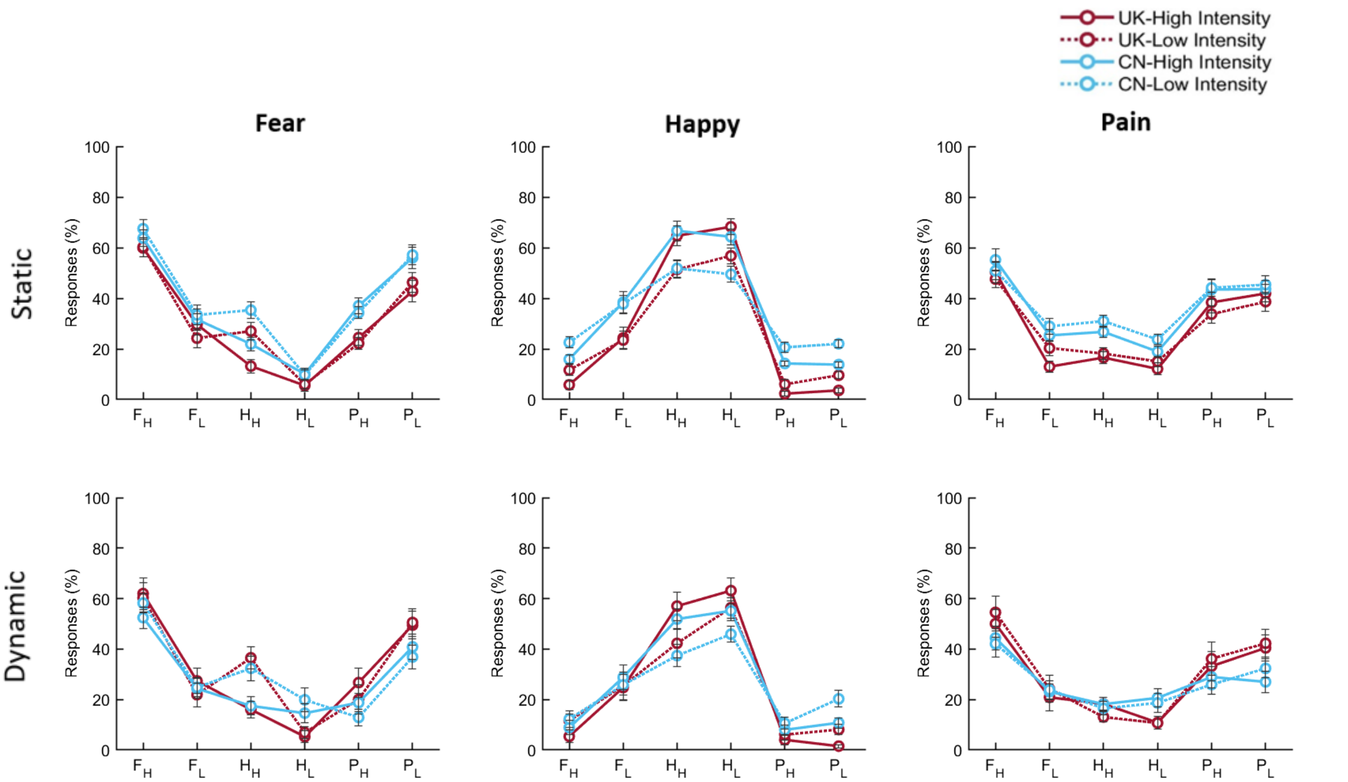

Figure C.3. **Context Profiles in response to Social scenarios.** *Responses averaged across Chinese (in blue) and British (in red) participants to the Context profiling task for Social scenarios. Responses to Static (top row) and Dynamic (bottom row) facial emotions of High-intensity (solid line) and Low intensity (dashed lines) Fear, Happy and Pain (columns).*