

Automatic Facial Expression Analysis as a Measure of User-Designer Empathy

Aleksi Salmi

Design Factory,
Department of Mechanical Engineering,
Aalto University,
Espoo, Finland
e-mail: alekksisalmi@gmail.com

Jie Li

Design Factory,
Department of Mechanical Engineering,
Aalto University,
Espoo, Finland
e-mail: jie.li@aalto.fi

Katja Holtta-Otto¹

Mechanical Engineering,
The University of Melbourne,
Parkville, VIC 3010, Australia;
Aalto University,
Espoo, Finland,
e-mail: katja.holttaotto@unimelb.edu.au

In human-centered product design and development, understanding the users is essential. Empathizing with the user can help designers gain deeper insights into the user experience and their needs. However, a few studies have captured empathy real time during user interactions. Accordingly, the degree to which empathy occurs and enhances user understanding remains unclear. To narrow this gap, a study was performed exploring the use of video-based facial expression analysis during user interviews, as a means to capture empathy related to understanding vehicle driving experiences under challenging conditions. Mimicry and synchrony have been shown to be predictors of empathy in cognitive psychology. In this study, we adapted this method to study 46 user-designer interviews. The results show that the user and designer exhibited mimicry in their facial expressions, which thereby indicated that affective empathy can be captured via simple video facial recognition. However, we found that the user's facial expressions might not represent their actual emotional tone, which can mislead the designer, and they achieve false empathy. Further, we did not find a link between the observed mimicry of facial expressions and the understanding of mental contents, which indicated that the affective and some cognitive parts of user empathy may not be directly connected. Further studies are needed to understand how facial expression analysis can further be used to study and advance empathic design.

[DOI: 10.1115/1.4056494]

Keywords: empathy, facial expression analysis, user understanding, product development, user-centered design, design methodology, design theory and methodology, product design

1 Introduction

The user-related approach has been shown to be a key to successful innovation in general [1] and in mechanical products both at the time of the launch [2] as well as later in the market [3]. Developing these innovations is an iterative design process that includes many steps, such as user understanding, ideation, and prototyping. In this article, we focus on the user understanding phase. Several user-centered and empathic methods have been popularized in non-academic design method books. These empathic methods help designers to gain deeper user understanding, identify user needs, and potentially enhance their creativity [4–6]. However, past studies do not separate empathy (generally the ability to share and understand user feelings about the design to the designer) from the rest of the user-centered design process, and thus, the link between empathy and good design remains unclear. To explain how exactly empathy contributes to design is complex and requires also quantitative studies of how empathy manifests, develops, and impacts design outcomes.

The complexity stems from the fact that empathy is an umbrella term [7], which covers many different aspects of what different people and areas of research call empathy. In psychology, there are many different definitions and aspects of it, such as affective and cognitive empathy, conscious and unconscious empathy, perspective-taking, and emotional contagion [8]. Chang-Arana et al. [7] showed how these terms do not map well to how design research discusses empathy. Surma-Aho and Hölttä-Otto [9] conceptualized empathy in the context of design (Fig. 1). They separate empathy into the external and internal aspects that relate to building empathic understanding. Empathic design methods and

the resulting empathic decisions or designs themselves (design actions) are external, whereas the empathic orientation and the mental processes of the designers are internal aspects of empathy. Surma-Aho and Hölttä-Otto [9] also identified potential measures for each of these five aspects (Fig. 1).

We use this suggested operationalization and aim to explore a novel means to measure the empathic mental processes in the context of understanding user experiences under difficult driving scenarios. Common mental process measures include the use of various sensors to measure shared physiology [10]. Facial expression analysis using electromyography (EMG) is another common approach in social psychology to study mutual understanding between two people, and it has been adapted also to study user interviews [11]. We build on this past work but explore if we can measure empathic mental processes using automatic facial expression analysis from a video instead of physical sensors. Successfully measuring empathy with computer vision-based facial expression analysis could allow us to measure the undelaying subconscious mental processes that are involved in developing empathic understanding in more natural environments than in a laboratory. This could enable new types of empathy research to help build a more fine-grained understanding of it, why and how empathic design methods contribute to user understanding and design in general.

2 Literature Review

2.1 Forms of Empathy. Empathy is a term that has several alternative definitions depending on the context and application. For example, empathy has been defined as knowing the affective and cognitive state of another person, mimicking another, sharing another's feelings, or imagining what another person is feeling or thinking [12]. Often, empathy is seen as a three-component term consisting of cognitive empathy, affective empathy, and compassionate empathy [13]. Affective empathy is an identification of another's emotional states [14] and is elicited by emotional

¹Corresponding author.

Contributed by the Design Theory and Methodology Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received June 3, 2022; final manuscript received December 11, 2022; published online January 9, 2023. Assoc. Editor: Katherine Fu.

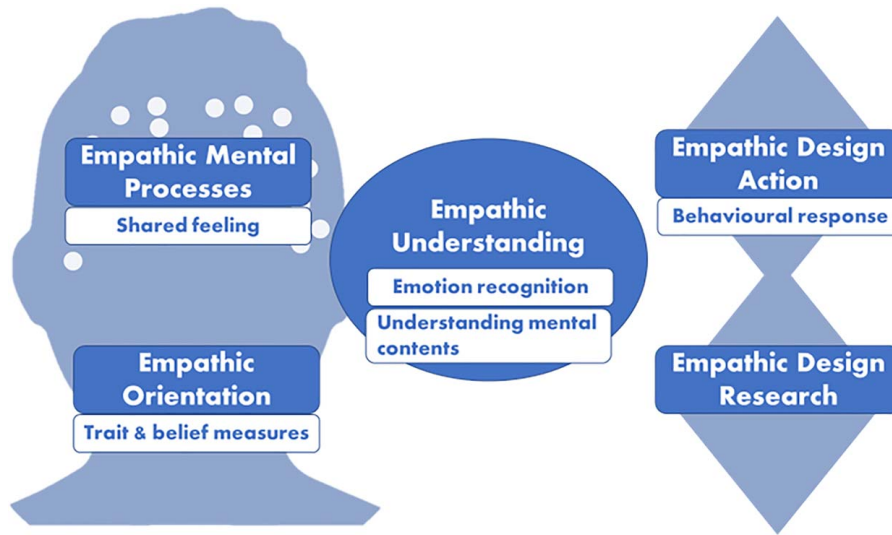


Fig. 1 Five conceptualizations of empathy (dark blue boxes) and their potential operationalizations (white boxes). Adapted from Ref. [9].

stimulus [8]. Cognitive empathy is our ability to correctly identify another person's mental contents, which is related to the theory of mind in psychology [15]. It is ambiguous whether cognitive empathy is a separate process from affective empathy or whether they overlap. For example, recognizing the most basic emotions can be unconscious, but the understanding of more complex mental states requires cognitive processing [13].

As discussed earlier, the multidimensional nature of empathy in design has been conceptualized by Surma-Aho and Hölttä-Otto [9]. In their conceptualization, the goal of user research is empathic understanding, which describes the learning and new cognitive understanding that happens as a result of different empathic processes. In this article, we focus on the empathic mental processes as reflected on the person's face and their relationship with the overall empathic understanding.

2.2 Measuring Empathy. Measuring empathy is challenging. Empathy is a complex dynamic activity between individuals. When empathizing with another person, we will go through conscious and unconscious mental processes, which often refer to affective and cognitive empathy. Researchers believe that affective empathy is more automatic, while identifying cognitive elements help re-modulate affective elements that we have identified [16]. Chang-Arana et al. [7] showed how the different aspects of empathy identified elsewhere do not map well to how it is understood in design, further complicating measuring empathy in design.

In general, several empathy measure types have been developed, and while their applicability in design has been discussed [9], there are no generally accepted means to measure empathy. Elsewhere, measures are generally divided into trait measures, behavioral measures, and physiological measures. We briefly introduce the types in the following sections and discuss the ones we choose for this study.

2.2.1 Trait Measures. Trait measures such as the interpersonal reactivity index (IRI), the empathy quotient, and the Balanced Emotional Empathy Scale (BEES) [9,17] are questionnaires that measure a person's self-reported empathic tendencies. Of these, IRI has been applied in engineering and design [18–20]. However, a trait measure can be prone to biases and somewhat static and may thus not describe any specific design context or point in time, but rather the person's tendency or belief of empathy in general.

2.2.2 Behavioral Measures. Behavioral measures are task-based measures that are less prone to biases than trait measures and can be used to measure either cognitive or affective empathy.

Affective empathy can be measured, for example, by exposing the participant to facial emotion and comparing the self-reported emotional state of the participant to the original emotional state in the picture [17]. Cognitive empathy can also be measured using behavioral measures. In such measures, the participant is exposed to a stimulus after which they conduct a cognitive empathy task. One example of a behavioral measure is MET (multifaceted empathy test) where the participant is exposed to photorealistic stimuli including complex emotions and context. Cognitive empathy is measured by asking the participant to infer the mental state of the pictured person [21].

Facial emotion recognition, the ability to identify an emotional expression, can be used for measuring both cognitive and affective empathy. While emotional contagion and the transfer of emotional expression is an automatic affective process that happens subconsciously, recognizing and identifying an emotional expression requires cognitive processing [22]. Facial emotion recognition skills might not fully correspond to cognitive empathy skills because facial emotion recognition is fully based on facial expression, whereas cognitive empathy is based on not only facial expressions but also other cues, such as speech and situational cues [17]. In addition to recognizing emotions, facial expressions can be analyzed for shared physiology, which is discussed in Sec. 2.3.3. An important aspect of emotion recognition is emotional tone, the extent to which a person's thought or feeling is positive or negative.

In this study, we use both emotional tone accuracy, or the recognition of the positive and negative tone of the emotion, as well as a cognitive behavioral measurement called empathic accuracy (EA) [11,23] and compare the novel automated facial expression analysis to these two measurements. Empathic accuracy measures how well a person understands another person's mental state. Thus, it measures cognitive empathy. Empathic accuracy has been previously used to measure how designers understand users and how empathy could be quantified in the design context [11,24,25]. Chang-Arana et al. [11] observed using empathic accuracy that designers are more accurate in inferring design-related thoughts and feelings than nondesign-related one, and Li et al. [24,25] found that accuracy is impacted by national cultural differences. Emotional tone accuracy, on the other hand, has been linked to improving the ability to identify important user needs [26].

2.2.3 Physiological Measures of Empathy. Shared physiology, measuring physiological signals such as facial expressions or bodily movements, is a common approach to measure affective empathy or

the mental processes related to empathy [27]. Mimicry is a related concept that refers to an unintentional and unconscious process where a person mimics accents, postures, facial expressions, or other behaviors from another person (e.g., Ref. [28]). Mimicry does not appear to be a learned behavior, but an evolutionary phenomenon that is also observed in infants and some animals [29]. It is suggested that facial mimicry supports emotional contagion, i.e., transferring emotions from one person to another. When exposed to a face with a certain emotional expression, a person tends to mimic that expression, which creates an emotional state that corresponds to the emotional state of another person [29]. Mimicry also has many different impacts on social interactions. People who are being mimicked by another person act more generously than people who are not being mimicked by another person [30]. However, mimicry leads to emotional contagion only with some emotions. According to Hess and Blairy [31], mimicry leads to transferring of emotion from person to person, only in the case of sadness and happiness. In other mental states, mimicry might still occur but does not lead to the transfer of emotional state. Similarly, Hinsz and Tomhave [32] found that smiling at another person leads to them mimicking your smile, but frowning does not lead to another person frowning back.

In physiological measures, an activity such as facial muscle movement or heart rate is measured during an interaction [12]. Mimicry can be measured using physiological measures, such as EMG. These EMG sensors can be placed on the two people of interest, and the sensor data can be analyzed for synchrony. Physiological measuring requires special equipment, which sometimes leads to small sample sizes [12]. Mimicry of facial expressions can be observed by measuring the activity of facial muscles. Different muscles are activated when a person is predisposed to a face with different emotions. For example, a happy face activates the zygomatic major muscle, angry face activates the corrugator supercilii muscle, and sad face activates the corrugator supercilii muscle. These changes in muscle activity happen even when participants are not aware that they have consciously seen a face with the specific mental state [33].

Similar measures can be taken from video recordings. For example, facial muscle activity can also be measured from video recordings using the facial action coding system [34]. Facial movements are presented using so-called action units (AUs). Action units are visually distinguishable movements in the face that correspond to different muscle activities. For example, movement, where lip corners are pulled upward, indicates the activation of zygomatic major, and it indicates that a person is smiling [35] (Fig. 3). The facial action coding system can now be readily used with automated computer programs that export the action unit data, instead of tracking action units manually from a video. Action units are calculated by tracking specific landmarks in face and observing their relative distances. For example, the lip corner pulling movement is observed by measuring the distance between the landmark points in the right and left corners of lips. This method has not previously been applied to measuring designer-user empathy. Multiple different projects and products offer software to analyze action units from a video. In this study, we used Open Face [36,37]. OPENFACE is an open-source framework that provides the user with information about action units, gaze, facial landmarks, and head pose (Fig. 2). In this study, we use the action unit data, which are captured as a CSV file from the software.

2.3 Relationship Between Empathy and Mimicry

2.3.1 Relationship Between Mimicry and Emotional Contagion. While there remains a debate over whether mimicry supports emotional contagion, there are several studies that have found such a relation. It appears that the relation between mimicry and emotional contagion is bidirectional. People with stronger empathic traits show more mimicry [38], and mimicking another's expressions creates a shared emotional state [39].

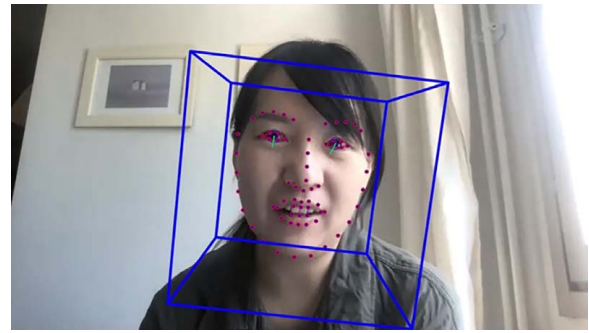


Fig. 2 Screen capture from OPENFACE software showing the purple dots representing the facial landmarks. Relative distances between specific landmark points are calculated to identify different action units




	Lip corner puller: AU12
	Brow lowerer: AU04
	Cheek raiser: AU06

Fig. 3 Examples of different action units

Deng and Hu [40] studied the relationship between mimicry and emotional contagion. They showed dynamic happy and angry faces to participants. Participants' reactions were measured using facial electromyography recorded from zygomatic major and corrugator supercilii. After seeing a happy or angry face, participants self-reported their feelings. The similarity of the emotion in the stimuli and the self-reported emotion was used to calculate emotional contagion. It was observed that emotional contagion happened both for happy and angry faces. However, only with happy faces did the participants react with corresponding muscle activity, which indicates that only positive emotional contagion happens through mimicry. They also observed social appraisal by creating stimuli that had happy and angry faces, which were targeted toward or away from another face [40]. It was discovered that there was a difference in self-evaluated anger when the angry face was or was not targeted at a person. This indicates that social appraisal affects the emotional contagion of anger. Such difference was not found in happy faces. This suggests that emotional contagion happens differently for positive and negative emotions.

There are a few factors that influence mimicry. According to Howard and Gengler [41], people tend to mimic more when they like the people. It also appears that people who are expressive mimic more. Both expressive and not expressive persons seem to mimic happy emotions, but negative emotions are mimicked mostly by expressive people [41]. This has led many studies in social psychology to select expressive people as participants [42], which might limit the applicability of the approach.

2.3.2 Mimicry and Empathic Understanding. Since mimicry is related to emotion transfer and transferring of emotions requires cognitive empathy, it seems that there is also a relation between mimicry and empathic understanding. Automatic muscle activities

trigger emotions, which helps to understand the emotions that another person is having [40].

However, the studies are inconclusive when mimicry leads to improved empathic understanding or cognitive empathy. Neumann et al. [21] found that the mimicry was correlated more with affective than cognitive empathy. They found that EMG measures of facial mimicry correlated with how the participants empathized with the shown images, but less so with the emotion recognition test. On the other hand, Soto and Levenson [42] found that mimicry relates to both affective and cognitive type empathy. Their study was based on two different experiments where the stimuli containing positive and negative, static and dynamic expressions were shown to participants. The results showed that the facial mimicry, measured via EMG, correlated with both the shared emotional tone and the recognized emotional tone in response to others' facial expressions. It is not known if mimicry in designer–user interaction leads to cognitive empathic understanding, affective empathy as a shared feeling, or neither or both.

In human-centered design empathy refers to user understanding in a given context; therefore, it involves a dynamic understanding process between designers and users. In this study, we aim to explore measuring empathy during this user interaction. Facial expression analysis using EMG sensors is performed in the most common approach to measure affective empathy, and it has been piloted in design [11]. Therefore, in our study, we also aim to explore facial expression analysis and affective empathy, except we present a novel automated computer vision-based approach as the means to measure affective designer empathic mental processes. We investigate the computer vision-based facial recognition relationship with cognitive empathic understanding measured as emotional tone accuracy and empathy accuracy.

3 Research Questions

While empathy has been identified as important and useful, there is no simple way to measure empathy during user interviews and evaluate the designer's empathic performance. Some aspects of empathy can be read by observing the facial mimicry of two people, but such techniques have been mostly used in psychology and neuroscience. Also, the relationship between empathic understanding and the physiological aspect of empathy is unclear. Chang-Arana et al. [11] are a notable recent exception. They piloted facial expression synchrony measures using EMG sensors and compared them with empathic accuracy. We build on their work but present a novel approach to use automated facial expression analysis from a video rather than a sensor-based approach to measuring empathy during a user–designer interview. Specifically, we will focus on the following research questions:

When using an automated video-based facial expression analysis instead of physiological sensors,

- (1) Is there facial mimicry between the user and the designer?
- (2a) Is there a relationship between the user's reported emotional tone and their facial activities?
- (2b) Is there a relationship between the user's facial activities and what emotional tone the designer thinks the user has?
- (3) Is there a relationship between the affective mimicry and the designer's cognitive empathic understanding?

The first question replicates measures taken in earlier studies since in earlier studies, there has been facial mimicry between an interviewer and participant, but when using physiological sensors. We replicate the measure to ensure we can use the video-based non-contact sensor method to measure the mimicry between the interviewer and the participant.

Our second set of questions relates to whether the user's reported emotional tone relates to their facial activities. This helps us to understand how well we can read a person's emotional tone by analyzing their facial expressions.

Our third question is if we can measure the correlation between designer's affective mimicry and cognitive empathic understanding. We measure emotional tone accuracy and empathic accuracy and compare them with mimicry in a video-recorded user interview. Emotional tone accuracy measures how well the designer understands if the user's thought or feeling is positive or negative, and empathic accuracy measures how well the designer understands what the user is thinking or feeling.

4 Approach

In this study, we explore a novel automatic facial expression analysis of mimicry and study its relationship with empathic understanding measured via a behavioral empathy measure called empathic accuracy [11,24,25]. This measure quantifies how accurately a designer can understand a user when the user comes up with a thought or feeling. Both empathic accuracy and automatic facial expression analyses enable recording empathic cues during a dynamic interpersonal activity such as a user interview. It will tell us the degree to which designers understand the users during an interview. This measure has three parts: user task, designer task, and rating part.

4.1 Participants. In this study, three different designers conducted semi-structured interviews with 46 users. The interviews were about the vehicle driving experience; thus, the common criteria for all participants were that they have a driving license and driving experience. There were no other preconditions for the users such as age, gender, or nationality. In the user group, there were 26 males and 20 females. Nineteen of them were Finnish, 15 were Chinese, and 22 were from other countries including the United States, Brazil, Pakistan, India, Thailand, Turkey, Spain, Netherlands, Germany, and France. The three designers all have master-level degrees or above in design (product development or industrial design). All participants were living in Finland when participating in the experiment, and their working or studying language is English. The interview language was English. Each participant received 15 euros per hour as compensation.

4.2 Structure of the Interviews. A total of 46 participants were interviewed by zoom since physical meetings were avoided due to the COVID-19 restrictions. The interviews lasted approximately 30 min. The conversation was about the driving experience, more precisely driving in the city centers, crossroads, inclement weather, and other circumstances where there might be visibility concerns. The interviews were semi-structured. Each designer followed the same interview guide that contained introductory questions, questions about driving, and questions about specific situations while driving. The introduction lasted for a couple of minutes depending on the interview, and they were meant to let the user and the designer warm up and get to know each other before the questions related to driving. Driving-related questions were about the participants driving history, driving habits, and their thoughts about driving. Driving-related questions gave us more insight into the users' relationship with driving, and they helped the users to freshen their memories about driving a car. This was followed by imaginary scenarios that aimed to gain information about how the user would act in different situations. The designers presented a situation such as driving in heavy rain and poor visibility in a residential area, and the participant was asked about what they would do and what kind of information they would need in such a situation. These questions provide information for another study, not reported here, where a driving-related product is being designed.

4.3 Empathic Accuracy Task for the User. The interview was recorded, and the video was sent to the user immediately after the interview. The user was asked to do the empathic accuracy

Table 1 Example sheet used in the empathy task for the user participants

Time	Thought or feeling (Mental content)	Emotional tone +, 0, -
6:47	<i>I was: feeling happy because the best part of driving is that you do not need to wait.</i>	+ 0 -
10:25	<i>I was: thinking what is a possible way to see in a heavy rain.</i>	+ 0 -
12:04	<i>I was: feeling terrifying to drive in a city center.</i>	+ 0 -

task following the same procedure as used in other studies [27–29]. Briefly, the user watched the recording and paused it when they remembered having a specific thought or feeling during the interview. That thought or feeling, along with the timestamp, was then marked down by the participant on an Excel sheet (Table 1). The participant also evaluated whether the feeling was negative, neutral, or positive. This is called emotional tone (Table 1).

The users went through the whole interview collecting all the thoughts and feelings they remembered having. Finally, the user was asked to write down a few needs they would have for a driving-related product that was discussed during the interview. Altogether, the user activity lasted approximately 1.5 h.

4.4 Empathic Accuracy Task for the Designer. After the participants recorded their thoughts and feelings, the designers conducted a similar task to find out the empathic accuracy of the designers. The designers watched the recorded user interview and paused at the time when the user came up with a mental content; and the designer then inferred the user’s mental content and emotional tone. As an empathic assessment, we used the quick empathic accuracy (QEA) [26] measure. This measure shortens the original empathic accuracy measure [23] by randomly selecting only ten users’ mental contents. Usually, each user will provide 10–20 mental contents during an approximately 30-min interview. To infer these mental contents, it will take each designer 1.5–2 h per user. By using the QEA, it saves approximately 1 h per user. Its reliability and validity were confirmed in a previous study [26]. QEA results in lower accuracy scores, but the accuracy distribution is similar to the full EA [26]. The randomly selected ten thoughts or feelings were not selected from the first few minutes of the interview to increase the number of thoughts or feelings that are related to the context of the interview and not to the introductory questions.

Next, we cut the video based on selected entries. To reduce the amount of work required from the interviewers, the videos were processed automatically to a form that contained only the parts of the videos that the interviewer had to see. The processed videos contained a 30-s clip for each thought or feeling. The clips started 30 s before each thought or feeling and ended at the time where the participant had marked down the thought or feeling. Processed videos also had 30-s pauses between each clip to give the

interviewer time to mark down their answer. During this step, the designers watched the whole video first and then watched the cut vision and completed the inferring task. The whole video provides complete information about the user interview. The cut version keeps the content before and when the user comes up with a thought or feeling and saves time and effort.

Finally, we calculate designers’ empathic accuracy by having three independent raters rate the similarity between the answers from the interviewer and the participant (Table 2). The three raters are well trained and familiar with the empathic accuracy rating. Cross-rater’s Cronbach’s alpha is 0.823, which indicates a good rating reliability. The criterion for choosing the reviewers was that they use English as their working language. The similarity of each sentence was evaluated by a three-point Likert scale (0 = completely different content, 1 = somewhat similar content, 2 = equaled similar content). The average from the three evaluators was calculated for each thought or feeling, and an average of those was calculated for each designer user pair. This empathic accuracy was expressed in percentage.

4.5 OPENFACE Measurements for Facial Expression Analysis. The interview videos were recorded using Zoom. These videos were used to analyze the action units in the designer’s and the user’s faces. Each video contains the faces of the interviewer and the participant side by side. The videos were cropped into two separate videos, where each contained only one face. This was done so that we can process each face separately and that we know which action units belong to which face. The cropping of the videos was done automatically using the FFmpeg library in PYTHON. OPENFACE software was used to track facial expressions. OPENFACE tracks head pose and gaze, but in our study, we focused on facial expressions. For this, OPENFACE tracks 68 different facial landmarks and estimates the so-called action units based on the relative distances of the landmarks.

Usually in empathy-related studies, mimicry is measured by observing zygomaticus major and corrugator supercilii muscles with an EMG. Zygomaticus major and orbicularis oculi are the muscles that activate during a smile. There are several action unit patterns of these muscles that can be captured. We looked for Action Unit 12, lip corner puller, which approximates the activation

Table 2 Example of empathic accuracy rating

User’s entry	Designer’s inferences	Similarity score
“I was thinking of parking garages, which makes me anxious.”	<i>He was thinking about the difficulties in a parking garage and how it is similar to parking somewhere else.</i>	1
“I was satisfied of having found some solutions.”	<i>He was wondering how to trust the device.</i>	0
“I was thinking I hate driving in the rain so bad.”	<i>He was feeling so difficult and anxious to drive in the rain, and he hates it.</i>	2

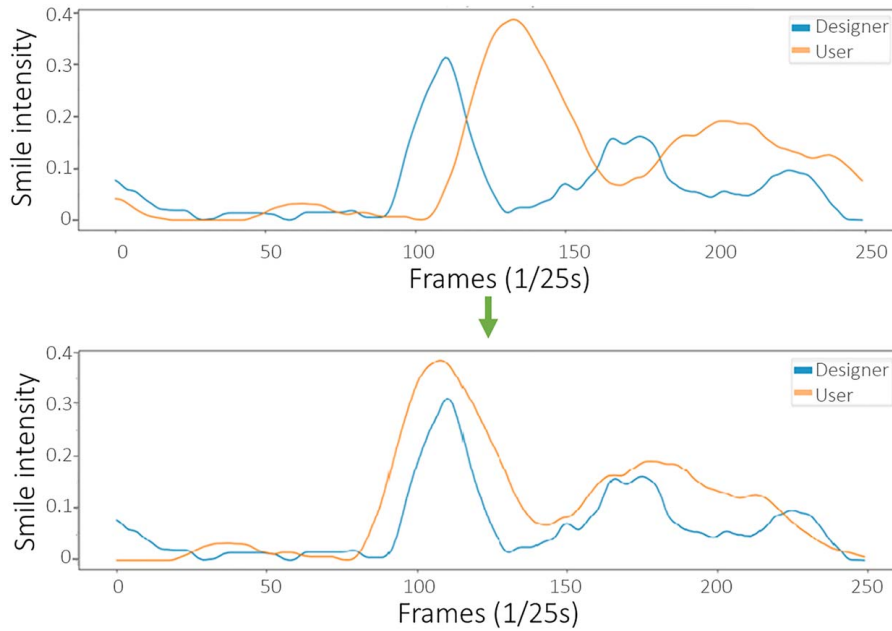


Fig. 4 Example of shifting the data when using windowed cross-lagged correlation.

of zygomaticus major for use in the empathy study. We also observed Action Unit 6, cheek riser, which approximates the activation of orbicularis oculi. We observed Action Unit 6 since the observation does not require any sensors, and it helped us to monitor how consistent our measurements are. Another muscle that is useful for empathic studies is the corrugator supercilii, a muscle that activates during frowning [43]. Corrugator supercilii activates during the expression of suffering [44]. Finally, another muscle is the depressor glabellae, which contributes to the feeling of anger [44]. Action Unit 4 describes both corrugator supercilii and depressor glabellae. Figure 3 shows examples of images, where OPENFACE detected particular action units. We monitored Action Units 6, 12, and 4 to capture both positive and negative emotions.

The action unit data are exported as an Excel file from the OPENFACE. The Excel file consists of 40 columns and a varying number of rows. Each row represents a frame in the video, and the columns contain information about the intensity and presence of 17 different action units. The values for presence are either 1 or 0, which indicate whether an action unit is present or not, and a value for intensity is a positive decimal number and it indicates the intensity of the particular action unit.

OPENFACE reads the action units from the person's face. Some factors make reading impossible such as objects blocking the face or technical issues caused by bad internet connection. The exported file contains a confidence value that tells how confident the neural network is, and that there is a face in the picture. By visually observing the frames with a low confidence value, we noticed that in frames with a confidence value less than 0.95, OPENFACE failed to correctly identify the location and orientation of the persons' face, and such frames were removed from the data.

4.6 Data Processing. To smooth the data, we calculated the moving average from the time series data. In a similar study [45], a moving average window of 0.5 s was used. We used a similar window. Because our data are from an interview, there were also some dynamics to consider. One outcome in particular that we were looking for was mimicry, more precisely here the mirroring of different expressions. However, the mirroring of expressions is delayed from one person to the other and so causes the time series not to be aligned. We took two approaches to deal with the delay between the signals. When we calculated the mimicry during the whole interviews, we used windowed cross-lagged

correlation and when we looked at the individual reported thoughts and feelings, we looked at the similarity between the facial expressions of the user and the designer.

In windowed cross-lagged correlation, the two time series are divided into smaller windows, where one of the series is shifted. After the shifting, the correlation between the two time series is calculated. An example of shifting is shown in Fig. 4. Riehle et al. [45] conducted a similar analysis, where the mirroring was expected to happen within 1 s.

Figure 4 shows an example of a case, where the designer is smiling, and the user is mirroring that smile after about 0.8 s (20 frames). In the lower image, the designer data are shifted 0.8 s forward. We calculated the correlation between the user and the designer in 7-s windows.

During the individual thoughts and feelings, the mimicry was measured by observing the similarity of the user's and the designer's facial expressions, which indicates the shared feeling between the designer and the user. These measures were conducted in the ± 5 s range relative to timestamps of each thought and feeling. The ± 5 s time frame was to make sure that we record the facial expressions where the thought had appeared, even if the user did not mark down exactly the right time stamp. We also did not assume that the user and the designer were smiling at the exact timestamps when the user had a thought or a feeling. However, by visually studying the data, we observed that most of the peaks in the action unit data that happen near reported timestamps happen within a time range of ± 5 s from the timestamp.

4.7 Statistical Analysis. The data in this study are nonparametric. In other words, all data failed normality tests and are analyzed distribution free. We also tested for the homogeneity of variance of all data and report those in the results. For RQ1, we use windowed cross-lagged correlation to investigate the mimicry between the user and the designer. To test the significance of this mimicry, we used a Monte-Carlo test (explained in the following section). For RQ2 and RQ3, we use rank analysis of covariance (ANCOVA) [46], namely, Quade's nonparametric ANCOVA in SPSS to investigate the difference between user-detected and self-reported emotional tone as well as designers' correct and incorrect understanding of user's emotional tone. We add both the user and the designer as categorical covariates in the analysis since we have ten data points (the individual thoughts and feelings from

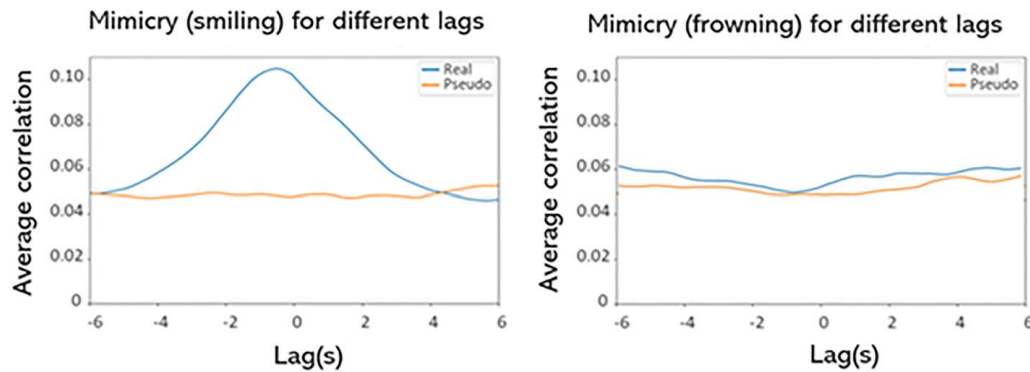


Fig. 5 Comparison between real data and pseudo data for smiling (left) and for frowning (right)

each interview) from each design–user pair. A categorical covariate is similar to a block in analysis of variance [47]. Spearman correlation between the designers’ empathic understanding (empathic accuracy) and mimicry of facial expressions is used in RQ3.

5 Results

5.1 Mimicry Is Detectable in Smiling, not in Frowning. Our first research question aimed to see if we can find evidence of mimicry in designer user interviews using automatic facial expression analysis. We observed the mimicry between the user and the designer by calculating mimicry with different lags in all 7 s windows in all user–designer pairs. The correlation between the users and the designers smiling is highest ($r=0.1$) when the lag between the user and the designer is approximately -0.5 s. The correlation was overall highest when the lag was around a few seconds (Fig. 5). This supports our hypothesis that there is mimicry that happens within a few seconds.

We used a Monte Carlo test to check whether the synchrony of smiling is statistically significant. We created 1000 sets of pseudo data that were sampled from the original data such that the designer’s and the user’s smiling data were from different interviews or from different parts of the interview. In the pseudo data, the correlation was approximately constant (0.05–0.06) and lower than in the real data (Fig. 5). We calculated the best correlation for each window in the real data and in the pseudo data sets to observe whether there is a statistically significant difference. Windows with no significant mimicry ($p>0.1$) were deleted from the data, and the remaining values were normalized using Fisher’s z -transformation [45]. The average correlation for all windows in the real data was higher than in all the permuted datasets ($p=0.00$). This suggests that there is mimicry of smiling in the real data but not in the pseudo data.

On the basis of the past research, we expected that smiling leads to mimicry, but frowning does not. We tested this by repeating the real–pseudo data comparison for frowning (AU04). The average correlation for different lags was approximately 0.05–0.06 for both pseudo and real data. There was no significant difference between the pseudo and the real data. The average correlation of pseudo data windows was higher than from the real windows in 40% of 1000 ($p=0.4$). This indicates that there is no noticeable mimicry when the user and the designer frown. This coincides with the past work [40,48], which showed that mimicry happens with smiling but not with frowning.

5.2 User’s Smiling and Frowning Intensity Is Not Correlated With Their Reported Emotional Tone, but Is Related to the Designer’s Guesses. Research question 2a aimed to investigate if there is a relationship between the detected facial expressions and users’ self-reported emotional tone. During the empathic accuracy task, the participants recorded the emotional tone of each thought and feeling they wrote down. They also

noted the timestamp for each thought and feeling. In total, there were 419 timestamps for all 46 user designer pairs. We read the user’s facial expressions around those timestamps and estimated how well the expressions matched with the reported emotional tone. The emotional tone of the user was assessed by measuring the user’s smiling (AU12, lip corner puller) and frowning (AU04 brow lower) intensities. The intensity of the action units was measured in a 10 s period around the exact time of the user’s self-reported emotional tone.

To answer RQ2a of smiling intensity, we considered both the user and the designer as categorical covariates. The homogeneity of variance assumption is violated for the smiling intensity data ($p=0.037$) but not for the frowning intensity data ($p=0.399$), but since both data are nonnormal, we use Quade’s nonparametric ANCOVA. The dependent variables are smiling intensity and frowning intensity, and the independent variable is users’ self-reported emotional tone. We only consider positive and negative emotional tone since neutral emotional tone usually relates to no facial expression. We did not find any significant difference in smiling intensities between positive ($N=134$, $M=1.54$, Std.=1.23) and negative ($N=126$, $M=1.29$, Std.=1.09) self-reported emotional tones ($F [1, 258]=0.674$, $p=0.413$). Similarly, we did not find any significant difference of frowning intensities between positive ($N=134$, $M=0.80$, Std.=0.653) and negative ($N=126$, $M=0.77$, Std.=0.580) emotional tones ($F [1, 258]=0.017$, $p=0.896$).

In research question 2b, we consider smiling and frowning intensities as dependent variables, designers’ guess as an independent variable, and both user and the designer as covariates. The homogeneity of variance is violated for the smiling intensity data ($p=0.044$) but not for the frowning intensity data ($p=.779$), but given both data are nonnormal, we use Quade’s nonparametric ANCOVA. In total, 186 guesses are included. A total of 112 of them are negative and 74 of them are positive. Neutral guesses are excluded. The difference in smiling intensities between positive ($N=74$, $M=1.88$, Std.=1.20) and negative ($N=112$, $M=1.23$, Std.=1.02) guesses was significant ($F [1, 184]=10.874$, $p=0.001$). The difference in frowning intensities between positive ($N=74$, $M=0.664$, Std.=0.528) and negative ($N=112$, $M=0.803$, Std.=0.554) guesses was not significant ($F [1, 184]=2.995$, $p=0.065$).

In conclusion, users did not smile significantly more when they reported a positive feeling nor did they frown more when they reported a negative feeling. However, a strong relationship was found between smiling and the designer’s guesses. The smiling of a user did not indicate that the user themselves were happy, but it seems to be related to the designers’ guess that the user was happy. The designer seems to be fooled by false smiles.

5.3 Mimicry Is Related to How Well the Designer Recognizes the User’s Emotional Tone but Not How Well They Understand Mental Contents. The third research question

focused on the relationship between affective mimicry and cognitive empathic understanding, which is measured as the designers' emotional tone accuracy and empathic accuracy. The emotional tone accuracy refers to how correct the designers can recognize the user's emotional feeling. The empathic accuracy indicates how accurately the designer understands the user's mental content.

When testing the relationship between emotional tone accuracy and mimicry, three dependent variables are analyzed: mimicry of all entries, mimicry of positive emotional tone entries, and mimicry of negative entries. The independent variable, the emotional tone accuracy, is coded into a binary variable (0=wrong, 1=correct). The homogeneity of variance assumption of over-all entries is met ($p=0.954$). The positive entries pass the homogeneity of variance test ($p=0.101$), but the negative entries do not ($p=0.004$). But since all data are nonnormal, we use Quade's non-parametric ANCOVA with the designer and the user as covariates.

Overall, the average mimicry of correct predictions of all entries is 17.19 ($N=86$, Std. = 26.96), while of wrong predictions is 16.83 ($N=174$, Std. = 28.81), and the difference is not significant ($F [1, 258]=0.224$, $p=0.637$). For positive entries, smiling mimicry is higher when the designers' emotional tone predictions are correct ($N=37$, $M=29.21$, Std. = 32.07) than wrong ($N=97$, $M=17.86$, Std. = 29.13), but the difference is not significant ($F [1, 132]=3.811$, $p=0.053$). Given the results are close to significance, we looked at the top and bottom 15% mimicry data and found that for top 15% mimicry, the designer is able to guess the positive tone correctly 50% of the time, whereas the bottom 15% can guess only 10% correct. The average rate of guessing positive emotional tone is 28%.

For negative emotional tone entries, we did not find a significant difference. Higher smiling synchrony is related to smaller accuracy in designer's guesses about the user's emotional tone, but there was no significant difference in the level of mimicry during correct ($N=49$, $M=8.11$, Std. = 17.90) or wrong guesses ($N=77$, $M=15.54$, Std. = 28.53), ($F [1, 124]=0.741$, $p=0.391$). Here, the average rate of guessing the emotional tone correctly is 39%, higher than guessing the positive entries on average, but it might be more static since it does not depend on the mimicry between the user and the designer as much it did for the positive entries mentioned earlier.

When testing the correlation between empathic accuracy and mimicry, we aimed to understand the correlation between empathic accuracy and mimicry of all entries, positive entries only, and negative entries only. There is no strong correlation ($r=-0.045$, $p=0.540$). We also looked separately at positive and negative thoughts and feelings, and there was no correlation between empathic accuracy and mimicry (positive cases: $r=-0.067$, $p=0.44$, negative cases: $r=0.032$, $p=0.67$).

6 Discussion

We explored the use of automated facial expression analysis to measure mimicry and the potential emotional contagion during a user interview as well as how this mimicry may be related to the designer's empathic understanding. This research builds on the early studies in design aiming to measure affective empathy using physiological measurements [11] and begins to study the relationship between empathic mental processes and empathic understanding [9], specifically how both could be measured in a design-relevant way.

We found mimicry between the designer and the user when they were interacting with each other. However, mimicry was detected in smiling, but not in frowning, which is in line with earlier studies [40,48]. The presence of mimicry indicates that the measure can be useful in user-designer interviews. We showed how it is possible to capture at least smiling mimicry via automated facial expression analysis, which simplifies the measurement compared to the use of EMG sensors. This will enable the use of such measures also outside the laboratory and in remote video interviews such as in

this study. However, the ability to detect smiling but not frowning mimicry significantly limits the applicability of the method as it is currently.

We found that users' reported emotional tone does not relate with their smiling or frowning intensity. However, the smiling intensity was related to the designer's prediction of the user's emotional tone. This was interesting since the users' facial expressions did not reveal their reported emotional tone but may have affected the designer's guess. For smiling, this might be due to many factors, including the social norm of smiling politely during a conversation. It should also be noted that the emotional tone is based on three different sources of data: The user likely bases their self-evaluation on their experiences related to the thought or feeling in question, whereas OPENFACE uses only the landmarks on the face, and the designer likely uses a combination of what the user says and how they act at the time possibly relating back to their own experiences as well. Further, while we stayed away from emotion recognition and only focused on the emotional tone, it could be that how we perceive emotions, possibly their tone as well, is culturally dependent [49], which might cause inaccuracies in our study.

Further, while it is an interesting insight into how the designer misunderstands the user's smiling, the results should be ensured with more testing because the results are sensitive toward issues in the video recording. For example, speaking affects the facial expressions because the mouth moves both when smiling and when talking. In fact, this "speaking effect" has been identified by others as well [44,45] and should be further researched. The designer naturally uses the content of the speech to help them guess the users' thoughts and feelings, but the facial expression analysis only focuses on lip movements. Frowning is less sensitive to the speaking effect, but it was not detectable for mimicry (RQ1) and did not show the significance for understanding emotional tone RQ2.

It also remains unclear which way of measuring emotional tone, if any, is relevant in design. Facial expressions are, for the most part, subconscious, and thus, it can be logical that the conscious self-reported tone might be different. It is also logical that the designer may not be able to guess the user's emotional tone (whether measured via facial expressions or self-report), but in practice, the designer is the one aiming to understand the user, and their understanding will impact the design. In psychology, often when physiological synchrony is studied, it sometimes [50] but not often [42] compared with self-rated assessment, leaving it unknown if the conscious and subconscious parts relate.

We found only a weak connection between facial expression mimicry and designers' empathic understanding of emotional tone and only for positively reported entries. We find that during positively reported thoughts and feelings, the designer was more likely to correctly guess the users' emotional tone if they both were smiling than when they were not. However, this occurred only during positively reported thoughts and feelings, which might be because mimicry was not detectable for negative emotions. Interestingly the designer, independent of mimicry, is able to guess the negative tone more often (39% versus 27%), but when there is high mimicry between the user and the designer, the designer is able to guess the positive feeling more often correctly (50%). This could also be due to smiling causing the designer to more often to guess that the user is feeling positive.

The difference between positive and negative emotional tone is another potential area for future research identified in this current work. In design, we often look for the so-called pain points and aim to solve user problems, but our study shows that we might be better able to detect and sometimes understand the user's positive experience. We found that designers' ability to correctly identify emotional tone depended on mimicry, which we were only able to capture for smiling. This study explored users' driving experience under particularly difficult scenarios such as bad weather and poor visibility, with the expectation of finding those pain points. Despite the focus on possibly negative emotions, only the positive emotions showed detectable mimicry. It also could be

due to us not limiting our participants to particularly expressive people as done in some studies [42]. Overall, the mimicry and the difference between positive and negative thoughts and feelings should be studied further.

We also found that the mimicry between the user and the designer does not imply that the designer understands what the user is thinking or feeling. This indicates the affective and cognitive empathy, as measured in this study, may not be directly linked when forming user understanding. We found a relationship with the more abstract form of cognitive empathy, emotional tone accuracy, but not empathic accuracy of thoughts and feelings. However, these results should be verified with more data analysis because of the unbalanced nature of the data: the empathic accuracy data had a majority of cases where the empathic accuracy was zero, and the designer's guess was different from the original thought or feeling. In an early study, we found a strong correlation between designers' emotional tone accuracy and their performance in identifying needs that users feel are important and where they feel satisfied with a solution. These are related to users' positive judgments [51]. Affective empathy is more subconscious and comes from human instinct, while the cognitive part involves an active reformulation of the recognized emotion [8]. Mimicry, as a subconscious physical element, seems to relate to designers' guess of emotional tone. Whether it has a role in designers' performance in other design steps remains unclear. Also in earlier research, the relationship between mimicry and cognitive empathy has not been clear [11,17,52], and it may depend on other interpersonal factors [52,53]. Further, it remains unclear which aspect of empathy or empathic understanding, if any, is related to better design performance, although an earlier work identified a link between emotional tone recognition and need identification [51].

Another potential avenue for future research is to explore alternative measures of mimicry and compare them with empathic understanding. In this study, we measured action units approximating the zygomaticus major muscle and corrugator supercilii muscles, since they are usually measured in similar studies. This enabled us to compare the results with the past EMG studies, but action unit-based facial recognition also enables other ways to measure emotions and empathy. For example, current machine learning models have enabled emotional tone estimations from video recordings. These kinds of measures could bring more insights into the interviews and their dynamics. Further, other types of mimicry, such as gross body movements or heart rate variability [10], have also been shown to be relevant for empathy in neuroscience and psychology, so perhaps other types of mimicry than facial expression mimicry are more relevant in design. This study is only a beginning in building understanding what type of measure and what type of empathic mental processes are or are not meaningful to measure in design.

7 Conclusion

In summary, we explored the use of automated facial expression analysis to measure user-designer empathy from video interviews. We find that the method shows some promise, since there is synchrony between the two, possibly indicating empathy. However, this synchrony was detectable only during smiling, not frowning, and this automatic affective empathy does not fully reflect the designers' empathic understanding, which is a form of cognitive empathy. Further, we find that the user's self-reported emotional tone does not match the one estimated by the facial expression analysis or the designer, which might be due to several factors and thus remains an area for future research.

Acknowledgement

This study was possible thanks to the following funding sources: Technology Industries of Finland Centennial Foundation, Jane and Aatos Erkko Foundation, and the Chinese Scholarship

Council. The authors would like to thank all participants for their time, Dai Sheng's help with statistics and effort, as well as the entire Empathic Engineers team for continued feedback.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

- [1] Cooper, R. G., and Kleinschmidt, E. J., 1987, "Success Factors in Product Innovation," *Ind. Mark. Manag.*, **16**(3), pp. 215–223.
- [2] Saunders, M. N., Seepersad, C. C., and Hölttä-Otto, K., 2011, "The Characteristics of Innovative, Mechanical Products," *ASME J. Mech. Des.*, **133**(2), p. 021009.
- [3] Hölttä-Otto, K., Otto, K., Song, C., Luo, J., Li, T., Seepersad, C. C., and Seering, W., 2018, "The Characteristics of Innovative, Mechanical Products—10 Years Later," *ASME J. Mech. Des.*, **140**(8), p. 084501.
- [4] So, C., and Joo, J., 2017, "Does a Persona Improve Creativity?," *Des. J.*, **20**(4), pp. 459–475.
- [5] Johnson, D. G., Genco, N., Saunders, M. N., Williams, P., Seepersad, C. C., and Hölttä-Otto, K., 2014, "An Experimental Investigation of the Effectiveness of Empathic Experience Design for Innovative Concept Generation," *ASME J. Mech. Des.*, **136**(5), p. 051009.
- [6] Raviselvam, S., Hwang, D., Camburn, B., Sng, K., Hölttä-Otto, K., and Wood, K. L., 2022, "Extreme-User Conditions to Enhance Designer Empathy and Creativity: Applications Using Visual Impairment," *Int. J. Des. Creat. Innov.*, **10**(2), pp. 75–100.
- [7] Chang-Arana, Á., Surma-aho, A., Hölttä-Otto, K., and Sams, M., 2022, "Under the Umbrella: Components of Empathy in Psychology and Design," *Des. Sci.*, **8**, p. E20.
- [8] Cuff, B. M., Brown, S. J., Taylor, L., and Howat, D. J., 2016, "Empathy: A Review of the Concept," *Emot. Rev.*, **8**(2), pp. 144–153.
- [9] Surma-Aho, A., and Hölttä-Otto, K., 2022, "Conceptualization and Operationalization of Empathy in Design Research," *Des. Stud.*, **78**, p. 101075.
- [10] Levenson, R. W., and Ruef, A. M., 1992, "Empathy: A Physiological Substrate," *J. Pers. Soc. Psychol.*, **63**(2), p. 234.
- [11] Chang-Arana, Á., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., and Hölttä-Otto, K., 2020, "Empathic Accuracy in Design: Exploring Design Outcomes Through Empathic Performance and Physiology," *Des. Sci.*, **6**(E16).
- [12] Neumann, D. L., Chan, R. C. K., Boyle, G. J., Wang, Y., and Rae Westbury, H., 2015, "Measures of Empathy: Self-Report, Behavioral, and Neuroscientific Approaches," *Measures of Personality and Social Psychological Constructs*, G. J. Boyle, D. H. Saklofske, and G. Matthews, eds., Academic Press, London, UK, pp. 257–289.
- [13] Decety, J., and Jackson, P. L., 2004, "The Functional Architecture of Human Empathy," *Behav. Cogn. Neurosci. Rev.*, **3**(2), pp. 71–100.
- [14] Kerem, E., Fishman, N., and Josselson, R., 2001, "The Experience of Empathy in Everyday Relationships: Cognitive and Affective Elements," *J. Soc. Pers. Relatsh.*, **18**(5), pp. 709–729.
- [15] Frith, C., and Frith, U., 2005, "Theory of Mind," *Curr. Biol.*, **15**(17), pp. R644–R645.
- [16] Lamm, C., Batson, C. D., and Decety, J., 2007, "The Neural Substrate of Human Empathy: Effects of Perspective-Taking and Cognitive Appraisal," *J. Cogn. Neurosci.*, **19**(1), pp. 42–58.
- [17] Holland, A. C., O'Connell, G., and Dziobek, I., 2020, "Facial Mimicry, Empathy, and Emotion Recognition: A Meta-Analysis of Correlations," *Cogn. Emot.*, **35**(1), pp. 1–19.
- [18] Apfelbaum, M., Sharp, K., and Dong, A., 2021, "Exploring Empathy in Student Design Teams," *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Virtual, Online, Aug. 17–19, Vol. 85406, p. V004T04A002.
- [19] Surma-Aho, A. O., Björklund, T. A., and Holttä-Otto, K., 2018, "Assessing the Development of Empathy and Innovation Attitudes in a Project-Based Engineering Design Course," 2018 ASEE Annual Conference & Exposition, Salt Lake City, UT, June 23–27.
- [20] Alzayed, M. A., McComb, C., Menold, J., Huff, J., and Miller, S. R., 2021, "Are You Feeling Me? An Exploration of Empathy Development in Engineering Design Education," *ASME J. Mech. Des.*, **143**(11), p. 112301.
- [21] Drimalla, H., Landwehr, N., Hess, U., and Dziobek, I., 2019, "From Face to Face: The Contribution of Facial Mimicry to Cognitive and Emotional Empathy," *Cogn. Emot.*, **33**(8), pp. 1672–1686.
- [22] Schirmer, A., and Adolphs, R., 2017, "Emotion Perception From Face, Voice, and Touch: Comparisons and Convergence," *Trends Cogn. Sci.*, **21**(3), pp. 216–228.
- [23] Ickes, W., 1993, "Empathic Accuracy," *J. Pers.*, **61**(4), pp. 587–610.

- [24] Li, J., Surma-Aho, A., Chang-Arana, ÁM, and Hölttä-Otto, K., 2021, "Understanding Customers Across National Cultures: The Influence of National Cultural Differences on Designers' Empathic Accuracy," *J. Eng. Des.*, **32**(10), pp. 538–558.
- [25] Li, J., and Hölttä-Otto, K., 2020, "The Influence of Designers' Cultural Differences on the Empathic Accuracy of User Understanding," *Des. J.*, **23**(5), pp. 779–796.
- [26] Li, J., Surma-Aho, A., and Hölttä-Otto, K., 2021, "Measuring Designers' Empathic Understanding of Users by a Quick Empathic Accuracy (QEA)," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Virtual, Online, Aug. 17–19, Vol. 85420. American Society of Mechanical Engineers, p. V006T06A027..
- [27] Kleinbub, J. R., Palmieri, A., Orsucci, F. F., Andreassi, S., Musmeci, N., Benelli, E., Giuliani, A. and de Felice, G., 2019, "Measuring Empathy: A Statistical Physics Grounded Approach," *Physica A*, **526**, p. 120979.
- [28] Chartrand, T. L., and Bargh, J. A., 1999, "The Chameleon Effect: The Perception–Behavior Link and Social Interaction," *J. Pers. Soc. Psychol.*, **76**(6), pp. 893–910.
- [29] Hatfield, E., Cacioppo, J. T., and Rapson, R. L., 1994, *Emotional Contagion*, Cambridge University Press; Editions de la Maison des Sciences de l'Homme, Cambridge, UK.
- [30] Van Baaren, R. B., Holland, R. W., Kawakami, K., and van Knippenberg, A., 2004, "Mimicry and Prosocial Behavior," *Psychol. Sci.*, **15**(1), pp. 71–74.
- [31] Hess, U., and Blairy, S., 2001, "Facial Mimicry and Emotional Contagion to Dynamic Emotional Facial Expressions and Their Influence on Decoding Accuracy," *Int. J. Psychophysiol.*, **40**(2), pp. 129–141.
- [32] Hinsz, V. B., and Tomhave, J. A., 1991, "Smile and (Half) the World Smiles With You, Frown and You Frown Alone," *Pers. Soc. Psychol. Bull.*, **17**(5), pp. 586–592.
- [33] Lundqvist, L.-O., and Dimberg, U., 1995, "Facial Expressions Are Contagious," *J. Psychophysiol.*, **9**(3), pp. 203–211.
- [34] Ekman, P., and Rosenberg, E. L., 2005, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, 2nd ed., Oxford University Press, Oxford, UK.
- [35] Ekman, P., Freisen, W. V., and Ancoli, S., 1980, "Facial Signs of Emotional Experience," *J. Pers. Soc. Psychol.*, **39**(6), pp. 1125–1134.
- [36] Baltrušaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P., 2018, "OpenFace 2.0: Facial Behavior Analysis Toolkit," IEEE International Conference on Automatic Face and Gesture Recognition, Xi'an, China, May 15–19.
- [37] Baltrušaitis, T., Mahmoud, M., and Robinson, P., 2015, "Cross-Dataset Learning and Person-Specific Normalisation for Automatic Action Unit Detection," IEEE International Conference on Automatic Face and Gesture Recognition, Ljubljana, Slovenia, May 4–8.
- [38] Hofelich, A. J., and Preston, S. D., 2012, "The Meaning in Empathy: Distinguishing Conceptual Encoding From Facial Mimicry, Trait Empathy, and Attention to Emotion," *Cogn. Emot.*, **26**(1), pp. 119–128.
- [39] Dimberg, U., Andréasson, P., and Thunberg, M., 2011, "Emotional Empathy and Facial Reactions to Facial Expressions," *J. Psychophysiol.*, **25**(1), pp. 26–31.
- [40] Deng, H., and Hu, P., 2018, "Matching Your Face or Appraising the Situation: Two Paths to Emotional Contagion," *Front. Psychol.*, **8**, p. 2278.
- [41] Howard, D. J., and Gengler, C., 2001, "Emotional Contagion Effects on Product Attitudes," *J. Consumer Res.*, **28**(2), pp. 189–201.
- [42] Soto, J. A., and Levenson, R. W., 2009, "Emotion Recognition Across Cultures: The Influence of Ethnicity on Empathic Accuracy and Physiological Linkage," *Emotion*, **9**(6), pp. 874–884.
- [43] Marur, T., Tuna, Y., and Demirci, S., 2014, "Facial Anatomy," *Clin. Dermatol.*, **32**(1), pp. 14–23.
- [44] Gray, H., 1918, *Anatomy of the Human Body*, Lea & Febiger, Philadelphia, PA, Bartleby.com, 2000. www.bartleby.com/107/
- [45] Riehle, M., Kempkensteffen, J., and Lincoln, T. M., 2017, "Quantifying Facial Expression Synchrony in Face-to-Face Dyadic Interactions: Temporal Dynamics of Simultaneously Recorded Facial EMG Signals," *J. Nonverbal Behav.*, **41**(2), pp. 85–102.
- [46] Quade, D., 1967, "Rank Analysis of Covariance," *J. Am. Stat. Assoc.*, **62**(320), pp. 1187–1200.
- [47] Schwenke, J. R., 1997, "Comparing the Use of Block and Covariate Information in Analysis of Variance," *Conference on Applied Statistics in Agriculture*, Manhattan, KS, Apr. 27–29.
- [48] Sullins, E. S., 1991, "Emotional Contagion Revisited: Effects of Social Comparison and Expressive Style on Mood Convergence," *Pers. Soc. Psychol. Bull.*, **17**(2), pp. 166–174.
- [49] Gendron, M., Roberson, D., van der Vyver, J. M., and Barrett, L. F., 2014, "Perceptions of Emotion From Facial Expressions Are Not Culturally Universal: Evidence From a Remote Culture," *Emotion*, **14**(2), pp. 251–262.
- [50] Lee, J., Zaki, J., Harvey, P. O., Ochsner, K., and Green, M., 2011, "Schizophrenia Patients Are Impaired in Empathic Accuracy," *Psychol. Med.*, **41**(11), pp. 2297–2304.
- [51] Li, J., and Hölttä-Otto, K., 2022, "Does Empathising With Users Contribute to Better Need Finding?" International Design Engineering Technical Conferences, Conference on Design Theory and Methodology, St. Louis, MO, Aug. 14–17, Paper No. DETC2022-89413.
- [52] Zaki, J., Bolger, N., and Ochsner, K., 2008, "It Takes Two: The Interpersonal Nature of Empathic Accuracy," *Psychol. Sci.*, **19**(4), pp. 399–404.
- [53] Li, J., and Holttä-Otto, K., 2023, "Inconstant Empathy–Interpersonal Factors That Influence the Incompleteness of User Understanding," *ASME J. Mech. Des.*, **145**(2), p. 021403.