FINAL REPORT — Facial and bodily temperature maps of emotions – 279/14

Lúcia Garrido (PI) and Nicholas Pound (CI)

1. Background and aims

William James (1884) argued that emotions are the feelings of bodily change, which happen in response to things, ideas, or other people. According to this view, different emotions are associated with distinct bodily changes or physiological responses. For the past 100 years, however, the question of whether the physiological responses that accompany emotions are specific for each emotion or are instead generalized arousal responses has remained unsettled, and it continues to attract interest from physiologists, psychologists, and neuroscientists.

Several studies have investigated whether there are physiological responses that are specific to certain emotions, and while some measures seem to consistently distinguish between emotions, the results are not always clear and continue to be debated intensely (e.g. Ekman et al., 1983; Feldman-Barrett, 2006; Kreibig, 2010; Rainville et al., 2006; Shacter & Singer, 1982). Most studies, however, have focused their investigation on 'overall' physiological responses (such as heart rate or respiration rate) or physiological changes in single points in the body (such as temperature changes in a finger).

It is possible, however, that we need much richer datasets to characterize physiological responses in the face and body and only spatial and temporal patterns of responses will distinguish between emotional states (Kreibig, 2010; Rainville et al., 2006). A recent paper addressed this issue by using an interesting and novel method that measured subjective sensations in the full body, thus allowing the creation of a rich spatial pattern of responses (Nummenmaa et al., 2014). These authors asked participants to color body silhouettes that either felt activated or deactivated in response to emotions. They found that participants' bodily sensation maps were consistent across people and types of stimuli used to elicit emotions.

Nummenmaa et al.'s (2014) results were based on subjective reports of physiological changes, so a crucial outstanding question is whether it is possible to *objectively* measure the patterns of bodily sensations. In our project, we aimed to do that by investigating the spatial distribution of physiological responses to different emotions across the face and body. More specifically, we aimed to use skin temperature changes of the whole face and the whole body and investigate whether those temperature change maps are distinct for different emotions. Thermal imaging, which measures emitted infrared heat, seemed ideally suited to address our aims. It has only been recently applied to research in psychology and physiology, but has already shown to be a valuable means to measure autonomic responses in emotional states (Ioannu et al., 2014). In particular, changes in temperature captured by thermal imaging can be related to vascular changes, perspiration, or muscular activity associated with emotional states.

We thus proposed to use thermal imaging to capture temperature changes in response to each of six basic emotions (happiness, fear, surprise, anger, disgust, and sadness), compared to a baseline. We then planned to investigate both maps of *perceiving* and *experiencing* emotions. We proposed to develop innovative methods to analyze these maps and determine their overlap. If different emotions are associated with different physiological responses, the temperature change maps should be distinct for different emotions — and therefore allow for the classification of emotional states based on thermal images. If, however, we only have generalized arousal responses to emotions, the temperature change maps should be completely overlapping for all emotions.

2. Summary of completed activities and overall progress

Our funded project ran from the end of January 2015 to the end of May 2017, but the PI was on maternity leave from August 2015 to January 2016. We obtained ethical approval for the studies in January 2015. Our experiments involved thermal imaging and we did not have experience with this method before the project started. We purchased a thermal camera in March 2015, obtained training on its use (April 2015) and completed a literature review of studies using thermal cameras in July 2015. Since then, we carried out three main experiments that involved collecting and analyzing a rich dataset of more than 2000 facial thermal images exposed to a range of emotional stimuli.

Experiments 2 and 3 involved eliciting emotions in participants, and it was important to conduct pilot work to identify procedures that reliably elicited emotions, preferably with high intensity. We conducted a literature review of methods for eliciting emotions (completed by March 2016). This review of published findings revealed that there was not sufficient evidence that stimuli from existent databases specifically elicited the target emotions with high intensity and with higher intensities than other emotions. Therefore, we conducted a number of pilot studies to determine the best procedures and stimuli for the main experimental work. We tested the use of (1) stories and personal experiences, (2) pictures from the Geneva Affective Picture Database, the Open Affective Standardized Image Set, and the Nencki Affective Picture System, (3) videos from Tettamanti et al. (2012), and (4) videos from Schaefer et al. (2010). In all these validation studies (completed by June 2016), each participant rated one stimulus at a time for each of six emotional labels. We selected the stimuli that consistently elicited each target emotion with high intensity and with specificity (i.e. with higher intensity than all the other emotions). Based on these pilot studies, we were able to select suitable stimuli for experiments 2 and 3.

We designed experiments 1 and 2, and collected data for these experiments from July 2016 to September 2016. We developed the methods to analyze the data on the spatial patterns of temperature change in the face and completed these by December 2016. These methods are one of the main innovations and an important contribution of our project, and we describe these methods in more detail below (please see section 'General Methods'). One crucial aspect of our methods is that they rely on the delineation of landmarks in each thermal image, so that we can establish correspondence of anatomical features in the face across time for the same participant and across participants. These delineations have taken considerably more time than we had initially planned. Typically, it takes between 10 to 15 hours to delineate relevant facial landmarks in all the thermal images for each individual participant tested. Our studies had about 20 participants each, so it took between 200 and 300 hours to delineate the data for each experiment. Consequently, delineations for experiments 1 and 2 were only completed in June 2017. We designed experiment 3 and collected data from April 2017 to May 2017, and completed the data analysis in July 2017. We described below the results that we obtained with experiments 1, 2, and 3. Our analyses so far indicate that we have obtained some interesting and promising results, and we discuss the results, progress, and future directions at the end of this report.

We had several research assistants working part-time on the project along the PI and CI. Lisa Kuhn worked from March to August 2015. Christine Girges worked in March 2016, and Chiara Casella worked from May to September 2016. All these research assistants gave fantastic support to the project, but they all left the project to start positions that were full time and that were thus better for their careers. Chiara, in particular, contributed greatly to the advance of the project, conducting most of the testing for the pilot studies, and testing all the participants for experiments 1 and 2 (22 participants in each experiment). After September 2016, we no longer had funds to support a dedicated research assistant. Instead, we employed undergraduate students as hourly-paid research assistants for 5 to 10 hours per week from the Autumn 2016 until the Summer 2017. These students were Emily Mitson and Andrew Phillips-Hird and they worked on delineating landmarks in the thermal images.

The PI and the CI worked on the experimental design of all the experiments, programmed all experiments, trained and supervised the work of the research assistants, developed the methods to analyze the thermal images, and analyzed all the data after delineation of landmarks. For experiment 3, the PI also collected all the data and carried out most of the delineations of the facial images.

3. General Methods

Experimental design and thermal imaging

Each experiment was divided in runs of about 10 to 15 minutes, and all conditions were always presented in each run. In each run, each condition was presented in a block of one minute. An emotional condition block (consisting of presentation of visual stimuli) was always preceded by a control condition block (consisting of presentation of objects). The aim of having this control condition block was to estimate the baseline temperature, and allow for temperature to return to baseline after an emotion condition was presented. There was also a control condition block at the beginning and at the end of each run.

All our experiments used a thermal camera (FLIR A655sc) to measure temperature in the face while participants were presented with pictures or videos (see experimental setup in Figure 1). We recorded the temperature of the face of the participant every second, from the start until the end of each run. To avoid participants moving their face, we used a chin rest specially built for this project (it only includes a chin piece to avoid covering the face).

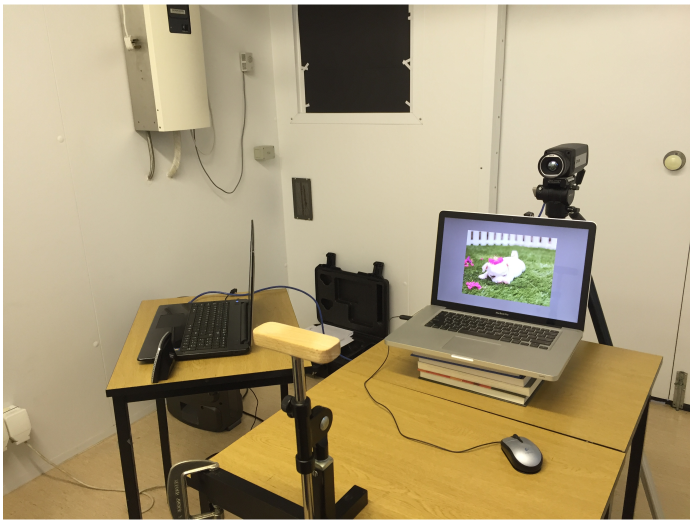

Figure 1: This was our experimental setup in the environmental controlled chamber at Brunel University London (we can control temperature and humidity). Participants sat with their chin on the chin rest, and the thermal camera recorded temperature above the laptop that presented the stimuli.

Data Analysis

In all three experiments, we analyzed only the thermal images taken at the end of each condition block (either emotional or control blocks). Therefore, so far, we have only analyzed the pictures taken every minute. Thermal images were saved as .png images (to be able to delineate landmarks in the face), and as Matlab (www.mathworks.com) matrices containing the temperature data for each pixel. On each .png image, we delineated 95 anatomical landmarks and semi-landmarks using Psychomorph (Tiddeman, Stirrat, & Perrett, 2005). All the delineations were carried out blind to the condition that the image was associated with. Using Matlab, we then used the delineated landmarks to divide the face in 68 polygons with each landmark as a vertex (see Figure 2). Each of these polygons corresponds to a region of interest. We used the same polygons for all participants to allow us to compare results across participants who naturally have different face shapes. We extracted the mean temperature inside each polgygon for each condition. Figure 3 shows an example of the delineation of one thermal image.

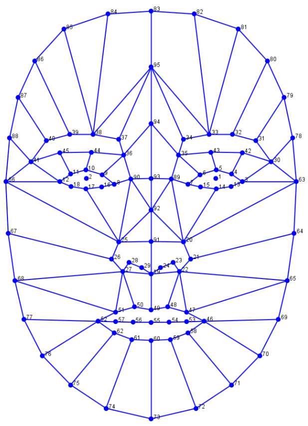

Figure 2: Scheme of the 95 landmarks delineated in each face and the 68 polygons derived from that delineation.

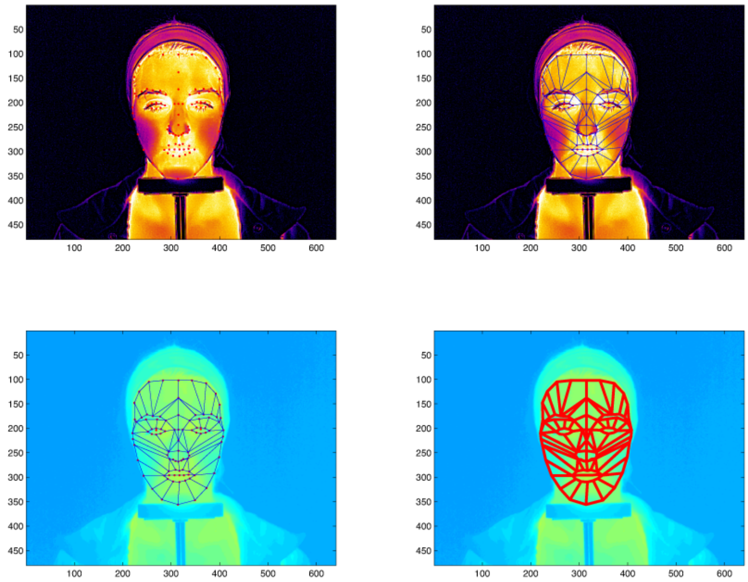

Figure 3: The two top pictures show one .png image that was used to delineate 95 landmarks (top left), and the 68 polygons that constitute our regions of interest (top right). The two bottom pictures show temperature images with the superimposed 95 landmarks (bottom left) and 68 polygons (bottom right).

So far, we have developed two main methods of analyses:

1. Main effects of emotions: For this analysis, we compared the mean temperature within each polygon of the face for each emotion condition with the temperature of that polygon during the baseline block that immediately preceded it. For each participant, we then averaged across repeated blocks for each emotion to create mean temperature face maps for each emotion across all runs. These maps show the increase or decrease in temperature in each polygon compared to baseline. We then tested whether the temperature within each polygon significantly increased or decreased for each emotion using one-sample t-tests at each polygon with the data across all the participants (Figure 4). This analysis allow us to test whether there are specific regions of the face that consistently increase or decrease in temperature depending on the emotion viewed or experienced by the participant.

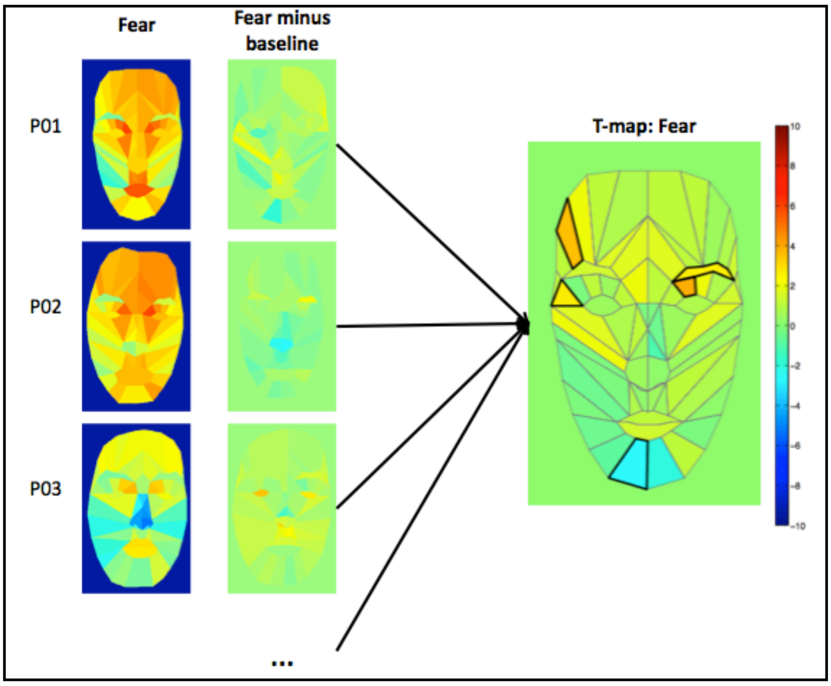

Figure 4: Example of main effect analysis. For each participant, we measure the temperature during an emotion condition, and calculate an average temperature in each polygon (first column). For each run and condition, we subtract the temperature map for the baseline condition, and obtain a map of all the differences compared to baseline. We average those maps for all the runs of the same participant. The second column shows these average

difference maps for three participants for the condition fear in experiment 1. We then compare the temperature within each polygon for all participants (i.e. vector of temperatures for each polygon with one value per participant) to zero using a one sample t-test. The resulting t-test map is shown on the right and the colors inside each polygon represent the resulting t-values (scaled from -10 to 10 according to the color scale on the right). Polygons with a bold line indicate that the t-test was significant ($p < .05$).

2. Classification of emotions: For this analysis, we investigated the spatial distribution of the temperature change associated with each condition and borrowed procedures from functional magnetic resonance imaging (fMRI) analyses (Haxby et al., 2001; Kriegeskorte et al., 2008). Specifically, we used classification analyses like the ones described in Haxby et al. (2001). We reasoned that, if the facial temperature maps are distinct for different emotions, then the correlation of two temperature patterns for the same emotion should be more similar across runs than the correlation across the patterns corresponding to different emotions (Haxby et al., 2001). For each condition, we analyzed the temperature data from the end of the block. We therefore extracted a vector of temperatures (i.e. vector consisting of the temperature in each of the 68 polygons) for each emotion. We then computed the correlations between vectors from two separate runs and used the same classification procedure as Haxby et al. (2001). Briefly, if the correlation between two patterns corresponding to the same emotion (e.g. fear run 1 and fear run 2) was higher than the correlation of one emotion with each of the other emotions (e.g. fear run 1 with happiness run 2, fear run 1 with anger run 2, etc), the classification would be one; otherwise, the classification result would be zero. We then looked at whether the classification of each emotion was significantly above chance across participants (by conducting a one sample t-test on the classification values comparing to chance performance).

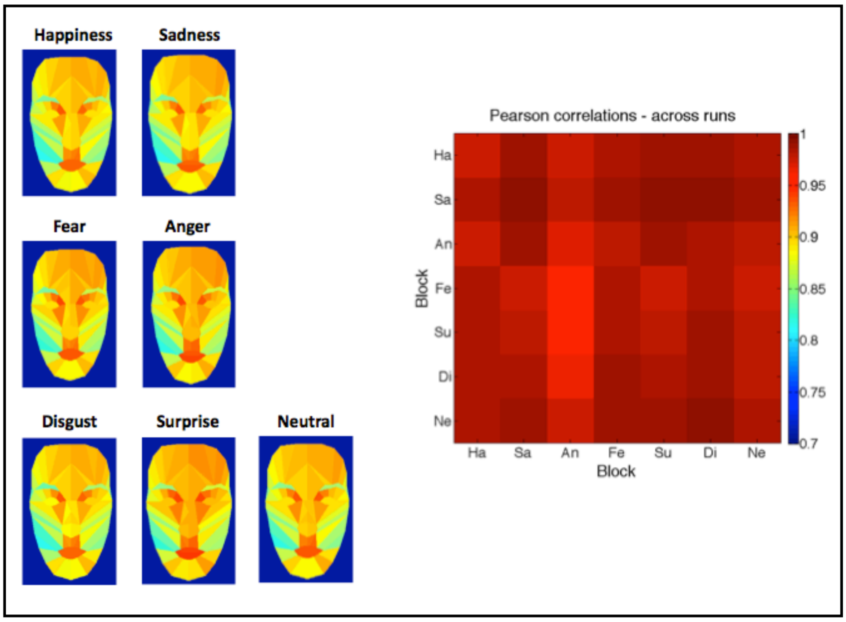

Figure 5: Example of classification analysis. The face maps on the left show the temperature maps for each emotional condition for one run of one participant. The matrix on the right shows the Pearson correlations of the temperature maps across two runs for the same participant. Each cell shows the correlation between each pair of emotions (Ha = happiness, Sa= sadness, An = anger, Fe = fear, Su = surprise, Di = disgust, Ne = neutral). For successful classification, we expected that the correlation of the temperature maps for the same emotion across runs (e.g. fear map run 1 vs fear map run 2) was higher than the correlations between maps of different emotions (e.g. fear map run 1 vs anger map run 2).

4. Experiment 1

Aims: We aimed to (1) create and characterize facial temperature change maps in response to *perceiving* each of the six basic emotions in the faces of others, and (2) determine whether those facial maps are distinct for different emotions.

Methods: Participants were presented with pictures of facial expressions of emotion (happiness, surprise, sadness, anger, disgust, fear) and neutral expressions from validated databases. We used a block design. Each block (lasting 1 minute) presented 10 different pictures showing the same expression, each picture presented for six seconds. We alternated blocks with faces with blocks with neutral objects (also ten different pictures of objects, each presented for six seconds), so that participants' temperature could return to a 'baseline state'. One run showed one block for each condition (facial expression), each preceded and followed by one block with objects. Expressions were presented in randomized order, and each participant completed three runs.

Data analysis: For each condition in each run, we analyzed the temperature data from the end of the block. We then conducted the analyses of main effects of each emotion plus classification of emotion maps — please see more detailed description in section 3.

Results: Figure 6 shows the results of the main effects analyses of each emotion. Our results showed that there were consistent effects of increased or decreased temperature in certain regions of the face for each emotion. For example, for fear, the temperature in several of the polygons increased compared to baseline, and some of these changes were significant (all $p< .05$ shown in the figure with lines marked in bold). In contrast, for disgust, the temperature seemed to decrease compare to baseline and many of these changes were significant.

Figure 7 shows the results of the classification analysis and we did not find that the classification of any of the emotions was significantly above chance. We discuss potential explanations for this in the final section of this report.

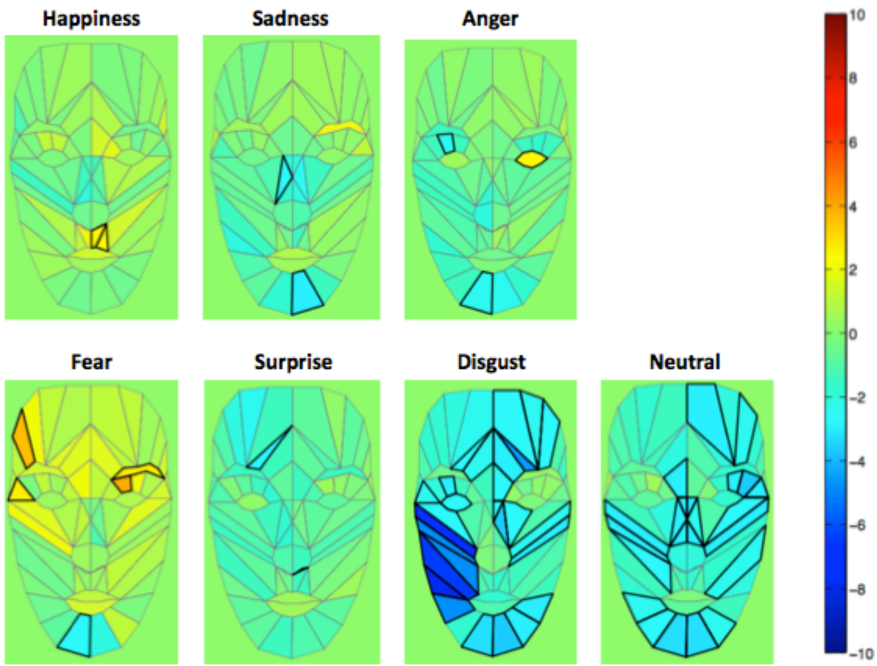

Figure 6: Main effects across participants for experiment 1. Each picture shows a map of t-values, with the color indicating the t-value (according to the color scale on the right). Polygons that have bold black lines indicate that the t-test within that polygon was significant ($p < .05$).

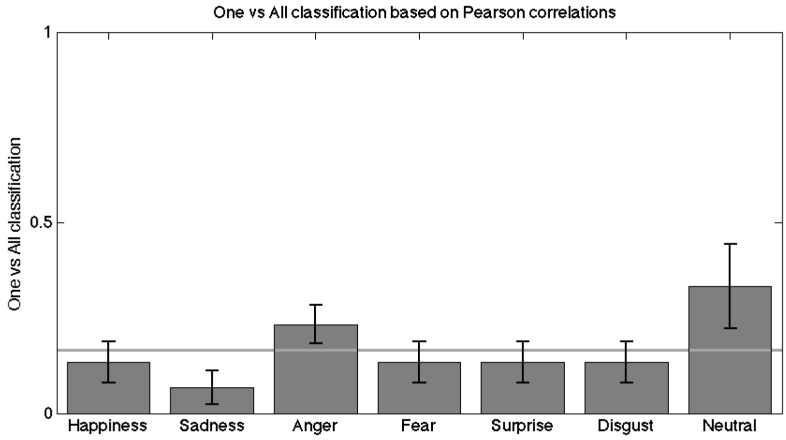

Figure 7: Classification results for experiment 1. The grey line indicates chance performance. Each bar graph shows the mean classification of each emotion. Error bars represent standard error of the mean.

5. Experiment 2

Aims: We aimed to (1) create and characterize facial temperature change maps in response to *experiencing* emotions elicited by viewing *pictures*, and (2) determine whether those facial maps are distinct for different emotions.

Methods: Our validation studies did not find sufficient pictures that consistently and specifically elicited anger, surprise, and fear. Therefore, we only used pictures that elicited happiness, sadness, and disgust, plus neutral pictures. Using a block design, participants were presented with pictures eliciting each of the emotional states. Each block (lasting 1 minute) presented 10 different pictures that elicited the same emotion, each picture presented for six seconds. We alternated blocks with emotional pictures with blocks with neutral objects (also ten different pictures of objects, each presented for six seconds), so that participants' temperature could return to a 'baseline state'. One run showed one block for each condition (emotion), each preceded and followed by one block with objects. Emotions were presented in randomized order, and each participant completed four runs.

Data analysis: For each condition in each run, we analyzed the temperature data from the end of the block. We then conducted the analyses of main effects of each emotion plus classification of emotion maps — please see more detailed description in section 3.

Results: Figure 8 shows the results of the main effects analyses of each emotion. Our results showed that there were consistent effects of increased or decreased temperature in certain regions of the face for each emotion. For example, for sadness, the temperature in many of the polygons increased decreased to baseline, and some of these changes were significant (all $p < .05$ shown in the figure with lines marked in bold). In contrast, for disgust, the temperature seemed to show both increases and decreases, with the lips showing a marked increase in temperature.

Figure 9 shows the results of the classification analysis and we did not find that the classification of any of the emotions was significantly above chance. We discuss potential explanations for this in the final section of this report.

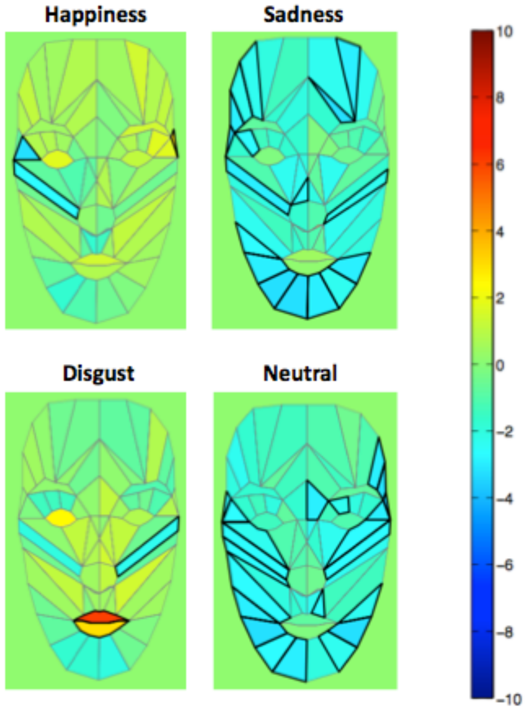

Figure 8: Main effects across participants for experiment 2. Each picture shows a map of t-values, with the color indicating the t-value (according to the color scale on the right). Polygons that have bold black lines indicate that the t-test within that polygon was significant ($p < .05$).

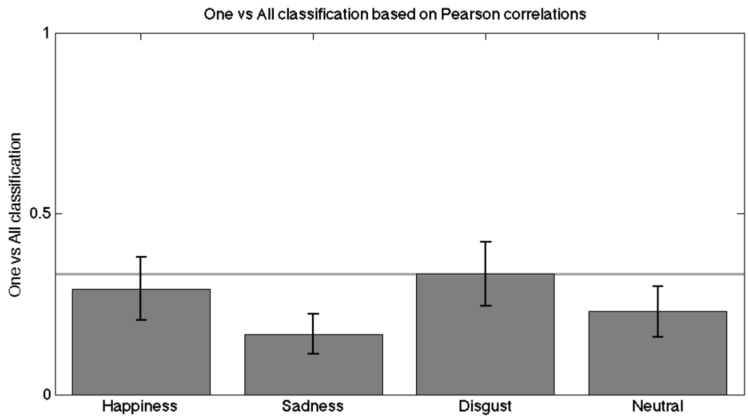

Figure 9: Classification results for experiment 2. The grey line indicates chance performance. Each bar graph shows the mean classification of each emotion. Error bars represent standard error of the mean.

6. Experiment 3

Aims: We aimed to (1) create and characterize facial temperature change maps in response to *experiencing* emotions elicited by watching *videos*, and (2) determine whether those facial maps are distinct for different emotions.

Methods: Our validation studies did not find sufficient videos that consistently elicited anger and surprise. Therefore, we only used videos that elicited happiness, sadness, fear, and disgust, plus neutral videos. Using a block design, participants were presented with videos eliciting each of the emotional states. Each block (lasting 1 minute) presented 6 different 10 second videos that elicited the same emotion. We alternated blocks with videos with blocks with neutral objects (ten different pictures of objects, each presented for six seconds), so that participants' temperature could return to a 'baseline state'. One run showed one block for each condition (emotion), each preceded and followed by one block with objects. Emotions were presented in randomized order, and each participant completed two runs.

Data analysis: For each condition in each run, we analyzed the temperature data from the end of the block. We then conducted the analyses of main effects of each emotion plus classification of emotion maps — please see more detailed description in section 3.

Results: Figure 10 shows the results of the main effects analyses of each emotion. Our results showed that there were consistent effects of increased or decreased temperature in certain regions of the face for each emotion. For example, for fear, the temperature in several of the polygons increased compared to baseline, and changes in the lips were significant (all $p < .05$ shown in the figure with lines marked in bold). For disgust, there was also an increase in temperature in the lips compared to baseline. In contrast with experiments 1 and 2, there were marked increases in the temperature also for neutral conditions.

Figure 11 shows the results of the classification analysis and we did not find that the classification of any of the emotions was significantly above chance. We discuss potential explanations for this in the final section of this report.

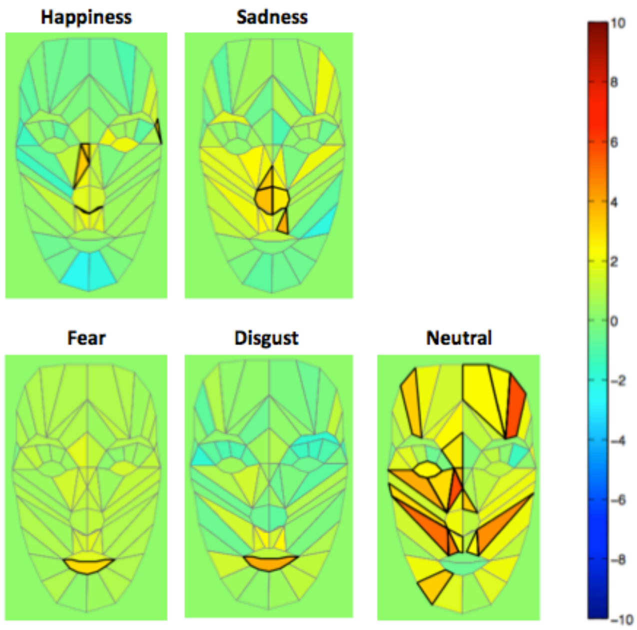

Figure 10: Main effects across participants for experiment 3. Each picture shows a map of t-values, with the color indicating the t-value (according to the color scale on the right). Polygons that have bold black lines indicate that the t-test within that polygon was significant ($p < .05$).

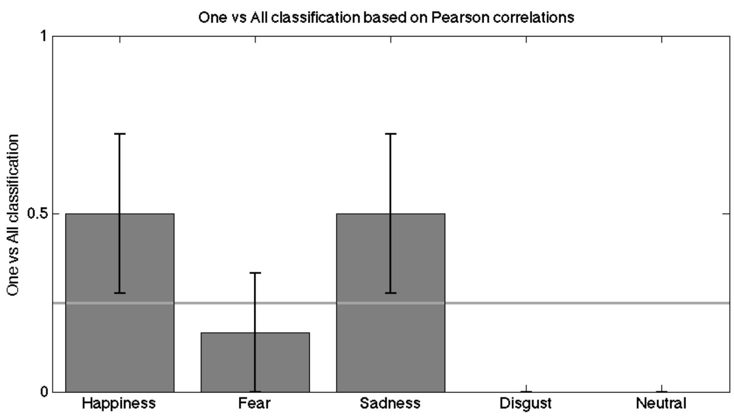

Figure 11: Classification results for experiment 3. The grey line indicates chance performance. Each bar graph shows the mean classification of each emotion. Error bars represent standard error of the mean.

7. Discussion of results, progress, and future directions

Our analyses of main effects in all three experiments showed that there are some regions in the face that consistently increase or decrease in temperature in response to perceiving or experiencing emotions. These results are interesting in that they show that it may be possible to observe physiological changes in response to emotions using our paradigm. These results raise a number of interesting questions regarding the specific regions that showed changes in temperature. For example, what are the physiological mechanisms that explain these changes and what is the time course of these changes (e.g. Do we need one minute of stimulation, or can we see these changes immediately after stimulus presentation? Could the time course itself be used to decode emotions?).

Our results, however, did not show consistent changes in temperature across experiments. For example, the temperature changes that we observed for fear in experiment 1 were not observed in experiment 3. This could be due to the difference between *perceiving* emotional expressions in experiment 1 versus *experiencing* emotions in experiment 3. But we also did not find consistent changes in emotional maps for experiments 2 and 3 (apart from an increase in the temperature in the lips for disgust), even if both were related to experiencing emotions. This could be due to the use of different stimuli. Nevertheless, we believe that the inconsistency of changes across different experiments needs to be treated cautiously, and that we will need first to replicate the effects that we have found so far before being confident that they are real and reliable. We may also need to increase our sample sizes to have more power to detect these differences.

Regarding our classification analyses, we did not observe any significant classification in any of the experiments. However, one possible explanation could be that we are using too many regions of interest (i.e polygons) in our analyses. Some of these regions carry out information about the emotion, but other do not, and instead just add noise to our estimates. One way of dealing with this would be to use feature selection procedures. Feature selection is often used in machine learning for similar problems, and we plan to implement that in the analyses of our experiments.

Regarding our progress, we made significant efforts to complete our initial plans. All the research assistants that we recruited were hard-working, competent, and committed to completing their tasks. The PI and CI have also spent many more hours in the project than had initially been planned. We were able to complete all of the empirical work described in our initial funding application involving the analysis of temperature changes in the face. We completed extensive pilot work and three main experiments, which required huge efforts in developing and carrying out the data analyses. However, we were not able to complete the final planned experiment examining temperature changes across the body. We believe that there are two main reasons for this. First, in addition to the extensive pilot research, our analyses of the face images took considerably more time than we had planned. As we described before, the delineations of the faces took between 10 to 15 hours per participant (i.e. around 200 to 300 hours per experiment). This required a huge effort from the student assistants and the PI who carried out this work. Because of all the time spent on these delineations, we did not have time to carry out the experiment looking at responses in the whole body. A second main reason was lack of funding. We had applied for €49,388 to complete all experiments but were only awarded €38.700 (i.e. more than 20% less than requested). Unfortunately this considerable difference of more than €10,000 has made it impossible for us to complete the experiment with bodies. Initially, we hoped that we were still going to be able to do it but we realized that this was impossible without funding for an additional dedicated research assistant who could recruit and test participants, and work on the delineations of the body images. We still think that the experiments with bodies would be very interesting and will pursue alternative funding opportunities to develop this project.

To conclude, we think that our work brings a highly innovative approach to examine physiological signatures of emotions. We used thermal imaging to investigate facial temperature change maps in response to emotions, and we developed novel methods to analyze these maps. So far, we have obtained some evidence that different emotions lead to specific changes in temperature of the face, but more work still needs to be done to assess the reliability of our findings. In any case, we think that our results open a number of interesting possibilities for future research.

We sincerely thank Bial Foundation for the bursary. Without this funding, we would not have been able to carry out any of this work.

References

Ekman, P., Levenson, R., & Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221 (4616)*, 1208-1210.

Feldman-Barrett, L. (2006). Are emotions natural kinds? *Perspectives on Psychological Science, 1(1)*, 28-58.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425-2430.

Ioannu, S., Gallese, V., & Merla, A. (2014). Thermal infrared imaging in psychophysiology: potentialities and limits. *Psychophysiology*, doi: 10.1111/psyp.12243.

James, W. (1884). What is an emotion? *Mind, 9*, 188-205.

Kreibig, S. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*, 394-421.

Kriegeskorte, N., Mur, M., & Bandettinii, P. (2008a). Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience, 2*(4).

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. (2014). Bodily maps of emotions. *Proceedings of the National Academy of Sciences, 111*(2), 646-651.

Rainville, P., Bechara, A., Naqvi, N., & Damasio, A. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology, 61*, 5-18.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69(5)*, 379-399.

Tiddeman, B., Stirrat, M., & Perrett, D. (2005). Towards realism in facial transformation: results of a wavelet MRF method. *Computer Graphics Forum, 24*, 449-456.