What's in a Face? Psychophysiological applications of neuroscience for diagnostics and therapies.

by

Javier A. Elkin

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Neuroscience

January 2018

University College London

Primary Supervisor: Geraint Rees

Secondary Supervisor: Parashkev Nachev

Declaration:

I, Javier A. Elkin, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

Date:

Abstract

The idea that the utility of research should be secondary to understanding its subject delays the extraction of potential value. A parallel translational approach to research was applied whereby discovering new findings and testing their validity was performed in parallel.

Research about the face was selected for translation as it provided the complexity, diversity, and fidelity necessary for multiple data-driven hypothesis exploration while remaining key to social interaction. For example, emotional contagion, the tendency for an individual to catch someone else's emotion has been linked to facial contagion: an automatic reaction whereby facial muscles adopt the expression of any emotional face.

Based on the reported exaggerated emotional reactions of patients with upper involvement in Motor Neuron Disease (MND) compared to lower involvement, an experiment was devised to make the difference through comparisons of facial contagion responses with recorded Electromyography (EMG) responses (chapter 3). As these patients were expected to have generally weak responses, it became necessary to increase the sensitivity of acquired signals to elucidate differences between subtypes. An adaptive filtering technique for signal processing was developed based on modelling methods and tested with support vector machines (chapter 2).

The therapeutic intervention (chapter 4) consisted of a series of experiments seeking to induce emotional contagion of happiness by presenting images of smiling faces through smartphones. This was also gamified in an experiment at the Science Museum in London to test whether the effect could be found over the short term. Lastly, I parametrised faces from a large population of Tibetan residents and predicted haematological and electrocardiographic measures with machine learning methods as a way of screening for cardiovascular disease through photographs of the face (chapter 5).

The results were analysed in relation to the field of cognitive neuroscience and the implications for a parallel translational and high-dimensional approach were discussed.

Table of Contents

List of Figures and Tables

Chapter 1: Foundations

Chapter 2: An adaptive filter to better capture the physiology of facial expression

Chapter 3: Quantifying emotional dysregulation in Motor Neuron Disease with facial electromyography

Chapter 4: Improving affective states through facial contagion

Chapter 5: Wearing your heart on your face: Predicting physiology from physiognomy with machine learning

Chapter 6: General discussion

This work is dedicated to all those patients struggling with Motor Neuron Disease, Depression, and cardiovascular diseases. May those who conduct research always make your wellbeing their priority.

Acknowledgements

To my mother who supported me to begin this adventure and couldn't see it end. To my wife who believed in me and for whom words here cannot describe my appreciation for her patience and support. To my father, family, and friends, who provided the essential emotional and moral support throughout all aspects of life.

I am grateful to both my supervisors who always pushed me to be better and mentored me to succeed as a scientist, a thinker, and innovator.

A big thank you goes to the volunteer team that worked on the Science Museum experiment without whose dedication, passion, and inspiration it would have been impossible to succeed. A special thank you to Filippo Aiello for developing and designing the tablet experiment and to Ben Howitt for his help uncovering patterns in the data.

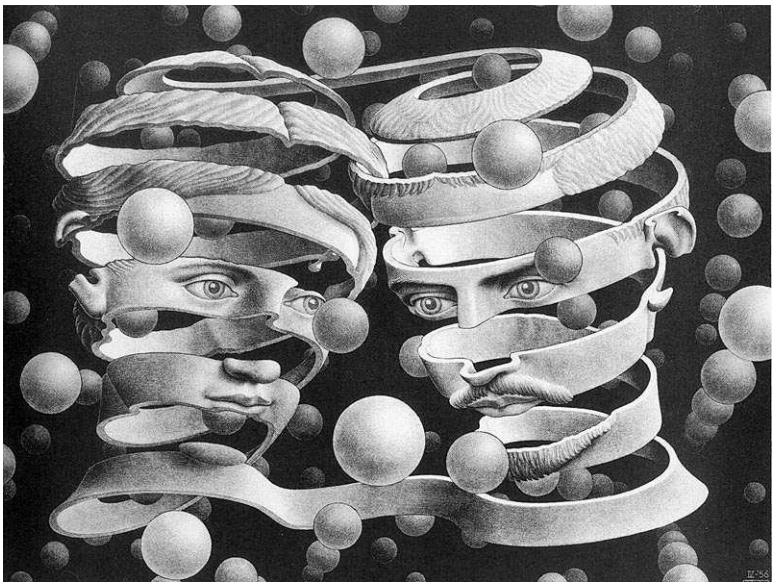
I want to thank the high-dimensional neurology and awareness lab members for their camaraderie and help. None of this work would have been possible without the assistance of support staff at the Institute of Cognitive Neuroscience (esp. Dean Lambert for setting up the IT infrastructure for chapter 4), the Institute of Neurology, and the collaboration of the MND clinicians of the National Hospital for Neurology and Neurosurgery (chapter 3).

Special gratitude must be given to the Engineering and Physical Sciences Research Council and the UCL Institute of Biomedical Engineering for funding my work and pushing for innovative and risky worthwhile research.

Thanks for all your encouragement!

"Anything in existence, having somehow come about, is continually interpreted anew, requisitioned anew, transformed and redirected to a new purpose by a power superior to it". (Nietzsche 1887)

Chapter 1: Foundations



M. C. Escher, Peeled Faces.

1.1 Introduction

The brain is a system so complex that after decades of dedicated research programs worldwide, the scientific community agrees that there is much yet to discover. To close the gap between research and application and extract benefit from this knowledge, it is therefore necessary to exploit findings prior to reaching full understanding of their underlying biological explanations.

The main goals of the research undertaken below is to add to the body of neuroscientific tools and knowledge and develop novel solutions for the diagnosis and therapy of patients with brain disorders. These efforts began by building the necessary conceptual foundations that guided the selection of research findings in neuroscience ripe for translation into practical applications. Ultimately, the face was selected as an appropriate concept as there is enough research that provides the necessary complexity, diversity, and fidelity for multiple data-driven hypothesis exploration.

The face is a key component in social interaction because of its informational value to emotional communication, perception, and understanding. It is a part of the body that is nearly always exposed and an accurate reflection of the intrinsic high dimensionality that reflect humans' complexity as biological social and emotional agents. Since it has been a research topic of interest throughout history, vast amounts of evidence from many independent methods are available to translate into beneficial health applications. Although other variables could be equally appropriate for this mission, selecting a single concept will narrow the scope of this thesis. In so doing I can ask a clear question: *What information from the face can be extracted to generate health outcomes?* Rather than concentrating on a function or process and changing the topic of study I can apply the translational approach to all aspects of human neuropsychology with respect to face research including perception, processing, and physiology.

A prerequisite to empirical study is an understanding of relevant notions of the key issues concerned. If we are mistaken in our theoretical foundation, any parameterisation of the data will also be wrong, and therefore any inferences may be rendered invalid. Since this PhD concerns itself with both elucidating biological questions and the construction of experiments attempting to influence positive change,

it is important to ensure ensuing manipulations are based on conceptually sound foundations.

Here, there are two distinct reasons to launch in this preamble. The first is that the face is intrinsically hard to reduce in its dimensionality, and therefore hard to study with the simple, low-dimensional methods most natural to neuroscience. The second is the opaque complexity of the psychological concepts surrounding emotion: they are difficult to see clearly precisely because we are so familiar with them. But being familiar with the use of a concept is not equivalent to having a synoptic view of its relations, which is what is needed to accurately characterise its function in human behaviour.

1.1.1 A new dimensional landscape

Biological entities vary between each other. Unlike when comparing two hydrogen atoms which are identical, one human being is different from another and this principle extends to all biological life. Furthermore, two biological entities vary along a large number of dimensions. If you were to take one mouse and compare it to another mouse, then there would be a great deal of variability between the two as they will differ across many visible and invisible features.

The notion of variability is premised on the scale at which one compares features and assumes the idea of parametrisation. To illustrate this, let's consider that the scale of parametrisation that is applied to a set of biological entities is on a species level, for example a mouse. If you consider the parameterisation of all individual mice at a spatial scale equivalent to the spatial dimensions of the particular species you are studying, you will identify all exemplars of the species as *mice*. The level of variability in the information captured in this example is at the species level so individual differences between mice will not be parametrised. Therefore, the scale of capturing the variability depends on parametrisation.

Parameterisation is the number and nature of the features chosen to describe an entity. These may vary with scale but need not necessarily do so. The choice of features is essentially arbitrary because biological entities have impossibly many features (if you count atoms, say). So, a level of parameterisation is established that seeks to satisfy two complementary criteria: first, to capture the features that matter for the issue of interest, and second to avoid capturing features that are irrelevant, simply noise that will obscure the exploratory task. For example, if one is interested in parameterising the sex

of a mouse all that is required is to look at the genitalia, or whatever feature best distinguishes male from female (e.g. chromosomes). But what should be measured if interested in parametrising the impulsiveness of the mouse? Evidently, capturing features of its brain and behaviour could provide an answer, but at what resolution? Since there is no way to know, first all that can be measured (e.g. brain function) is measured, and second anything deemed irrelevant is discarded. The first step is limited by the capacities and reach of current technology and is not under the researcher's control. The latter step however is a matter of mathematics and computation and can therefore be tuned to increase its accuracy.

Defining a reasonable level of parametrisation requires careful consideration as it is highly context-dependent. For the present purposes, reasonable is defined as the level necessary to enable a coherent explanation of the behaviour of interest in an animal. Whether this is at the tissue, cell, intracellular, or behavioural level, one will always find that (say) a mouse differs from another mouse on many dimensions. The scale will not be in the tens but likely in the thousands. This is true not only at the whole subject level but also for individual organs or tissue. If one takes the forebrain of one mouse and compares it to the forebrain of another, overall they might have the same shape and number of cortical layers, but the individual structure, the cytoarchitecture, will be radically different, just as in humans (Brett et al 2002). The transformation between the two cannot be explained in any simple or reductive way.

Scientific reductionism approaches complex problems by separately identifying, investigating, and summing their constituent parts. This idea reduces complex processes to more tractable subsets and facilitates scientific investigation. However, recognising that biological entities are composed of parts can lead to a misplaced tendency to look for a small set of explanatory features. Ultimately all physical entities and the universe must be explainable through one unified science of molecular biology and physics (Bechtel and Hamilton 2007). For example, a reductionist would consider that by researching underlying chemical reactions in the brain, they would be able to explain the complexity of the human brain and its functions including emotions, intelligence, and other human behaviours. The presumption is that the underlying mechanism is simple, and the observed complexity is essentially noise.

How complex are models of the brain expected to be? Biological organisms exhibit complex behaviours and physiology that cannot be fully explained if studied as

isolated parts. Can it be that accurate models of the brain are as simple as low-dimensional models assume? It is not tenable to expect reductionism to accurately characterise a complete entity as a reduced model cannot be as good as a *full* one (Bloom 2001; Strange 2005). Within Magnetic Resonance Imaging (MRI) research, the brain is crudely parametrised into chunks of brain or voxels that are about 2 mm across and contain over a million neurons each (Aguirre 2012). Even at this level comparing one brain to another requires a huge number of parameters because the non-linear transformation that describes one brain to another is in voxel space and is huge.

The implication that this holds for the variability in the biological domain is that they require many dimensions to specify but also to differentiate between each other, and between different states at different times. This is because the separation between different entities is not explained by any single variable or by any large or consistent agglomeration of variables. To make comparisons across individuals one needs to factor all the individual features and the more possibilities there are about where the boundary between those features falls the more replications are needed to ensure that is the case and the larger that set of features the harder it is to do this empirically.

If, for example, two samples are completely segregated by 1000 features, a single or small number of observations cannot explain differences accurately without knowing which set of features are critical. The traditional approach to what is known as the *curse of dimensionality* (Bellman 2015) presumes in biology that the underlying organisation is low dimensional. For example, say a researcher is interested in exploring the causes of hemispatial neglect, where following a stroke or damage to one hemisphere the patient stops attending to stimuli on one side of their body or visual field while retaining sensation (McFie et al 1950). The traditional approach would consider there is a network in the brain controlling attention and if this network is disturbed, hemispatial neglect ensues and this network can be adequately described with a relatively small number of parameters (Figure 1.1.A and 1.1.B). In neural network terms, instantiations of this problem are presented by computational models that can be composed of as little as 3 modules, yet are responsible for the whole of *attention* (Figure 1.1.C).

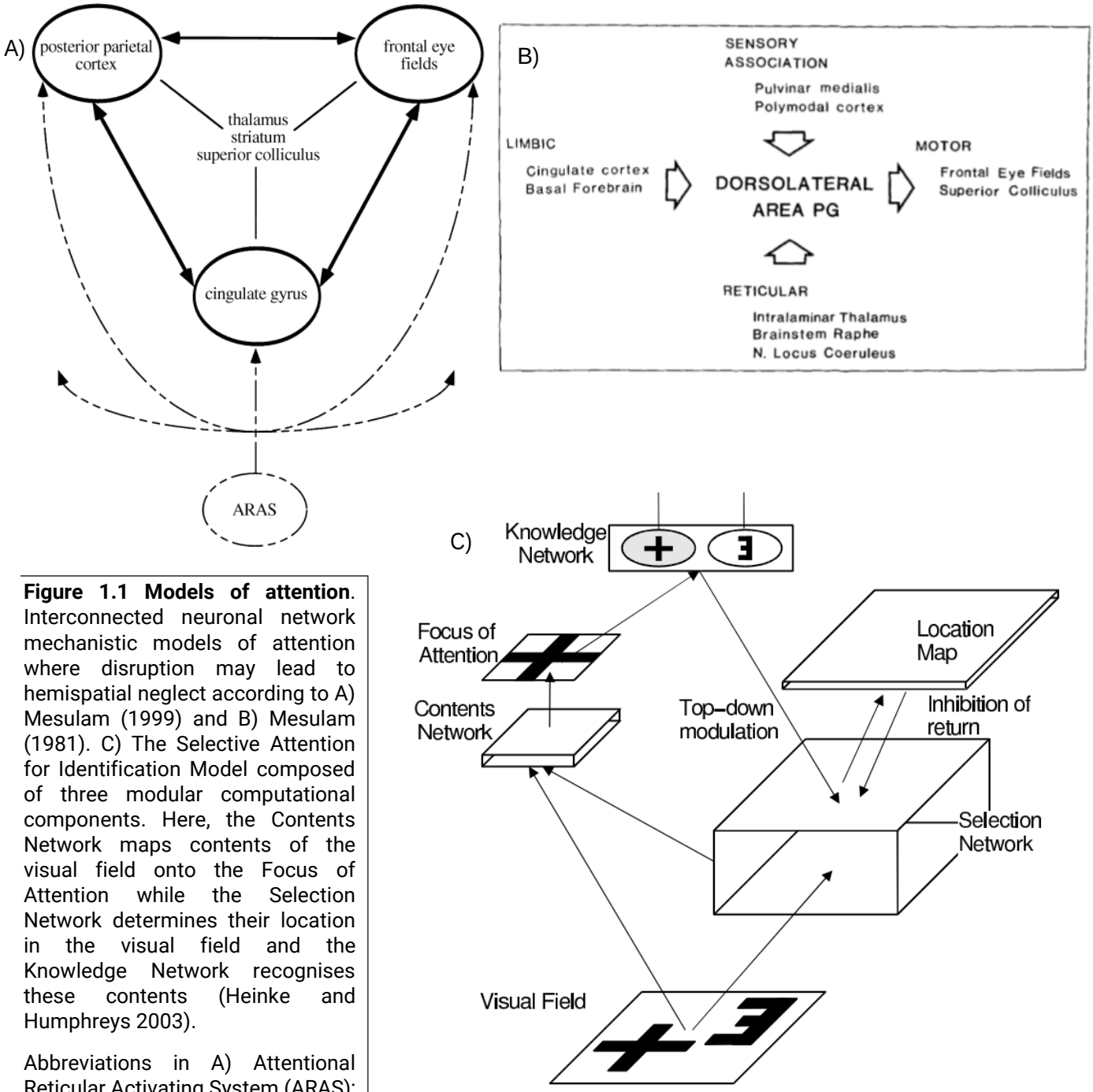


Figure 1.1 Models of attention. Interconnected neuronal network mechanistic models of attention where disruption may lead to hemispatial neglect according to A) Mesulam (1999) and B) Mesulam (1981). C) The Selective Attention for Identification Model composed of three modular computational components. Here, the Contents Network maps contents of the visual field onto the Focus of Attention while the Selection Network determines their location in the visual field and the Knowledge Network recognises these contents (Heinke and Humphreys 2003).

Abbreviations in A) Attentional Reticular Activating System (ARAS); B) Posterior Gyrus (PG)

There are two main problems with this reductive approach. First, it assumes that *high-dimensional variation across individuals is accidental* and what holds informational power are the low dimensional factors. Conceptually, it is akin to considering the biological entity as a watch mechanism, where the wheels might change slightly in shape or the springs might be in different positions, but the mechanism is essentially always the same. This is an assumption that considers the world as simple and this position discards all the high-dimensional features as noise and constitutes an assumption that may not be justified. If complexity is mislabelled as noise, individual variability and uniqueness may be lost and the cases could be thought to differ along a fewer set of dimensions than that reflected in reality.

The second problem with assuming that high-dimensional variation across individuals is accidental is that it can be empirically shown to be false. To go about testing the assumption there are two possibilities. The first is to see if that high-dimensional variability is indeed noise. If it is noise and one were to try to predict something about a system (e.g. will it live or die, is it a man or woman) as we add features to the model, the accuracy of prediction should not increase. So, after adding (say) 10 important features, thereafter adding more should not improve the power of prediction. In some domains, this was examined, and increased dimensionality adds predictive power provided one uses the right inferential approach.

For example, Chu et al (2012) found that data driven region of interest (ROI) selection of features was no better than using whole brain MRI data to predict which patients with mild cognitive impairment were to develop Alzheimer's Disease (AD), the most common type of neurodegenerative dementia associated with ageing. Furthermore, correctly predicting which MRI scans belonged to AD patients or healthy age-matched controls was more effective using data form the whole brain rather than a literature based ROI (Cuingnet et al 2011). Finally, in an explicit challenge to the modular notion of brain functional organisation for higher cognitive processes, Nawa and Ando (2014) found that whole brain functional MRI scans were superior in classifying activity of subjects recalling positive or negative autobiographical memories than ROI specific activity. Consequently, this shifts the notion that a particular dataset may contain hardly any *non*-informative features, to only *less*-informative ones. Thus, although the reductive approach was able to make some progress and general observations, high dimensionality captures more of the informative variability.

Equally, the more cases are added, the better the heterogeneity of feature conjunctions will be captured and predictive power increases. To illustrate this point, consider the problem of face recognition and classification. When one applies machine learning techniques (explored in more detail in section 1.1.3) the predictive power is higher as increasing amounts of data are available. This is best exemplified in the battles between the big internet companies to optimise face recognition. Whereas Facebook trained their algorithms on 4 million photos and achieved 97.25% accuracy (Taigman et al 2014), Google used 200 million to achieve 99.96% near perfect classification (Simonyan and Zisserman 2014). In other words, when one adds millions of faces to a dataset, the performance becomes better and so on until maximum accuracy is achieved. Of course, performance will always depend on different algorithms and some added features will sometimes inevitably be noise. In general however, whenever a biological system is heavily parameterised predictive power increases well beyond the level of parameterisation of most mechanistic models which do not go beyond a handful of factors.

The second way to test the assumption is to explore the possibility that the same function may be subserved by different mechanisms. At its most basic a biological mechanism is an entity with a range of powers allowing it to act upon the world and solve problems. In this context different mechanisms could result in the same solution. There is no need for a brain to add 2 and 2 the same as another does, the mechanism to produce the solution '4' can be different. If sets of mechanisms can be different, then looking for a *common* solution must be wrong because the process of getting there is different. The traditional approach to discovering the multiplicity of possible solutions would involve averaging and finding what is common across all mechanisms. However, this will ignore the heterogeneity in the population as it assumes that the mechanisms are all comparable and therefore the same, thus raising the risk of false negatives (not finding a mechanism where one exists).

Multiple solutions in biology are however expected in some domains. This could be because the final biological form is not definitively specified in the genes. If there is not enough information in the gene, then the final product needs to arise from elsewhere and it is here proposed to arise through learning. The term *learning* is about any input-output transformation at any level. This could be, for example, how an enzyme performs a certain task and the way it is instantiated at a particular level through feedback. So, an enzyme will typically be under feedback control from another agent at the molecular level

where some would suppress it and another will increase it and the system finds an optimum. When a multi-parameter system is finding an optimum, there is no reason why the system should yield the same solution and all might be equally good in multiple individuals.

A classic empirical example of this process is found in that of the gastric mill motor patterns of the lobster's gut (Marder and Bucher 2007). Though simple in anatomy – about 30 motor neurons – and in behaviour – open and close the muscles to mechanically break food – its extensive study has shown that different mechanisms may generate similar gastric mill motor patterns and different neuronal connectivity may subserve the production of similar motor patterns. It also shows evidence of learning as though it may be fully formed early in development, as an adult it will generate different motor patterns. If such complexity can be true of the gut, is it not plausible that it is also true of the brain?

There is an emerging trend in the life sciences to embrace complexity and diversity in its many forms leading to a higher dimensionality within the field. An increasingly established principle finds that biological phenomena can in most cases be better explained by taking a systems approach (Kesić 2015). This reflects a movement towards considering functionally critical information is not concentrated in single genes or specific brain areas, but it is rather distributed across loci in networks, an emergent product of their interactions. For example, mirror neurons were a class of cells originally discovered in a macaque's ventral premotor cortex and inferior parietal lobule that were thought to become active when the macaque performed an action and when the same action was executed by another visible macaque or person (Rizzolatti et al 1996). Over two decades of human brain imaging studies later, mirror neurons are now considered part of a system where a network of over 20 anatomically interconnected cortical regions are coordinated in matching actions and observations (Bonini 2016). Although a lot more was learned about the systemic nature of organisms, it is nevertheless unreasonable to expect the information contained in a small number of loci to fully account for the complexity of the mechanisms studied.

The intrinsically complex and multi-faceted nature of the biological will require amassing large datasets with descriptors of their interactions that are composed of multiple attributes. This is a problem that will necessarily require a high-dimensional analysis to answer. Across many biological disciplines this anti-reductive approach is

accepted and holist and reductionist methods are balanced to yield the most accurate representation of phenomena (Romero et al 2006; Anon 2014).

1.1.2 The quality of data

Appreciating complexity is accompanied by a recognised need for a different type of data that enables useful high-dimensional analyses. This should not be surprising: without fully exploring the space that defines a complex system we cannot expect to understand it. One of the main consequences of adopting a non-reductionist approach in the life sciences, is the incredible amount of data generated.

Ushered by the development of more affordable and capable technology, copious and substantial datasets are now routinely produced across many life science fields. The most common experimental solution to dealing with so called *big data* is to explore the average of the behaviour in question. Although this may be enough to superficially understand a biological system by powering small effects through better definition of the mean, it cannot fully characterise the disease processes or physiological responses which can be obscured by random variance that occurs in a biological system. Accounting for the presence of inherent stochastic components that govern interactions between processes within biological systems are now being recognised as a major step towards greater clarity (Quackenbush 2007). These large number of different cases allow the capturing of consistencies in variation in the high-dimensional space. By gathering enough data to identify the heterogeneities in its complex structure, the clusters best illuminating each individual are revealed through a high-dimensional approach.

In the field of neuroimaging for example, a study of patients where sufficient data exists to capture the complexity of the lesion distributions using multivariate inference based on machine learning has proven quite disruptive. With the largest sample of structural imaging post-stroke lesion data published to date (N=581), Mah et al (2014) quantified and modelled the mislocalizations found in conventional mass-univariate lesion mapping distributions. By showing that co-occurrence of damage across voxels leads to systematic biases in this approach to lesion mapping, the authors demonstrated the superiority of combining high-dimensional data and analysis. In other words, making comparisons across *whole* brains rather than individual voxels captured the pattern of damage in the underlying anatomy contributing to the functional deficit more accurately.

Where a reductionist approach would have scientists discard much of the features as noise, the high-dimensional model interprets what was redundancy as informational and incorporates it towards fuller explanations. The better data is available the more likely it is to accurately describe the complexity of the processes and avoid mistaking meaning for noise. This effort should not be confused with likelihood of falling into type-2 statistical errors, rather it is about discovering the underlying dynamics between the variables at play. Indeed, a higher dimensionality in data structures has not solely yielded datasets composed of a larger *N*, this drive has truly pushed towards greater quality as evidenced by the use of this data to generate personalised treatment models.

Where conventional clinical trials track single measures among thousands of people, personalised approaches require probing many factors - including genetics, behaviour, and environment – to develop precise interventions customised to a particular person (Schork 2015). These single subject or *N-of-1* trials are multi-cycle within-patient, randomized, double-blind studies that use copious information on individual response to treatments and have proven useful for research studies (Guyatt et al 1990; Lillie et al 2011; Woodworth et al 2016) and to optimize the management of chronic diseases (Scuffham et al 2010; Mitchell et al 2015; Punja et al 2016). These trials consider differences between individuals as informational, and best courses of treatments are developed based on data and evidence rather than dismissing variance as noise. Although on the surface it may seem oppositional to the idea of big data, it is simply a shifting of the space where the high-dimensionality model is applied.

Consumers today have greater facility to gather data about their health from wearables and mobile technology, holding the potential to truly shift healthcare models towards patient-driven targeted preventions (Swan 2009). Embracing the advent of such technologies to complement the ongoing self-monitoring of many categories and variables (see Table 1), led to the generation of massive datasets by people in both good and ill health (Swan 2013). This high-dimensionality at the individual level, composed of multiple diverse measures, has led to a more comprehensive view of diseases and is challenging the *one-size-fits-all* approach to solving some of the greatest global health challenges (Adams 2016). The implications for concentrating on single people as *worthy* of particular study rather than in aggregate renews the dignity of the individual and the conception of wellness and healthcare in society. It accepts that people reflect the

polymorphic variations found across individual physiology and cannot simply be represented as accidental departures from the canonical form.

Table 1. Categories and Variables of Self Quantification.

Physical activities: miles, steps, calories, repetitions, sets, METs (metabolic equivalents)
Diet: calories consumed, carbs, fat, protein, specific ingredients, glycemic index, satiety, portions, supplement doses, tastiness, cost, location
Psychological states and traits: mood, happiness, irritation, emotions, anxiety, self-esteem, depression, confidence
Mental and cognitive states and traits: IQ, alertness, focus, selective/sustained/divided attention, reaction, memory, verbal fluency, patience, creativity, reasoning, psychomotor vigilance
Environmental variables: location, architecture, weather, noise, pollution, clutter, light, season
Situational variables: context, situation, gratification of situation, time of day, day of week
Social variables: influence, trust, charisma, karma, current role/status in the group or social network

Source: http://measuredme.com/2012/10/building-that-perfect-quantified-self-app-notes-to-developers-and-qs-community-html/

1.1.3 Following the data

With more available higher quality datasets key research findings become more reliable, disturbing the structure of research and allowing scientists to bridge the gap from discovery to implementation. Historically, science involves investigating and recording physical phenomena and conducting trial-and-error experiments driven by hypotheses. One shouldn't however presume the existence of how the mechanism works. That is what the traditional reductive approach would do: there is a hypothesis and it should be tested. However, this introduces a form of bias as it enforces hypotheses that are plausible within the current limited understanding of a field, safely embracing the familiar and accepted. The task in science is not hypothesis testing but hypothesis *comparison* and if there are no means to define the space of available hypothesis, then this is not a possible task.

Pushing the boundaries of scientific enquiry towards embracing this complexity leads to *data-driven* approaches that require high-dimensional solutions to make these tractable. Machine learning methods are algorithms that learn from experience and discern information within the variability of the data from which scientists can discover the spectrum of possible patterns within the data. In this way, the process of hypothesis generation becomes driven by the data and one can undergo a process of hypothesis selection. Although this should mean one is less likely to be biased by one's own imagination, it is not because the data fits a particular hypothesis that another hypothesis could not fit the data just as well. An independent measure of the quality of how good a hypothesis is, can be measured in how useful it is in predicting outcomes with machine learning methods.

There are many artificial intelligence modelling methods falling under the umbrella term of *machine learning* and reviewing these is beyond the present scope as these are best treated by specialised textbooks on the subject (Alpaydin 2014). However, all problems that will need to be solved by machine learning techniques in this thesis share a common goal: to predict whether a data point belongs to a particular group or *class* with the highest accuracy and speed. Since the actual true class or label of the data point will also be known at the time of testing, all relevant machine learning problems are presently to be considered as *supervised learning classification* problems. This therefore limits the choices of potential relevant techniques to a specific subset by way of excluding certain techniques. For example, unsupervised learning algorithms aim to cluster data points into groups sharing highest similarities or reduce dimensionalities to limit the number of variables under consideration. As these techniques are most useful when the labels are unknown or when high dimensional variables are considered to be irrelevant, these are not the most appropriate to solve the tasks that will be laid out in subsequent chapters.

Furthermore, within the subset of supervised learning, further requirements dictated a preference for speed and accuracy and also, importantly, ease of use. This led to steering away from the more complex and resource-heavy artificial neural network algorithms which became best at solving unsupervised "deep" problems such as feature extraction (Schmidhuber 2015). Similarly, random forest algorithms randomly combine data features to create "trees" and compare subset of trees (i.e. forests) to generate predictions of actual class belongingness (Breiman 2001). This method was not opted for considering random forests are generally best suited for multi-class problems and

the solutions in this thesis require simple binary two-class prediction. Finally, one of the simplest and seemingly most commonly used machine learning methods[1], the support vector machine (SVM) was selected as ticking the most requirements for present purposes.

One of the greatest advantages of SVMs is the relatively "out-of-the-box" ability to be implemented without requiring delicate and computationally expensive tuning. This allows the method to be implemented successfully by non-computer scientists such as the author and be applied to other fields outside artificial intelligence. In addition to being simple to grasp theoretically and implement within a coding environment, SVMs have been established as excellent binary classifiers when benchmarking them with other methods where labels are known (Meyer et al 2003) and produced such results across a range of disciplines (Statnikov et al 2008; Nitze et al 2012), making them an adequate choice to implement in this thesis.

Specifically, the form of the technique used in this thesis is a binary SVM which, at its most basic, is a learning algorithm that classifies data points according to their particular group (Cortes and Vapnik 1995). The decision to classify is based on finding a separating line or *hyperplane* that can distinguish between classes with the largest margin between instances of both groups closest to the borders (aka the "support vectors"). To achieve this, it is necessary to first obtain a dataset which is large enough to separate the instances into training and testing sets. Each instance in the training set should contain a *label* that classifies it within a particular group and be described by high-dimensional descriptor variables consisting of *features*. This process of modelling the matching of features to labels is called *training* and is how SVMs "learn". In cases where a clear line is not identifiable between the input data, SVMs will use a set of mathematical functions called *kernels* to re-arrange the objects in high-dimensional space based on the degree of similarity between the features describing the data. Rather than constructing a complex curve between the data points kernels can map the data so that the optimal linear hyperplane can then be used to separate the categories. The accuracy of the SVM model produced in training can be tested by predicting the label of new points of features that are found in the testing set. Furthermore, the labels of the

[1] A rough indication of method adoption can be gained by searching for specific terms in google scholar and observing the number of published scientific works for a particular time period. A search from 2013 onwards for "support vector machine" yielded 69,400 results while "random forest" yielded 20,300 results (Dec 2017).

testing set are used to compare the predicted labelling (ensuing from the SVM) versus the expected outcome (the actual labels) to generate a measure of generalisability of the SVM model of accurately classifying unknown data.

SVMs operate in contrast to hypothesis-driven methods such as logistic regression which fit the input data to a curve defined by a pre-determined model to predict the occurrence of a binary event. Since the SVM generates a model based on the data, it is a powerful agnostic classification tool that provides an unbiased alternative to elucidate patterns in data where samples sizes are small and a large number of high-dimensional variables are involved (Yu et al 2010). These machine learning models bestow the freedom to enquire along emergences outside of the constraints of conventional hypothesis-formation and draw useful conclusions directly from the data itself (Anderson 2008). With the greater availability of complex electronic health records and surge in computational power, SVM classifiers have gained momentum in health research during the past years.

For example, reframing medical challenges such as seizure detections and onsets in epilepsy from electroencephalography (EEG) recordings as a classification problem can be very useful to notify caregivers and better manage these episodes (Shoeb and Guttag 2010; Fergus et al 2016). Achieving robust seizure detection, avoiding false positives, and generating a model with good inter- and intra-patient generalisability is difficult considering the high variability in seizure morphology. By treating each individual seizure as a separate classification task, Van Esbroeck et al (2016) generated a multi-task model that captured seizure-specific characteristics and patterns shared across all types of seizures within the same patient. As the results generated a single shared hyperplane, this was used to separate specific seizures while optimizing performance across the high-dimensional shared seizure types, thereby reducing false positives significantly as the more common seizure types are not over-represented in the final hyperplane.

In an exemplary demonstration of the benefits of collaborative science, Jongkreangkrai et al (2016) utilised an SVM approach with open-source software and open-access repositories of health data, to inform the description of conditions such as AD. The authors inputted cortical thickness measurements obtained from MRI scans of several brain areas implicated in AD (including the hippocampus, amygdala, and entorhinal cortex) into an SVM to explore which of these were most useful to use as

markers of degeneration to predict presence (or absence) of the disease. The classifier was able to discriminate from as little as 100 healthy controls and 100 AD patients whether the scans belonged to a normal individual or an AD patient. Although the study suffered from low sample sizes, it holds potential to explore the patterns of degeneration that may characterise AD and create an algorithm that can predict whether an individual may develop the condition.

1.1.4 Re-structuring scientific research

By recognising that accurately parametrising complexity demands large datasets, one recognises the priority of considering how large datasets can be assembled. It is always hard to do this purely for research purposes, which is why a dual-track parallel system is required to apply what is known, and the utility of the result can be used to justify (and enable) the collection of large scale data. By following the data where enabled by innovative technological developments, scientists re-structure the practice of research and blur the gap between the basic and applied domains.

Most scientific endeavour can be dichotomised into fundamental, purely curiosity driven research (sometimes referred to as *basic*) and applied research, where the effort is aimed at solving a real-world problem. Within the corporate and government sector, these are unified under an innovation cycle termed Research and Development (R&D). Although in these sectors there is usually an emphasis on commercialisation which is considered dangerous by many academics due to risks of interference by the corporate funders (Olivieri 2003), it can be a useful motivator to explore how research may be put into practice for the benefit of the public. This is where an alternative conceptualisation of doing science, a *parallel translational approach* to research has provided a more seamless integration of the classically separate paradigms of basic and applied research (Ashburner and Klöppel 2011).

Although some consider basic research as “not-yet-applied” (Coward and Franklin 1989) it is implicit within such frameworks that it be *others* who should apply it. However even when translating clinical research is considered a national priority and responsibility (Zerhouni 2003), it is a difficult endeavour to achieve as there are more difficulties to this kind of approach than in traditional scientific research (Zerhouni 2005). If there is no intent within the field to do so however, it may not necessarily happen and if the experts who develop the innovation do not make it easily communicable and

accessible beyond their specialty, it will hardly happen at all. Where there is no natural connection between the field of study and practical applications of the research, as there is in (say) engineering or pharmacology, establishing a parallel translational approach or culture within an academic field requires an innovative shift in attitude that embraces uncertainty. The parallel element emphasises the dual-track system whereby large datasets may be enabled and justified through exploring the contained utility of the results which in turn allow for greater data-driven hypothesis explorations.

The researchers need not shy away from the complexity and time required in becoming familiar with the apparatus outside of conventional research methods and accepting a high risk of failure for venturing out of the established paradigm within their field. It is also a risk that parallel translational research end in a limbo between a well-established structure of a university department that may further confuse the efforts. In the current case, the uncertainty of not sitting precisely within UCL's Institute of Neurology, Institute of Biomedical Engineering, or the recently established Institute of Digital Health informed and directed the current projects by providing a flexible network of knowledge and opportunities. Ultimately, the crux of the parallel translational approach hereby adopted is to progress from efficacy into effectiveness by exploring how findings can translate into real-world health solutions and collect larger datasets that further research.

As a pure curiosity-driven exercise the parallel translational approach provides a powerful test to the quality and confidence of our knowledge about the phenomena being studied. To understand a system, it is argued that one needs to demonstrate "fixing" a broken instance (Lazebnik 2002; Jonas and Kording 2016). The parallel translational approach puts the knowledge base to the test by applying the accepted concepts to generate change in the world. Should such a soundly reasoned and developed intervention fail, it could lead to a re-assessment of the extent to which the knowledge base truly relies on the mechanisms and processes it purports to define or its completeness. As with any innovative approach that aims to impact people's lives, an unwanted consequence of failure could be the potential for negative outcomes. Of course, great care should be taken during the conceptualisation stages to ensure that these are minimised, and a high ethical standard is ensured throughout.

The parallel translational approach can demonstrate mastery of phenomena by generating clinical impact from the application of findings within a field of specialty. As

demonstrated in the examples discussed, the implications of data-driven parallel translational science are the predictive power that is generated by the results of such research. This model is also being more recently recognised as a way forward to create sustainable global health interventions that inform policy (Neufeld et al 2016). One of the aims of this thesis is to generate such impact from neuroscience by employing the high-dimensional dataforms and methods presented above.

1.1.5 The face

To demonstrate clinical and scientific impact, the current approach individuates a high-dimensional unifying concept that will serve as the central theme from which to apply such methods and upon which enough theory and knowledge is available to derive concepts of interest and exert influence: the *human face*.

The human face is a high-dimensional feature that holds potential to translate what is known about it into applications with health outcomes. This is not to mean that there are no other variables which could be equally appropriate for this mission, however selecting a single concept that is well-established will narrow the scope of this thesis. In so doing we can ask a clear question: *What information from the face can be extracted to generate health outcomes?* Doing so also allows the researcher to apply the translational approach from all conceivable aspects of human neuropsychology with respect to face perception, processing, and physiology rather than concentrating on a function or process and changing the topic of study.

The face is undeniably key to human social interaction because it holds special informational value to emotional communication, perception, and understanding. Owing to the central role of faces to social, cognitive, and emotional communication in humans, extensive research on the face has revealed how information such as identity, age, intention, and gender are conveyed within them (see Calder (2011) for a review). As it is involved in a wide variety of important human phenomena, there is an equally large scope to generate meaningful impact from translating research within the field. Indeed, the face is a readily accessible part of the body that is always exposed and may thus be accessed non-invasively. As it has been a topic of study for centuries, there are large amounts of evidence acquired from independent methods about the way faces are processed and identified (Farah et al 1998) and how the muscles (Moody et al 2007) and brain (Lee et al 2006) have selectively specialized to enable these processes.

Due to the highly complex and varied information the face provides on specific emotions and behaviour, the face can be considered high-dimensional yet specific. When dysregulated, this can hold severe implications for successful social interaction and integration within society as demonstrated by associated dysfunctions in patient populations. For example, autistic individuals demonstrate atypical visual and neural mechanisms to determine facial identity (Mukerji et al 2013) including poorer memory for faces (Gelder et al 1991) and marked difficulty at identifying negative emotions (Enticott et al 2014), which may contribute to the social deficits presented in autism spectrum disorders. In other patient populations such as psychopathic individuals, it is possible that their lack of ability of recognising the emotional information transmitted in facial expressions may disturb the development of healthy interpersonal relationships and social adaptation (Contreras-Rodriguez et al 2014). Deficits in facial responding can also occur automatically as in borderline personality patients who demonstrate abnormally increased non-voluntary facial responses for negative emotions (Matzke et al 2013).

This contributes to the face as an accurate reflection of the intrinsic high-dimensionality of biological social and emotional agents. The field could therefore contain enough theory and data to derive concepts of interest that can be used to exert influence and generate impact. As a field of study, the face has provided copious amount of multivariate, independently acquired, evidence of its systems and functions thereby requiring the reader familiar themselves with this variable and associated inseparable concepts as these are the tools ripe for translation.

1.1.5.1 Conceptualising emotions

> “We see emotion." As opposed to what? We do not see facial contortions and make inferences from them (like a doctor framing a diagnosis) to joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give another description of the features.” (Wittgenstein et al 1967)

It would be difficult if not impossible to better emphasise the intricate role of emotions in characterising a face than the quote above. Since faces provide the single most informative conduit through which humans perceive emotional states it is important to re-evaluate the conceptual aspects surrounding these.

Detailed treatises on the conceptual structure of emotions have placed them along with agitations and moods, as subtypes of affections, which further fall under the umbrella term: feelings (Hacker 2009). Emotions are active and intertwined modes of engagement with the world that reflect a people's beliefs and interpretations (Marks 1982; Oakley 1992; Hacker 2015). These have a duration and course, stem from a specific object and cause, and can be associated with distinct facial expressions and somatic sensations. Hacker (2009) further differentiates between internal dispositions or emotional attitudes, such as pride for a childhood achievement or love for one's partner which may last a lifetime, and short emotional perturbations or outbursts. Distinguishing emotions from moods and agitations is also important. Moods can be described as a proneness to feel a particular state of mind during waking hours though lacking specific object while agitations are short term affective disturbances caused by unanticipated events. The boundaries between these are often blurred as (say) a dog barks and begins chasing a person whose initial reaction may be to *cry in startlement* (agitation), which crystallises into a specific *fear of the dog* (emotion) and once as the dog is muzzled and led away, eventually dissipates into *objectless anxiety* (mood).

Evidently, emotions cannot be considered as disjointed from a person's wishes, imagination, deliberations, and beliefs about the world (Scherer 2005). Emotions are dependent on multidimensional behaviours, beliefs, specific contexts, and prior events that compose the complexity of human lives (Bedford 1956; Oakley 1992). For this reason, emotions are flexible, as they need to accurately account for people's actions and reactions to our diverse and unpredictable world. This flexibility translates into emotions not being reported equally by observers and those experiencing it even when it is manifest in the face (Prkachin et al 1994). This reflects a constitutional indeterminacy which is central to our concept of emotions (Wittgenstein 1958) and presents challenges to establishing a uniform and rigorous criteria to ascribe emotions to individuals.

Although this line of reasoning logically dismisses single simplistic criteria such as self-report or non-verbal behaviour measures as incapable of capturing the full range of emotional experiences (Cannon 1927) these have been and remain the most popular method of enquiry into this subject. One's emotional state in psychology is traditionally communicated through verbal self-report: asking the patient to introspect about how they feel, and their justifications for feelings. Verbal report is useful because it provides *specificity* of language to differentiate between different emotional states, however it

ignores non-verbal cues that are also informative. Therefore, it is only partial- though important - for report is merely one manifestation of emotion. To address this shortcoming, many turned to taking physiological measures such as pulse, heart rate, or brain activation as markers for emotions. Although this approach provides a more objective measure of the emotional response, it *lacks specificity*, as it is incapable of differentiating in a univariate way between the valences of emotions (Schachter and Singer 1962). For example, quantitative differences in heart rate or pupil dilation cannot dichotomise between being scared or sexually aroused. Thus, neither can form logical criteria to ascertain that someone is feeling an emotion.

While there is a connection between emotions and their behavioural, physical, and verbal manifestations, it is unhelpfully reductive to contextualise emotions as only one of these. But how can we combine the sensitivity of the physiological with the specificity of the cognitive methods in one high-dimensional variable? The answer is: the face. The muscles of the face are capable of minute alterations that are immediate, direct, and specifically configured to every emotion which humans are naturally sensitive. Unlike the other involuntary physical manifestations of emotion, the face has specificity, for the expression of an emotion is tightly constrained by its valence (Ekman et al 1972). Studying the face thus combines much of the specificity of verbal reporting with the directness of physiological measures.

1.1.5.2 Face processing and neural architecture

The search for understanding how neural systems are organised to achieve face perception has stimulated intense debate. The story of this debate demonstrates parallels to the higher dimensionality achieved by models of mirror neurons describing action observation and matching (discussed in 1.1.1).

Originally the neural underpinnings of face processing were thought to rely mainly on the fusiform face gyrus (FFA) as it preferentially activates during face perception as opposed to in response to other stimuli (Kanwisher et al 1997). A more recent review of the neuroimaging data however proposes that face perception relies on a distributed and highly interconnected system that can be modelled as a Core and an Extended System (Haxby and Gobbini 2011).

Both of these systems are composed of anatomically distinct regions, each responsible for the distributed and complementary processing of the various aspects involved in interacting with faces. The core system is posited as responsible for analysing the visual information from faces and recognising facial identity. Within the system, the FFA and the occipital face area are considered more responsive to constant face features while the posterior superior temporal sulcus is more involved in analysing dynamic social information of faces such as eye gaze and expressions (Engell et al 2006). The extended system on the other hand is more involved in recognising specific information from faces including familiarity, facial expressions, speech, and includes subcortical areas for emotional association of the faces with episodic information. This system is generally accepted as the most accurate representation of our current knowledge related to face cognition and is well summarised in Figure 1.2.

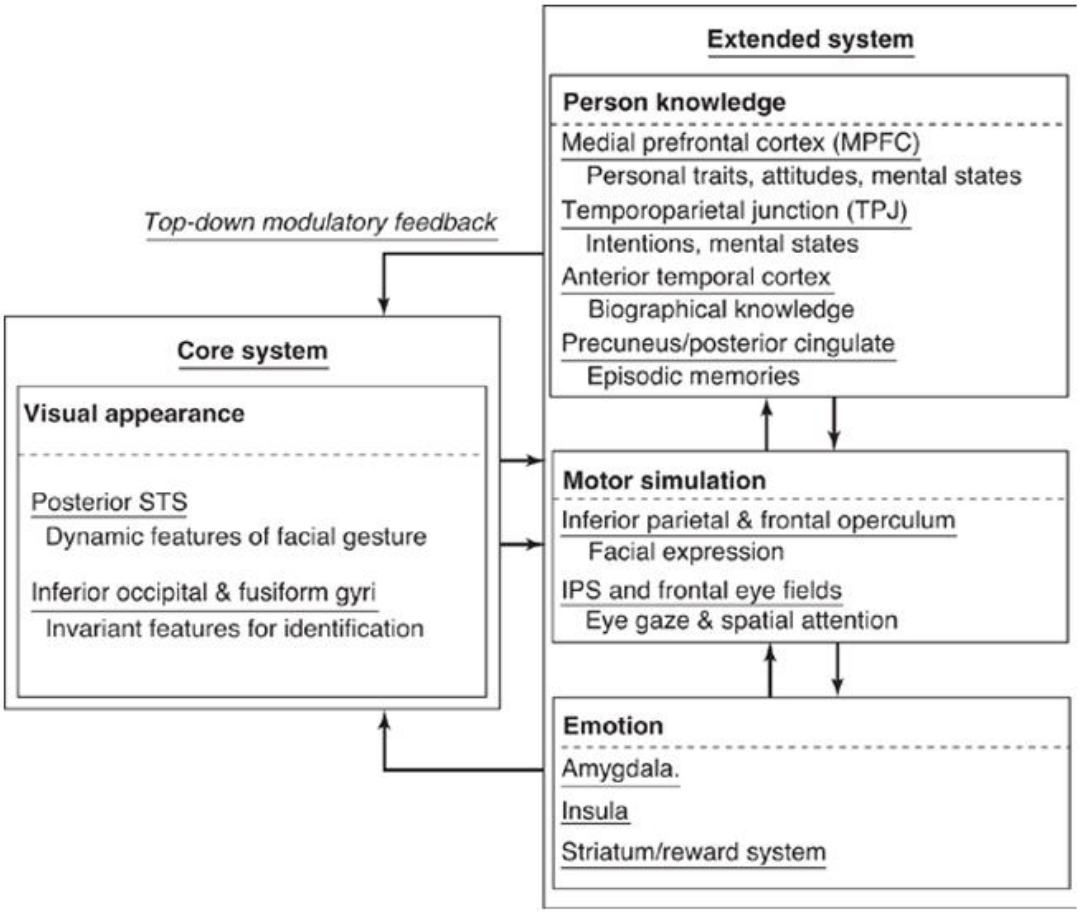


Figure 1.2. A neural model of face processing (from Engell et al (2006))

It is no surprise that the high-dimensional approach to solving the problem of face processing and identification relies on algorithms inspired by neuronal function and SVMs that perform almost as well as humans (Taigman et al 2014). Indeed, Nachev (2015) suggests that when assessing the individuality of a face, a low-precision high-dimensional parametrisation (e.g. compressed and pixelated low resolution whole face image) will be more superior than a high-precision, low dimensional parametrisation (such as interpupillary distance). Importantly, the average of other faces in a similar category will, though providing a more immediately recognisable face, be equally of no value in assessing identity (see Figure 1.3).

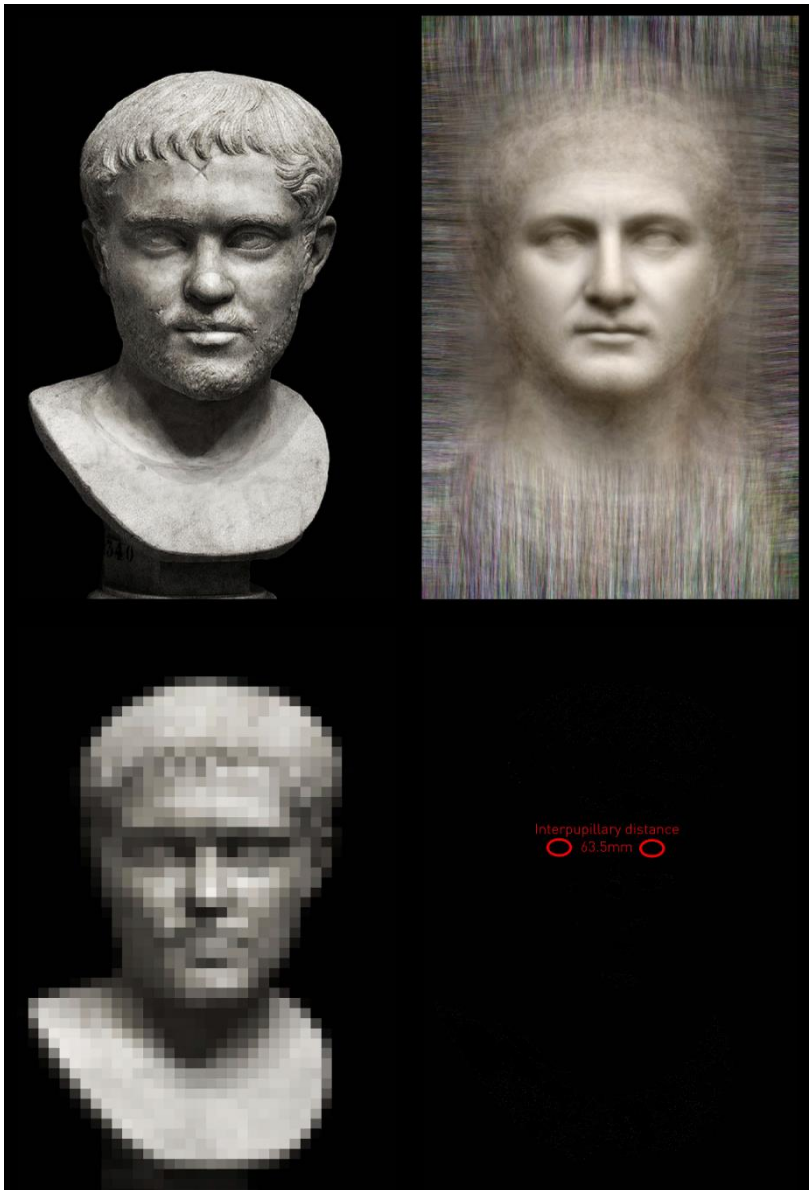


Figure 1.3. The superiority of higher dimensions in face processing. The bust of the Roman Emperor Hostilian (top left) and the average of all Roman Emperor busts (top right). A low precision high-dimensional parametrisation (bottom left) and a high precision low-dimensional one (bottom right). From Nachev (2015).

1.1.5.3 From faces to emotions or emotions to faces

Charles Darwin (1965) was a pioneer of research on the effects of the face on emotion activation and regulation by proposing that voluntary outward mimicking of an emotion may intensify it while repression of a felt emotion may weaken its feeling. Shortly after, William James (1884, 1894) and Carl Lange (1885) independently affirmed that it is our natural physiological reaction to situations which bring about the experience of a specific emotion. Openly admitting to contradicting common intuition, which stated that emotion gives rise to a congruent physical reaction, they took the inverse position that emotion is derived from perceiving one's own physiological state. They proposed a feedback-loop mechanism that began with the initial perception of an emotional situation or object with the sensory organs followed by the processing of the content and associated visceral sensations in the brain, finally transforming the situation from "simply-apprehended" to "emotionally-felt" (James 1884).

The most famous critique of the James-Lange theory considered that the autonomic nervous system would be too slow to account for the speed and lability of arising emotions and that visceral sensation lacks valence specificity to account for the range of human emotion (Cannon 1927). Subsequent self-attribution theories of emotions shifted the sensations from the viscera and re-emphasised the importance of the face while adding a cognitive dimension through requiring interpretation of others' and one's own facial expressions (Schachter and Singer 1962; Laird 1974). Making visceral changes unnecessary circumvented Cannon's criticisms and lead Tomkins (1962) to propose that the uniquely fast facial muscles could communicate one's emotional state to others and even the self without conscious awareness through a feedback mechanism which he coined: the facial feedback hypothesis (FFH).

All versions of the FFH state that facial expressions influence the experience of emotion in a positive feedback loop, however there are only two versions that survived empirical testing (Schneider 2008). The "weak" version considers that already felt emotions can be intensified by proprioceptive feedback stemming from the muscles involved in creating the facial expression, while the "strong" version considers facial action can, on its own, lead to the feeling of an

emotion where there none existed (McIntosh 1996; Rutledge and Hupka, 1985). Although there has been extensive research performed and evidence found in support of the weak version, the stronger version has not received equal attention though it has been identified as "a promising avenue for future research" (Schneider 2008) and will therefore be prioritised in this thesis.

Although the FFH took different forms, *facial contagion*, or the tendency of humans to spontaneously adopt the facial expressions of others, remained central to understanding the relation between a facial expression and a feeling. Nevertheless, it is perfectly possible for all of these processes and competing theories run in parallel and interact at multiple levels. Neither need be secondary to the other.

1.1.5.4 Contagion

Facial contagion has since been considered as part of a wide array of pervasive behaviours defining social interaction whereby people adopt features common to other organisms that they are observing or otherwise sensorially aware under the umbrella terms of *contagion* or *mimicry*. Although the term 'mimicry' is prevalently used in the literature to describe this process (Hatfield et al. 2014b), the term may presuppose that someone is engaging in voluntary action (as in the case of a mime or a parrot repeating a sound that we recognise as a word). Thus 'contagion' more accurately denotes the passivity involved in this process.

Contagion is found in many behavioural forms as people spontaneously copy gestures, postures, mannerisms, and facial expressions unintentionally (Chartrand and Bargh 1999). Empirical research established the universality of this phenomenon found equally in adult humans (Dimberg et al 2000; Hess and Blairy 2001) as in children (Beall et al 2008; Jones 2009; Deschamps et al 2012). These findings were also investigated across species with complex social interactions and found in Orang-utans (Ross et al 2008), Chimpanzees (Davila-Ross et al 2011), and Gelada baboons (Mancini et al 2013a, 2013b) in different forms. An example of this is contagious yawning, whereby seeing someone yawn can trigger a yawn in the observer (Platek et al 2005) and was found in several animal species (Paukner and Anderson 2006; Palagi et al 2009) and can even occur across-species (Buttner and Strasser 2014).

Although contagion appears as a more common feature of the emotional (e.g. laughing when another is laughing) than of the cognitive, it is certainly not unique to it. There is evidence of contagion at a *physiological* level, with electrophysiological and neuroimaging evidence in humans and monkeys demonstrating that perceiving others' actions activates congruent cortical motor representations in the observer (see Rizzolatti et al (2002) for a review). This type of "neural" contagion is seen in the domains of action, and it may well be that emotional contagion is a species of this phenomenon. Why neural contagion occurs is for the moment mysterious. It may be part of the brain's anticipatory mechanisms, preparing the agent for the circumstance that has presumably caused the behaviour in the other animal. Or it may just be an artefact of the way the brain is organised: we do not know and for the present purposes need not know as I will concentrate on the occurrence and the consequences of contagion in the face, or *facial contagion*.

1.1.5.5 Facial and emotional contagion

Early critics of the link between facial contagion and emotional contagion called into question the extent to which early studies could verify the intensity of the muscles, duration and stability of the expression, and the statistical effect size of said effects (Matsumoto 1987). However, these concerns dissipated when facial electromyography (EMG) began to be used to elucidate the covert characteristics of facial contagion. Indeed, as EMG elucidated the covert characteristics of facial contagion, it became well established as an involuntary process that can be imperceptible to both subject and observer.

As a non-invasive technique, EMG involves measuring tiny electrical impulses produced by contracting muscles from surface electrodes that track the most subtle transient changes (technical details of this technique are provided in the introduction to Chapter 2). A seminal study employing facial EMG demonstrated the value of this method by differentiating between depressed and non-depressed patients when engaging in emotional imagery (Schwartz et al 1976). They found that imagining positively valenced thoughts induced higher zygomaticus activity in the normal population than the depressed patients while negative valence thoughts increased corrugator activity in the depressed patients. The zygomaticus major is mainly

responsible for pulling the cheeks into a smile whereas the corrugator supercilii is located between the eyebrows and is responsible for frowning (Hjortsjö 1969).

Ulf Dimberg (1982) later showed that spontaneous congruent facial contagion could be measured with facial EMG in healthy participants passively looking at pictures of happy and angry facial expressions. Subsequently, a series of studies elucidated the intricacies of facial contagion and were demonstrably more pronounced in women than men (Dimberg and Lundqvist 1990; Vrana and Rollock 1998; Wild et al 2001), higher in response to genuine depictions of emotions (Surakka and Hietanen 1998; Krumhuber et al 2013, 2014) and though observable also in response to virtually generated faces (Weyers et al 2006), dynamic stimuli induced stronger responses of the congruent muscles than still images (Weyers et al 2006; Sato and Yoshikawa 2007). For many years, researchers believed the muscles of facial expression were subcutaneous voluntary muscles (e.g. Waters 1987). Recently, facial contagion was demonstrated with facial EMG in response to backwards-masked stimuli presented too briefly to be consciously perceived (Neumann et al 2014), emphasising its immediate automaticity in humans while it was known early on that it could occur within 1000 milliseconds (Ekman et al 1972; Schwartz et al 1976). Perception is thus sufficient for facial contagion to occur involuntarily in humans without the requirement to mentally compare our facial musculature with resulting mimicked facial expressions.

Moreover, blocking contagion by manipulations which interfere with certain muscles such as biting on a pen (preventing the zygomaticus from responding to happy faces congruently), or through Botulinum toxin injections to the frowning muscles (preventing facial contagion of the corrugator), was found to selectively interfere with the recognition of the corresponding emotional expressions (Oberman et al 2007; Neal and Chartrand 2011), the speed of emotional expression recognition (Lydon and Nixon 2014), as well as decreasing the associated felt emotion (Davis et al 2010). These results highlight the important contribution of facial contagion to the emotional experience in humans as the contracted emotion appears qualitatively the same as a spontaneous one. Indeed, though facial contagion occurs unconsciously and automatically (Dimberg et al 2000) it is modulated by higher order cognitive and emotional mechanisms such as degree of empathy (Sonnby-Borgström and Jönsson 2003; Dimberg and Thunberg 2012) and representation of self and others (Sonnby-Borgström and Jönsson 2003), so that it has been suggested it may be a useful mechanism of emotional regulation through emotional contagion (Hatfield et al 1994).

Primitive emotional contagion refers to the tendency to automatically synchronize with another's emotions as a consequence of non-verbal cues. This could allow humans to share and understand others' emotions during social interaction (Hatfield et al 1994). This forms part of the greater umbrella term of emotional contagion which refers to a "multiply determined family of cognitive, psychophysiological, behavioural, and social phenomena" (Hatfield et al 2014a). This more general definition allows the inclusion of a range of its studied aspects including multimodality (e.g. seeing an angry face leads to raising once voice in anger), and multiple levels (as more than one individual may be affected by the process and to different degrees). For brevity's sake, the use of the term "emotional contagion" in the following thesis will be used as shorthand for "primitive emotional contagion" as is commonly employed in the field (Hatfield et al 2014a).

Many researchers consider that facial contagion is one of the primary mechanisms involved in emotional contagion (Bavelas et al 1986; Fischer et al 1990; Lundqvist 1995; Chartrand and Bargh 1999). Indeed, it has been described as part of a three-step process whereby 1) perceiving an emotional facial expression 2) triggers facial contagion in all its form including muscular and neural feedback, leading to the emotion felt in the perceiver (emotional contagion; Hatfield et al 1994; Schneider 2008). Nevertheless, facial and emotional contagion are not temporally dissociable in the process of arising emotional states. When you contract a smile, you seem to also contract the emotion that goes with it: facial and emotional contagion go hand in hand. However, we do not know what is driving what, and cannot assume that it is one rather than the other. Their temporal relationship thus remains indeterminate as facial and emotional contagion have not been established to belong to a serial system as is often assumed (e.g. Schneider 2008). No experimental manipulation of facial contagion has extinguished the ability to perform the emotional tasks in humans. Rather it has only attenuated success or ease to perform such tasks. Thus, considering them as interacting in a parallel rather than in a sequential system could more easily reconcile existing data.

Contagion can therefore be conceptualised as an anticipatory characteristic of the neural affordance action system. Since the brain is a predictive system, tuned to automatically generate affordances (potential actions) about every encountered situation, object, and individual (Gibson 1977; Cisek 2007), it is possible that contagion forms part of the anticipatory mechanism. In this context, facial contagion is to perceiving facial expressions what Pavlov's dog's salivation was to the bells: an

anticipatory reaction that could advantage the subject in integrating into the situation and extracting maximum potential benefit from such preparedness.

Facial contagion has therefore been established as a universal, rapid, and automatic motor reaction to the emotional facial display of others, occurring spontaneously irrespective of type of stimuli used and with a strong underlying physiological basis. It is therefore a good candidate phenomenon to translate into practical applications.

1.1.6 Scope of the thesis

In this section, I have introduced the general foundational aspects upon which the investigations in this thesis are based. Namely, a high-dimensional, data-driven approach to scientific research can create original contributions to both expand and test the current knowledge base while providing insight into the applications of well-established phenomena. The face is justified as the centre of this translational endeavour whereby the way it is parametrised, and its associated emotional and facial contagion processes will be tapped to exert effects of clinical significance.

The remainder of the thesis is composed of six experiments organised in the following four chapters and followed by concluding remarks. The first chapter proposes an innovative signal filtering technique for the detection of facial contagion as measured by EMG. The following chapter presents the first investigation of the facial contagion response within a sample of patients afflicted by Motor Neuron Disease and explores the potential for this to be translated as an early marker diagnostic for a subtype of the condition. The chapter that follows explores whether emotional contagion can be hijacked and delivered through a smartphone application to enhance positive feelings as a freely available public intervention. This section contains three experiments depending on the type of feeling being affected (mood vs emotion), duration of the intervention, and studies also the potential to gamify these interventions. The last chapter explores whether SVMs can be useful in predicting the physiology of a large Tibetan population by parametrising the face itself as a high-dimensional predictor variable. All chapters will follow a standard scientific journal format whereby every chapter will be introduced by a treatise of the specific subject matters, followed by methods, results, and a discussion of the implications.

Chapter 2: An adaptive filter to better capture the physiology of facial expression

A statue in Pompeii excavated in 1978 where the archaeologists followed the expected shape of the statue to remove the ashes.

2.1 Introduction

The surface Electromyography (EMG) signal is widely used to measure facial contagion responses in research environments. EMG is utilised in a number of clinical and research settings beyond the scope of this PhD, so this chapter will only refer to facial contagion research as measured by EMG. Basic clarification of the underlying principles of the EMG signal and recording techniques are essential to clarify the logic of the methodological innovation presented in this chapter. The overarching goal of this chapter is to explore if the physiology of facial expression as captured by EMG can be made more sensitive to allow for better prediction.

2.1.1 Principles of EMG

The EMG signal measures electrical currents generated by contracting muscles as a result of neuromuscular activity over time. It is a complex biological signal that is described in terms of amplitude, frequency, and phase as a function of time. This signal can be invasively measured by needle electrodes intramuscularly or with minimal intrusion with sticky electrode pads on the skin surface over the muscle. When mentioned subsequently EMG will refer to the latter method, surface EMG which was selected in this study due to its non-invasive and safe qualities.

2.1.2 The Neurophysiology of EMG

Facial muscles are striated skeletal muscles that contract as a result of action potentials firing from motor neurons. Motor neurons relay messages from the spinal cord to innervated muscles to initiate contractions and control relaxations. Activity in these muscles originates from the centre of the fibre and propagates down the length of the muscle through polarisation and de-polarisation. Specifically, the voltages between the extra- and intra- cellular cell environments are different and when a neuron stimulates the muscle fibres, it causes depolarisation. This is defined by a movement of ions that constitute the signal that propagate along the membrane surface and generate an electric field near each fibre that is picked up by the electrodes. This source of electrophysiological current is defined as linear quadrupolar (Daube and Rubin 2009) consisting of two adjacent current dipoles of opposite orientation placed end-to-end (←[-+] origin [+-]→).

Prior to amplification by recording devices, the EMG signal ranges in amplitude between 0-10 mV (-5 to +5). EMG recordings measure extracellular group currents termed motor units - rather than single cell activity - whereby the fibres of different motor units are intermingled. When summed across larger parts of the muscle, these motor unit action potentials (MUAP) are measured indirectly as waves and constitute the EMG signal which can be further subdivided into individual MUAPs through decomposition (Fig. 2.1).

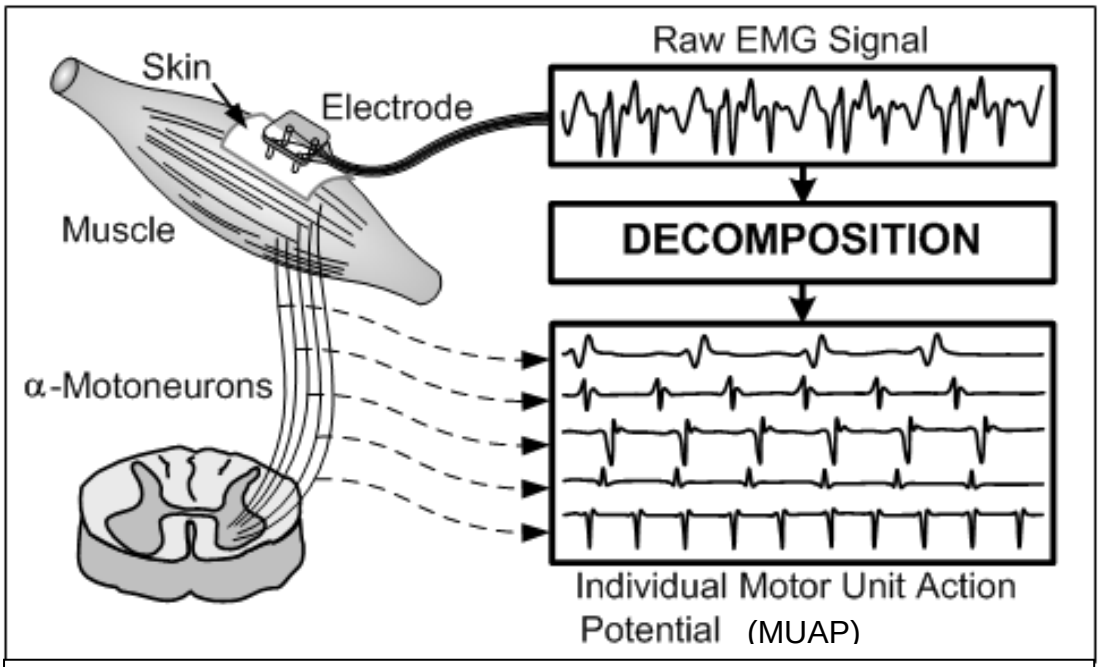


Figure 2.1. Acquiring the EMG signal and MUAP. Modified from http://www.bu.edu/iss/files/2010/12/EMG_signal_decomposition.png

Since MUAPs fire at random intervals, at any point, the measured signal can be either positive or negative in voltage. This renders the EMG signal non-stationary in nature as minute contractions translate into variations of the frequency content of the signal continuously varying over time.

2.1.3 Recording the EMG signal

Facial EMG recordings are commonly performed with electrode-pairs placed on the skin surface parallel to the fibres of the target muscle. Guidance on the positioning of electrodes over the Corrugator Supercilii, Zygomaticus Major, and ground electrodes were developed by Fridlund and Cacioppo (1986). Ground or reference electrodes are to be placed over electrical inactive tissue where the risk for common mode disturbance

will be minimal (Hermens et al 1999). Inter-electrode distance in bipolar montages, defined as electrode centre-to-centre distance, are to be kept between 1 and 2cm to minimise cross-talk and obtain stable recordings (Hermens et al 1999; Konrad 2005). This is preferred over single electrode, monopolar montages where one electrode is placed on the target muscle and the other is a ground electrode because the signal generated by facial muscles is very small.

Indeed, bipolar recordings were one of the most important developments when introduced to EMG research to decrease the level of noise as common signals present in both electrodes are eliminated and differences amplified (Basmajian and Stecko 1962). Additionally, specificity is gained when recording bipolarly as 2 waves are measured per MUAP firing while the signal is monitored propagating in both directions. Nevertheless, as the EMG signal is acquired after travelling some distance from the muscle fibre, through layers of skin and tissue there are many sources of noise that require solving.

2.1.4 The problem of noise

Investigating the EMG signal is complicated by a multitude of unwanted interference, obscuring the underlying facial muscle response. Although EMG has a reputation for being “too easy to use” (De Luca 1997), obtaining a useful and clean signal is quite difficult as evidenced by studies reporting corrupt data can be responsible of excluding upwards of 20% of participants (Ardizzi et al 2014; Hofree et al 2014).

Contamination of the signal can originate from many sources and electrical signals that are not part of the desired signal are defined as artefact or noise. These may include power line interference, electrode-to-skin impedance, motion artefacts, modulation with muscle contractions, and other instrumentation noise (Chen and Xie 2004; Reaz et al 2006; Chowdhury et al 2013a). Detecting a signal of interest from massive body muscles such as the biceps or quadriceps can be easily performed visually once digitally imported and amplified even if noisy (Konrad 2005). However, the face musculature requires more meticulous attention and preparation as the muscles here are significantly smaller in size and surface area and are thus more susceptible to interference.

Thorough cleaning of the skin, careful placing and securing of electrodes, proper isolation of the amplifier and shielding of other electronic equipment are all best practice to ensure a cleaner signal is acquired (Hermens et al 1999, 2000; Clancy et al 2002).

Despite preparatory measures, some noise always remains in the signal and is best tackled by the application of digital offline filtering procedures. The conventional approach to increase fidelity of the signal has been to remove the noise and extract the data of interest by conducting a series of pre-processing filtering steps.

2.1.5 Conventional approach to EMG signal processing

The conventional method to analysing EMG signals was identified from the most common steps undertaken by authors in the field[2]. Most authors applied the steps delineated below to identify the facial contagion response:

1. *Import of data*: the data is imported into a digital analysis software;
2. *Outlier removal*: Incomplete data and outliers are removed by digital or visual inspection;
3. *Epoching and filtering*: the data is epoched and digitally high- and low-pass filtered;
4. *Rectification*: the data is converted into positive polarity;
5. *Smoothing*: The data is smoothed with a moving average filter;
6. *Averaging*: The data is collapsed across left and right muscle channels and averaged across trials of the same stimulus presentation;
7. *Binning*: The data is binned over 100ms bins during the period of expected facial contagion (1-2 seconds after stimulus presentation);
8. *Statistical analysis*: t-tests are used to compare responses to emotional stimuli at every time point across muscle sites and the best sites are selected.

Many of these steps are important and practically necessary. Importing the data is a requirement to apply modern digital analysis techniques. Analyses of the signals post-processing performed on each muscle separately accounts for differences in responses due to muscle size variations and inherent differences in the automaticity and shape of the reactions (Rinn 1984; Larsen et al 2003).

Trialling by dividing the time-locked activity in epochs reflecting expected response time is also a requirement to individuate responses to the stimuli of interest. Subtracting between 100-1000ms of activity before stimulus presentation to account for variations in baseline can account for variable changes within participants' responses as the signal studied is non-stationary. Averaging results across the left and right side

[2] The criteria for selection of these papers will be discussed in more detail in section 2.2 and are summarised in Table 2.2.1

of the face in windows of 100ms (binning) for the duration of the expected response allows for stronger statistical comparisons.

Removing outliers (usually defined as being 2 standard deviations away from the mean) is also a sound step, because it is necessary to identify the inevitable noise contamination that occurs from swallowing, blinking, speaking, and other such unwanted artefacts. Smoothing the data in varying degrees can better reveal the underlying trend of the activity over time. This step effectively takes the envelope of the signal and removes some of the high frequency components and since the response of interest is inherently low frequency as it changes over 1-2 seconds, it is also adequate.

Full-wave rectification turns negative values in a signal into positive ones, yielding only a positive polarity signal (Fig. 2.2). This step is necessary to address the bipolarity of the chosen electrode montage and the depolarisation of the signal. In other words, signals acquired by the "**+**" electrode appear as positive voltage and those measured by the "**-**" electrode appear as negative de-activation which, when averaged, would have 0 mean by cancelling each other out (Reaz et al 2006). Since polarity of the signal is indicative of direction of the firing on the muscle rather than activation or inactivation, full-wave rectification is crucial to accurately display the power of the EMG signal. Therefore, rectification, or taking the absolute values of the signal, allows for the correction of the difference in potentials between the electrodes and constitutes a conceptually sound step to address issues of bipolar electrode recording.

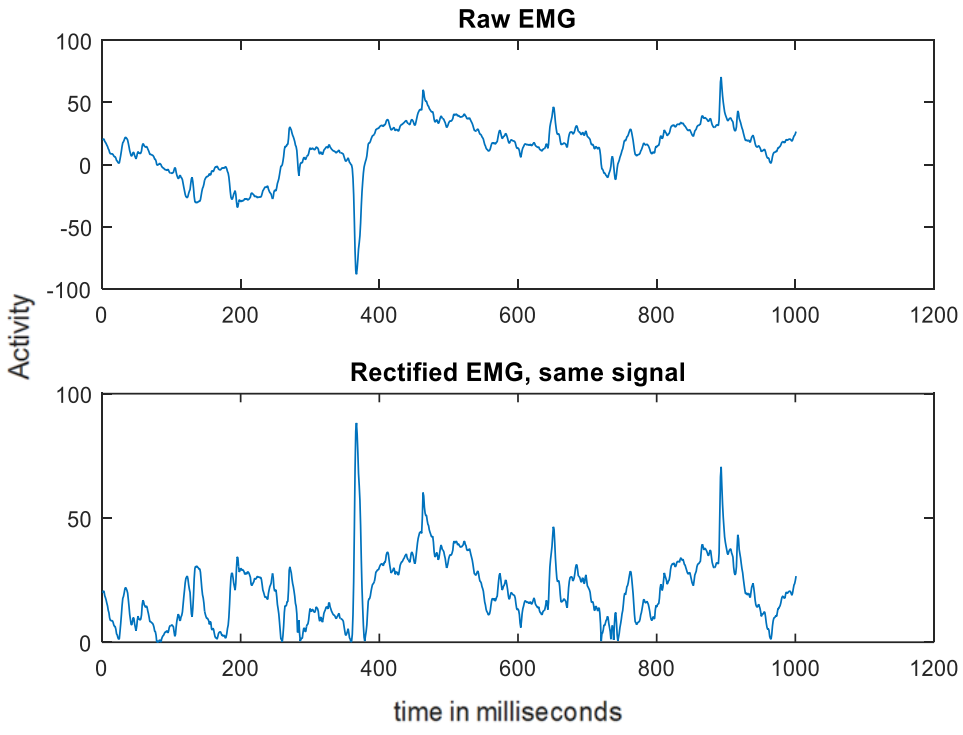


Figure 2.2 EMG raw recording (upper trace) and after full-wave rectification (lower trace)

Bandpass filtering of the activity is a process that has remained virtually unchanged for decades (Kreifeldt 1971) and is performed to increase the signal-to-noise ratio (Konrad 2005). The primary energy of the EMG signal is considered to lie between 10 and 200 Hz (Fridlund and Cacioppo 1986). Filtering is generally set to allow this spectrum of frequencies for the signal of interest to pass while removing data falling outside. The construction of the filters can take the form of separate low- and high-pass filters or a single bandpass filter with the same desired effect (Fig. 2.3). Additionally, a notch filter at 50 Hz is sometimes imposed to eliminate power-line noise and can be either applied during recording or afterwards (Mewett et al 2001).

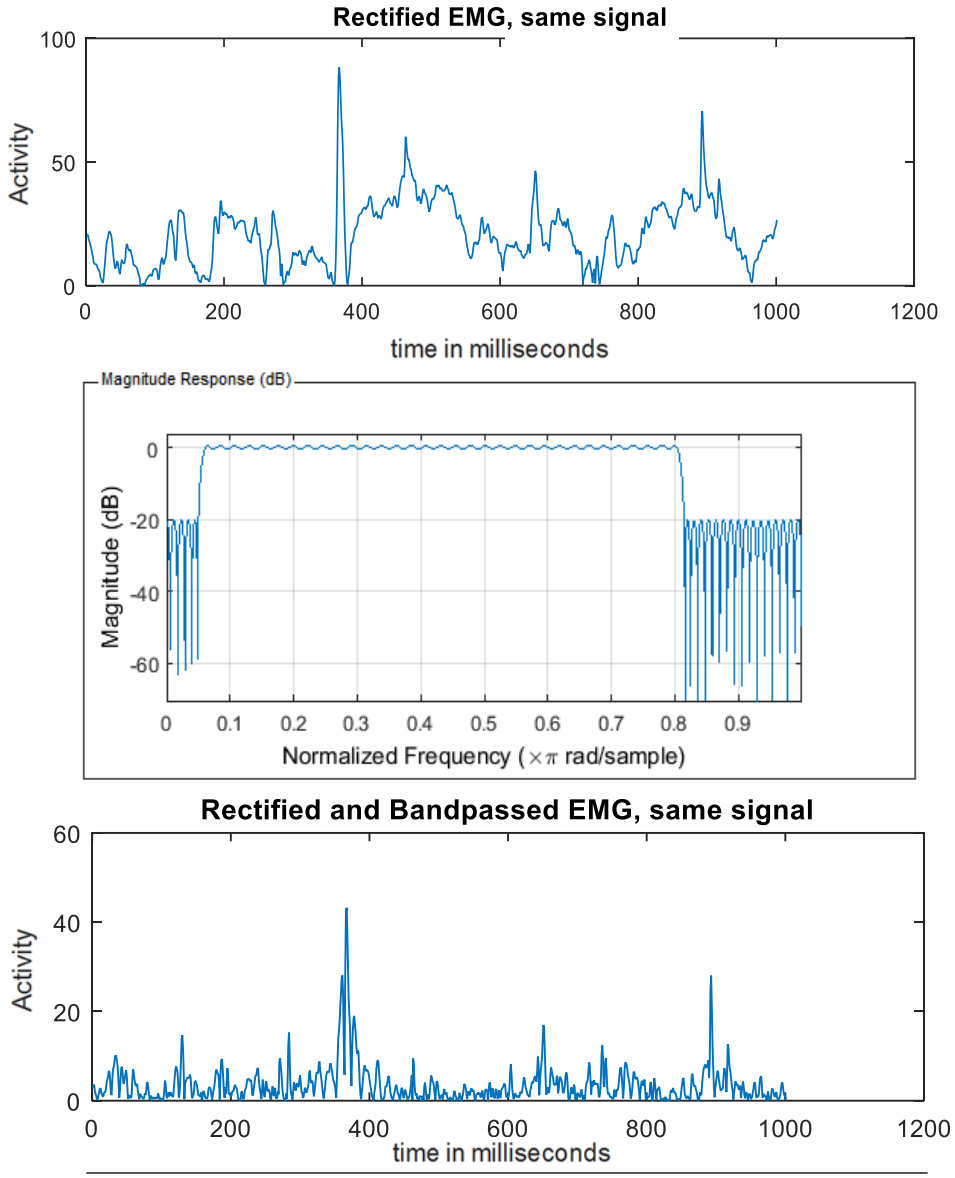


Figure 2.3. Conventional filtering in EMG. A rectified EMG signal (top), a Bandpass Equirriple filter set to allow frequencies between 30-500 Hz to pass (middle), and the resulting filtered signal (bottom).

2.1.6 Issues with the conventional approach to pre-processing EMG

As the EMG signal is non-stationary and non-linear in nature it is one of the hardest forms of time varying signals to study. When studying small facial muscles with EMG, its complexity is further complicated since it is not only a muscle responsive to voluntary control but also to emotional non-voluntary influence (Dimberg et al 2000). A closer inspection of the conventional method for analysis of the EMG signal reveals issues.

As discussed, most of the steps involved are conceptually sound, however *filtering* as it is conventionally performed when attempting to reveal facial contagion responses, is more problematic. By bandpassing a signal one is selecting which frequencies to allow through and which to stop. This process effectively defines which parts of the signal are most likely to construct the phenomenon of interest that is being investigated by discarding data considered to be "noise". Of all the steps, this is the most invasive one. This reductive approach assumes that noise reflects uninteresting data that is not characteristic of the signal of interest. This therefore discards the possibility that within that noise lies data that reflects individual variability that could inform the facial contagion response.

Moreover, there is no consensus concerning which exact frequencies truly constitute the facial contagion response. Although the majority of studies will extract signals between 30 and 500 Hz, some will consider lower or higher thresholds (e.g. Van Boxtel (2010) bandpassed between 15 and 400 Hz and Fridlund and Cacioppo (1986) suggested looking at 10 to 200 Hz). It is difficult to favour a particular range over another since none of these are empirically founded. Most of conventional practices are based on pioneering work and so practice is based on convention rather than empirical evidence. The small tweaks in values to these filters are generally performed without justification, leaving few authors performing studies that will extract informative guidelines based directly on the underlying physiology (see Van Boxtel (2001) for a unique effort).

There are therefore some points of contention with this method that would benefit from clarification prior to imposing such stringent criteria on data analysis. The present chapter cannot address all the issues that affect research attempting to elicit the facial contagion response. From inter-trial intervals varying between 500 milliseconds (Aguado et al 2013; Ardizzi et al 2014) to 34 seconds (Weyers et al 2006) and stimuli presentation times above conscious perception threshold ranging from 600 milliseconds (Künecke et al 2014) to 10 seconds (Matzke et al 2013), there are many potential areas for improvement upon the methods and guidelines to investigate the

facial contagion response. Therefore, this chapter will rather focus on proposing an alternative filtering method which aims to reveal greater differentiation of the signal in line with the current high-dimensional translational approach.

2.1.7 Model-based filtering

Since the signal of interest is non-stationary and non-linear but consistently characterised with EMG in the literature, the novel filtering method exploits a basis function that is itself empirically derived, falling within model-based analyses.

Model-based analysis has an implicit assumption that each response conforms to a certain class of shapes while the individual shape can still vary by a large degree. These models can specify how the underlying physiology generates the observed data in the form of a template response or Canonical Response Function (CRF). These CRFs capture the basic expected shape of the activity and may be useful in describing the underlying activity. Statistical inference based on these modelled responses provide certain advantages including more power as the models are more informed than unconstrained methods (Bach et al 2013). Additionally, since the parameters of the model provide a quantitative and explicit description of expected responses, precise testing of these assumptions is possible. It is useful to refer to similar fields of application in exploring the advantages of contemporary uses of model-based filtering methods.

For example, sympathetic arousal can be quantified by measuring event-related skin conductance responses (SCRs). The SCR is generated by perspiration resulting from instantaneous burst firing of the sudomotor nerve (Boucsein 2012) with higher amplitudes implying higher sympathetic arousal responses. Various methods adopting a model-based approach demonstrate increased sensitivity compared to their conventional counterparts (Bach et al 2010; Bach and Friston 2013). These methods elucidate arousal from SCR by positively relating the CRF to nerve activity and modelling ensuing SCR amplitude. The CRF is highly constrained by the well-established literature describing the SCR and incorporates a descriptive model of the sudomotor nerve sweat response and a model of the short neural bursts occurring immediately after each stimulus presentation (Bach et al 2009).

Similarly, the main inspiration for this type of analysis originates in neuroimaging research where changes in cerebral blood flow as measured with functional Magnetic Resonance Imaging (fMRI) rely on modelling a canonical Haemodynamic Response

Function (HRF). The details underlying biological processes that lead to this signal arising and being captured by these methods are beyond the current scope of this thesis though there are excellent works on the subject (e.g. see Friston 2005 and Ogawa & Sung 2007). In brief, the function is assumed to reflect how the system responds to a short, intense period of neural stimulation (Friston et al 1995). A Canonical HRF is characterised by a fast rise from baseline with a peak around 6 seconds, followed by descending activity that continues under the baseline after 12 seconds and readjusts slowly (at around 30 seconds; Gitelman et al 2003). Each part of this CRF models different but consistently reproduceable shapes of the observed neural activity as measured with fMRI over time. It is therefore possible to *convolve* or smooth a signal with such a function and obtain more reliable results that match the shape of underlying activity to the CRF and filter out more noise. However, common caveats to modelling approaches have been identified (Monti 2011) which assume homogeneity in HRF parameters across individuals and may lead to fitting physiologically implausible shapes. Nevertheless, once these CRFs are specified for each group-of-interest, these approaches best demonstrate their value and are useful in making differences between conditions or groups more evident (Steffner et al 2010; Arichi et al 2012).

Since the facial contagion response as measured by EMG from the Corrugator and Zygomaticus muscles also exhibits a stereotyped response across individuals and stimulus types and is similarly well-grounded in the literature, using such a CRF-driven approach is justified and should provide a more effective way of elucidating the response.

2.1.8 Pompeii adaptive filtering

The method that I developed is an automated model-based filtering method that is data-driven and therefore agnostic to where the best frequencies lie. In brief, it deconstructs the signal into a set of component functions and assesses all possible combinations to overlap them with a CRF derived from the literature in the field. The response most correlated with the template signal is then automatically extracted for each trial, thereby filtering the data while assuming as little as possible about which frequencies are to be discarded.

Whereas conventional approaches to extracting the facial contagion response would discard large parts of the signal *a priori*, the current filtering technique is informed by the evidence available in the field that characterises this response and uses it as a guide to explore the data. As the current technique adapts its filtering to follow the

constraints imposed by CRFs of the known responses, it is reminiscent of the excavations at the archaeological site of Pompeii that follow the guide provided by unearthed walls and artefacts to reveal an evident structure buried in volcanic ash rather than draw arbitrary limits as to where a structure begins and ends. Thus, if the noise in the recording can be considered the volcanic ash, the signal is the structure underneath, and the filtering innovation can draw inspiration from the careful methods applied at the site of Pompeii.

Pompeii adaptive filtering (PAF) aims to assess whether it is possible to make the detection of the underlying facial contagion response embedded in the ongoing EMG signal and blurred by noise more sensitive. Since there are less chances of the frequencies of interest to be discarded, PAF should prove more powerful. In this section the development of PAF is described in detail, followed by a test of efficacy where the PAF and conventional methods are compared in filtering simulated physiologically realistic data.

2.2 Methods

Here I specify the development of the PAF in detail. All computations were carried out in the MATLAB 2016a computing environment on Windows 10 Professional 64-bit with an Intel® Core™ i7 laptop @2.0 GHz with 6.0 GB RAM.

2.2.1 Meta-analytical CRF creation

The first step involved creating a biophysically informed model of the typical adult facial contagion response to be correlated with actual data. To generate this representative CRF for facial contagion, the literature was searched for relevant articles.

Only peer-reviewed published articles written in the English language were included. *Google Scholar* and the *PubMed* database were searched by entering the following keywords: *facial,* AND *electromyography,* AND *EMG,* AND *faces,* AND *emotion,* AND *Zygomaticus,* AND *Corrugator*. A set of criteria were devised to ensure the literature-based CRF (henceforth known as Meta-CRF) would be representative of a typical facial contagion response while viewing smiling or frowning faces. Minimum inclusion criteria included presentation of smiling or frowning faces while simultaneously measuring facial-EMG responses from the Corrugator or Zygomaticus and presentation of ensuing muscle activity as a graphical or numerical time series. Articles or time-series were excluded if 1) EMG measurements were obtained from clinical populations, 2) if no visual line graph depiction or values were provided for the mean activity responses per condition, 3) if the authors attempted to alter the typical facial contagion response by manipulating the visual procedure, timings, or using psychopharmacological substances or rewards. Figure 2.4 outlines the process and results of each step of the literature search.

This procedure yielded 15 articles whose mean EMG activity were extracted visually and recreated for each muscle and condition for the first two seconds after stimulus presentation. When specific values for each time point were not available, the values were visually matched by overlaying published line graphs with the estimated ones and modifying values accordingly. The values at each time point were averaged for each combination of muscle (Zygomaticus and Corrugator) and condition (angry and happy) yielding four Meta-CRFs. Table 2.2.1 displays a summary of the main characteristics of interest of the final selection of studies.

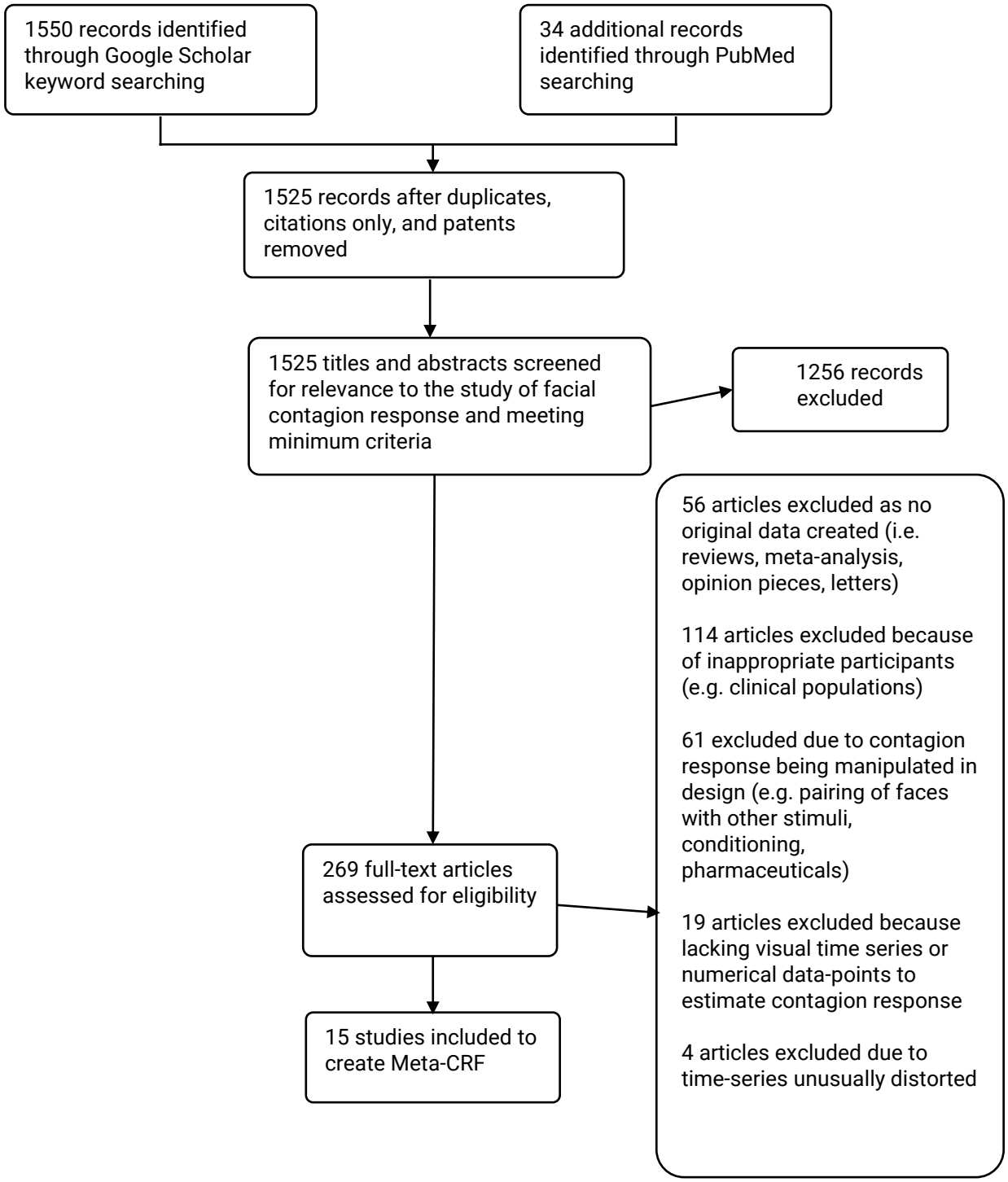


Fig. 2.4. Flow-chart of selection process to include or excluded studies used to generate the Meta-CRFs

Table 2.2.1

Characteristics of final 15 studies selected to generate Meta-CRFs

Study	Type of stimuli	N (females)	Stimulus presentation time in ms**	Inter-Trial Interval in ms	Epoch length / baseline in ms	Bandpass filtering spectrum in Hz	Percentage of N excluded due to EMG corruption
Achaibou et al (2008)	dynamic	15 (10)	1460	2800 - 5200	1400 / 1000	20-400	NA*
Aguado et al (2013)	static	57 (48)	1000	500	3000 / 1000	50-400	NA*
Ardizzi et al (2014)	dynamic	20 (11)	7000	500	5000 / NA*	20-500	22%
Harrison et al (2010)	static	40 (14)	NA*	3500	1600 / 1000	10-480	11%
Heerey and Crossley (2013)	dynamic	35 (29)	300	2000	2000 / NA*	10-400	2%
Hermans et al (2006)	dynamic	20 (20)	6000	6000 - 9000	5000 / 1000	30-NA*	NA*
Hofree et al (2014)	dynamic	48 (30)	6000	NA*	6000 / 2000	10-500	25%
Künecke et al (2014)	dynamic	269 (140)	600	1000 - 3500	1200 / 200	NA*	17%
Oberman et al (2009)	static	13 (0)	25 or 75 or 1000	NA*	2000 / 1000	10-500	NA*
Matzke et al (2013)	dynamic	28 (28)	10000	NA*	10000 / NA*	NA*	NA*
Moody et al (2007)	static	48 (42)	5000	6000	1000 / 500	10-500	NA*
Neumann et al (2014)	static	52 (NA*)	16.67	11250	1100 / NA*	NA*	1%
Rymarczyk et al (2011)	both	30 (15)	1500	15000 - 25000	1500 / 1000	30-500	10%
Weyers et al (2006)	both	48 (48)	1000	25000 - 34000	1500 / 1000	30-500	6%
Weyers et al (2009)	dynamic	49 (49)	8000	25000 - 31000	1000 / 1000	30-500	NA*

*NA = not available, not explicitly stated, or unclear from information provided in study

** ms = milliseconds

2.2.2 Simulated EMG dataset generation

The four Meta-CRFs were spline smoothed and upsampled to 2 KHz to increase comparability with acquired and artificially generated EMG signal simulations. Each Meta-CRF was duplicated 48 times yielding 192 total happy and angry trials from each muscle group. Random white and pink noise vectors of the same length as the Meta-CRFs were created separately and added to simulate an authentic muscular signal recorded with typical EMG equipment.

White noise is characterised by constant power at all frequencies and was introduced as it is generated by electronic recording equipment employed for EMG signal (Chowdhury et al 2013b; Supuk et al 2014). Pink noise was selected as the best type of noise to be mixed into the Meta-CRFs to simulate biological noise. Whereas white noise has constant power spectral density, pink noise has more power at low frequency than at high frequency as it decreases in power per Hz with increasing frequency (Bedard et al 2006). Although the origin of such noise is still unclear, it has been measured across many living systems (Szendro et al 2001; Ward and Greenwood 2007) and as such, is considered a good candidate in creating simulated EMG responses.

In practical terms, the MATLAB function *rand* was used to generate 96 noise signals and multiplied with the CRF to introduce constant machine noise. The *pinknoise* function (Hristo. 2016) was used to create an additional 96 noise signals. After these two vectors of equal length were added, they were multiplied by 12 ascending noise factor levels to simulate increasingly noisy conditions whose output with conventional filtering would replicate low (where the signal appeared very clean), medium (resulting in typical appearance of filtered signals found in the literature), and extreme (where no filtering method should differentiate between the conditions) noise levels.

This resulted in 192 unique noisy simulated EMG responses from both muscle to each condition ranging from very clean recordings to extremely noisy (see Figure 2.5 for depiction of process). This dataset was used to contrast the effectiveness of de-noising the simulated signals and revealing the underlying information between the conventional and PAF filtering methods.

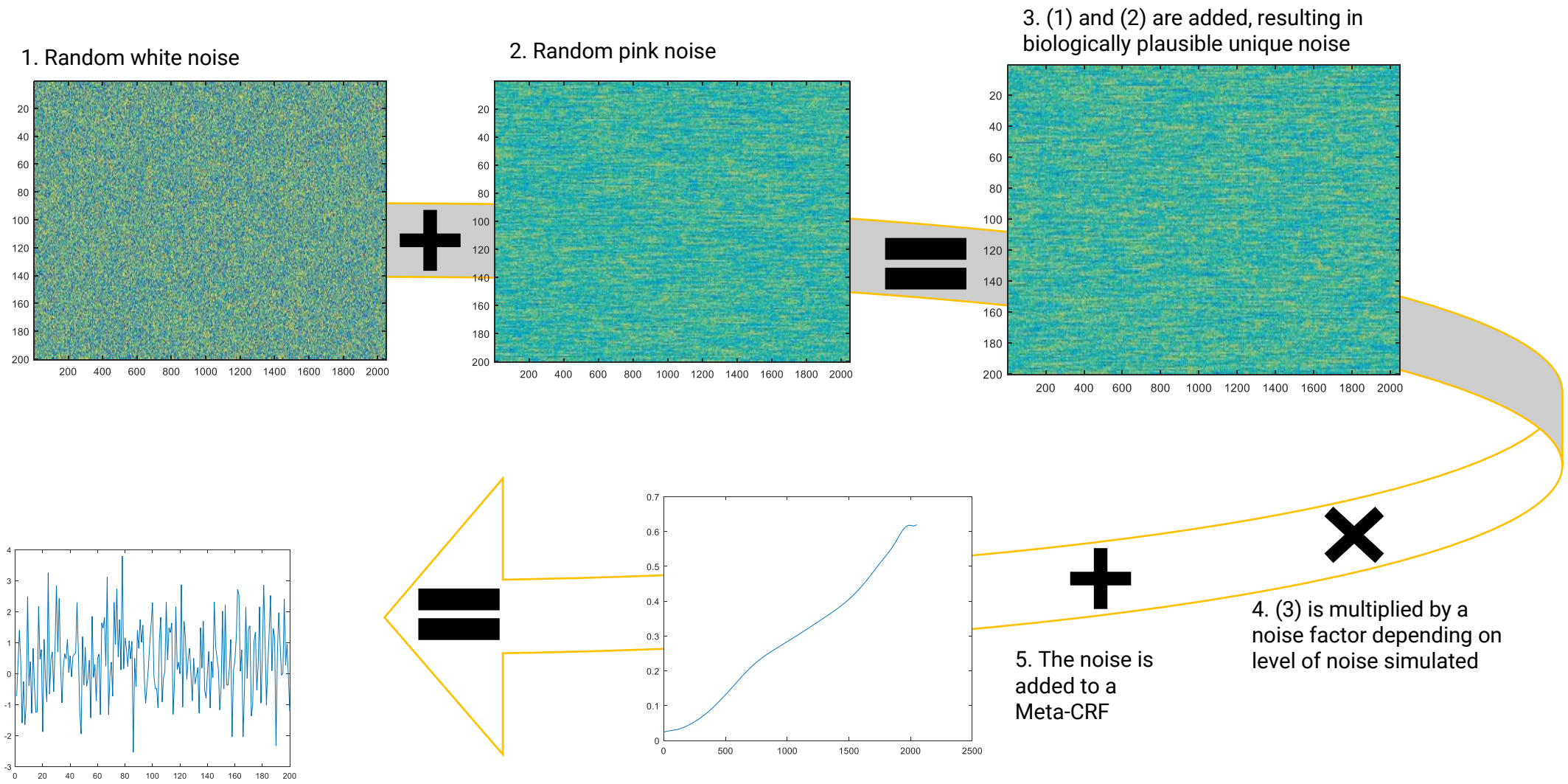


Figure 2.5. Process of generating simulated EMG trials. This figure describes the process by which each EMG simulated trial is generated. Vectors of equal length in milliseconds as the Meta-CRFs are generated containing 1) white and 2) pink noise. 3) These are then added, resulting in biologically plausible EMG noise and 4) the result is multiplied by a factor depending on the level of noise required to be simulated. 5) The resulting levelled noise is added to the CRF, resulting in 6) a unique signal resembling a single noisy EMG trial. This process is repeated 48 times for each of the four CRFs yielding 192 unique noisy trials.

2.2.3 Pompeii Adaptive Filtering and statistical analysis

The simulated trials underwent outlier removal and rectification and were automatically filtered with PAF which consists of three main steps: fission, fusion, and comparison with the Meta-CRF.

Fission, or decomposition of the signal involved the decomposition of the signal with variational mode decomposition (VMD). VMD is an adaptive algorithmic method that decomposes a signal into an ensemble of sub-signals or principal modes, that can reconstruct the given input signal optimally (Dragomiretskiy and Zosso 2014). The first step in this framework involves spline interpolating the local minima and maxima of the signal to estimate the lower and upper envelopes. The mean of the envelope between these extrema is then subtracted from the original signal as a mode where the centre frequency acts as the centre of gravity of the mode's power spectrum and limits its bandwidth (Figure 2.6.B). This sifting process is then repeated on the residual until the specified number of modes is reached. At this last stage, VMD re-balances the extracted modes to capture the full variation of the original signal so that the modes can be reconstituted into the original signal (see Figure 2.6.C).

This concurrent property of the VMD algorithm differentiates it from similar but recursive methods such as empirical mode decomposition (EMD). Indeed, VMD was selected as it fully captures the intrinsic variability of the signal and is superior to EMD which is linear, does not permit specification of the number of extracted modes, does not allow for backwards error correction, and may yield modes that do not necessarily reflect the principal components (Dragomiretskiy and Zosso 2014).

After this fission process, the signal pertaining to a particular trial is decomposed into 6 individual modes which vary from high frequency to a low frequency trend. To reveal which modes most resemble the Meta-CRFs, the first 5 modes are combined into all possible combinations, with the trend being added to each combination. By following a standard combinatorics formula, a total of 31 combination signals are produced, each containing unique information contained in the original signal (Figure 2.6.C).

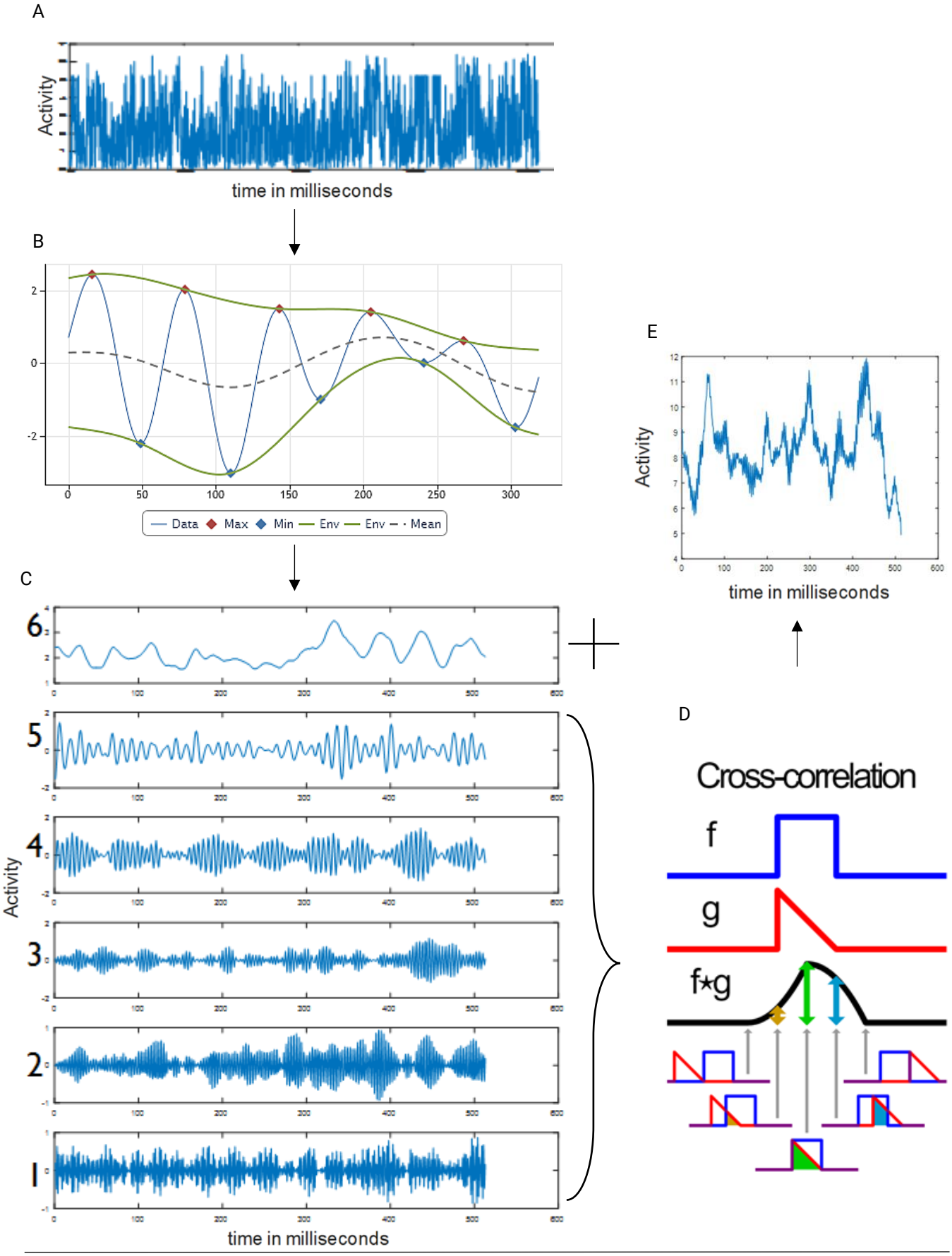


Figure 2.6. Pompeii Adaptive Filtering (PAF). A pre-processed signal (A) is decomposed into its component modes with VMD by spline interpolating the local maxima and minima and subtracting the mean of the envelop from the original signal (B). This is repeated on the residual until the preferred number of modes are extracted, in this case 6, and rebalanced so as capture the full variability of the original signal (C). (D) All mode combinations and the Meta-CRF are normalized between 0 and 1 and cross-correlated (Image from http://tinyurl.com/CrossCorrelation). (E) The highest cross-correlated combination is indexed and used to select the non-scaled combination as the filtered signal for that particular trial.

To compare the selected combination and the Meta-CRF, they were first normalized between 1 and 0 to allow for maximum comparability between the signals. Different methods of statistical dependence between the signals were tested to obtain the highest and most accurate degree of similarity between the CRFs and the resulting combination. The methods selected in this section were those most commonly used in the field and involved iterative testing of the whole process of the PAF with regards to different methods and observing the results as noise levels were increased. Initially, a very basic and simple method testing whether the signals followed the same absolute trajectory was performed using Pearson's correlation coefficient. This looked, in absolute terms, at the linear correlation between the two signals and produced a value between +1 and −1 inclusive (where 1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation). This method however did not produce results that allowed dissociating the signal with enough sensitivity. Therefore, a more complex extension of this method was applied to attempt searching for further correlations between the combinations of the correlated values. This method, called Canonical Correlation Analysis (CCA), finds the optimal linear combination between separate interrelated attributes within the signals and relates them to each other, thereby respecting the higher complexity within signal variations. Although this method was a slight improvement, it did not respect the temporal changes as much as desired as the noise introduced could also shift the signal's property in time.

Indeed, implementing a method called cross-correlation, allowed measuring the degree of similarity between the two signals with a shifting x-axis. After testing different lags in temporal shift, a lag of 500 milliseconds was found to consistently select the combination that most closely resembled the Meta-CRF as compared to manual selection (Figure 2.6.D and E). The key was in shifting the alignment of the signals as individuals may differ as to when the facial contagion response could begin. The index of that combination was then used to select the unscaled version of the signal and extract it as the filtered signal to be used in subsequent analysis.

Recall that there are two Meta-CRFs for each muscle site corresponding to the facial contagion response to congruent and incongruent emotional facial expression (Happy and Angry Meta-CRF). All responses underwent PAF to each CRF depending on muscle site to ensure the method did not over fit results to simply match noise that fits to the Meta-CRF. At the inference stage, this choice of Meta-CRF was modelled in the comparisons as a factor. Specifically, a 2x2 ANOVA was used to compare the means between the activity in response to happy and angry faces at each muscle site, looking

for the main effect of facial contagion related to expression emotion. The effect of CRF was also included in each comparison with the hypothesis that if no separation was found between the signals depending on Meta-CRF, then any difference between responses to angry and happy faces were unbiased by the modelling of the Meta-CRF at the filtering stage. All significance thresholds were adjusted for multiple comparisons with Bonferroni corrections.

2.2.4 Conventional Filtering

Steps 2 to 8 of the conventional approach to EMG signal process detailed in section 2.1.5 were separately performed on the twelve noisy simulated raw datasets to compare the performance of the two filtering methods. Outlier removal was performed by removing all values 2 standard deviations away from the mean. High pass and low pass filters were designed with MATLAB's Filter Design and Analysis Tool (FDATool) to limit the signal to frequencies of interest between 30Hz and 500Hz. Before applying the bandpass filtering, the signals were concatenated and zero-padded to create a single signal and avoid common edge distortions and allowed for epoching. The signal was then rectified and smoothed with a moving average filter (125ms) and binned in 100ms over 2 seconds. Simulated data were then analysed with two samples t-tests at each muscle site comparing angry and happy faces responses. All significance thresholds were adjusted for multiple comparisons with Bonferroni corrections.

2.2.5 Additional testing with Support Vector Machines

For every noise level, each result from the muscle sites were inserted into support vector machines (SVMs) to quantify the performance of the approaches given the levels of noise. LibSVM was used for developing SVM models (Chang and Lin 2016), with the size of the hyperplane margin (parameter *C*) varied from -5 to 15.

For both filtering techniques, the binned time-series from both conditions (happy and angry) were inserted into a continuous dataset for each muscle site and labelled accordingly (-1 for angry and 1 for happy trials). Five-fold cross-validation was performed by taking slices of 75% from each condition to train the model and classify features based on their labels. This resulted in the hyperplane model that fit the data best. This was then selected and used to predict the rest of the dataset over 50 iterations. At each iteration, the resulting predicted labels were compared to the actual labels and sensitivity (true positive / (true positive + false negative)) and specificity (true negative / (true negative + false positive)) were calculated. These assess the model's ability to correctly

identify members that belong to a certain class and quantify how many are falsely categorised. By taking the mean of these two measures, the balanced accuracy was calculated as a single criterion of the predictive performance of the models. The balanced accuracy combination compensates for any potential imbalance in the dataset.

This procedure resulted in 50 balanced accuracy values per noise level that were inserted into two-samples t-tests for each muscle site with Bonferroni corrections appropriately performed for multiple comparisons.

2.3 Results

All results compared the conventional and PAF filtering methods with the same input simulated data. Prior to the more exhaustive comparison with the simulated EMG data at 12 increasing noise levels, two sanity checks were performed to assure the methodology worked as expected. For reporting and presentation purposes, results are presented as graphical time-series where the p-values are represented at the time points were statistically significant for main effect of emotional expression (muscle response to happy vs angry visual facial stimuli).

For all PAF 2x2 ANOVAs, no main effect of Meta-CRF or interaction between the factors reached statistical significance at $p<0.0025$[3] and are therefore not reported below.

2.3.1 Sanity checks

The first sanity check involved running both filtering methods on simulated data with no noise added. This was performed by multiplying the noise by a factor of 0 which was then added to the data. Since the signals are constituted of clean Meta-CRFs, both filtering methods should perform equally well at perfectly dissociating the signals. This was confirmed as all main effects of emotion were significant at $p<0.001$ for both methods (Figure 2.7).

For both the conventional and PAF methods, significant facial contagion was found congruent with muscle site. Conventional filtering found higher activation in response to angry face stimuli than for happy faces for the Corrugator and vice versa for

[3] The threshold was set as $p<0.05$ with Bonferroni corrections for 20 multiple comparisons.

Zygomaticus. The same result was found with PAF independent of Meta-CRF. The first sanity check can therefore be considered *passed*.

The second sanity check involved running both filtering methods on low-level generated noise. If the PAF was imprinting a signal upon the noise, then this check should demonstrate it. In practice, random signals were inserted instead of Meta-CRFs to be multiplied by a noise factor of 1. This sanity check was also passed as no test reached statistical significance across the two filtering methods and no evidence of facial contagion was found (Figure 2.8).

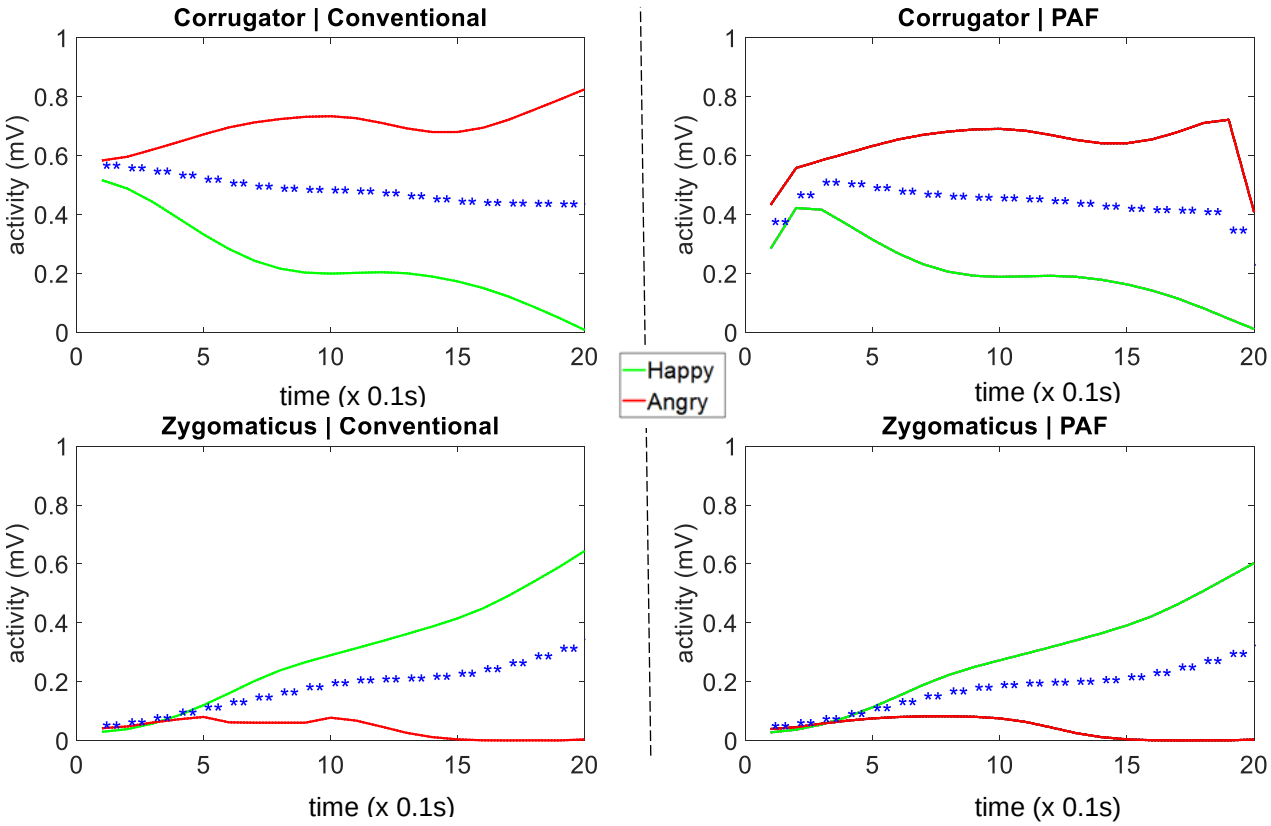


Figure 2.7. Sanity check 1 for PAF. This figure shows the facial contagion responses for 0-noise simulated data with conventional and PAF filtering methods. The legends provide information on different solid line colours for different conditions and dotted lines of the corresponding colour for upper and lower standard error margins. Two blue asterisks (**) at specific time points indicate statistical significance at p<0.001. For the PAF comparison (right), since no effect of Meta-CRF was found, the same legend applies as with conventional methods.

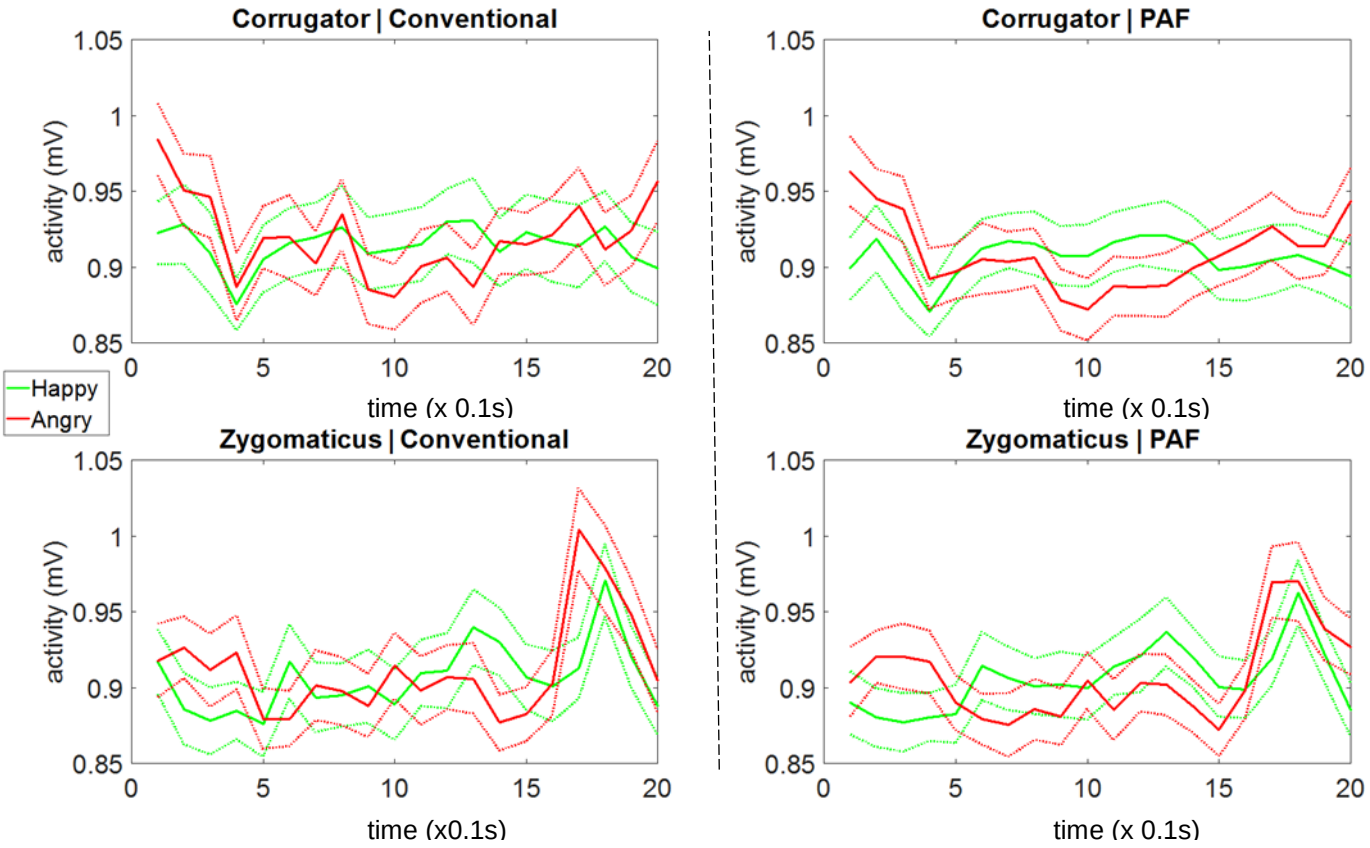


Figure 2.8. Sanity check 2 for PAF. This figure shows the facial contagion responses for simulated data with noise only for the conventional and PAF filtering methods. The legends provide information on different solid line colours for different conditions and dotted lines of the corresponding colour for standard error margins. No test reached statistical significance.

2.3.2 Validation across noise level comparisons

Noise level factor multipliers ranged between 1 and 6.5 rising in increments of 0.5 per factor. In the interest of conciseness, 6 of these will be presented as the intermediary follow the same linear pattern of signal-to-noise degradation.

At a noise factor of 1, both filtering techniques performed equally well at distinguishing the facial contagion response, with clear responses at each muscle site in response to congruent facial visual stimuli (Figure 2.9). Thereafter, with increasing noise levels, the facial contagion response began to degrade in both techniques (Figure 2.10), with the PAF demonstrating more resistance to higher noise levels. Ultimately, when the noise levels reached their higher disturbance levels, neither techniques could recover the facial contagion responses (Figure 2.11).

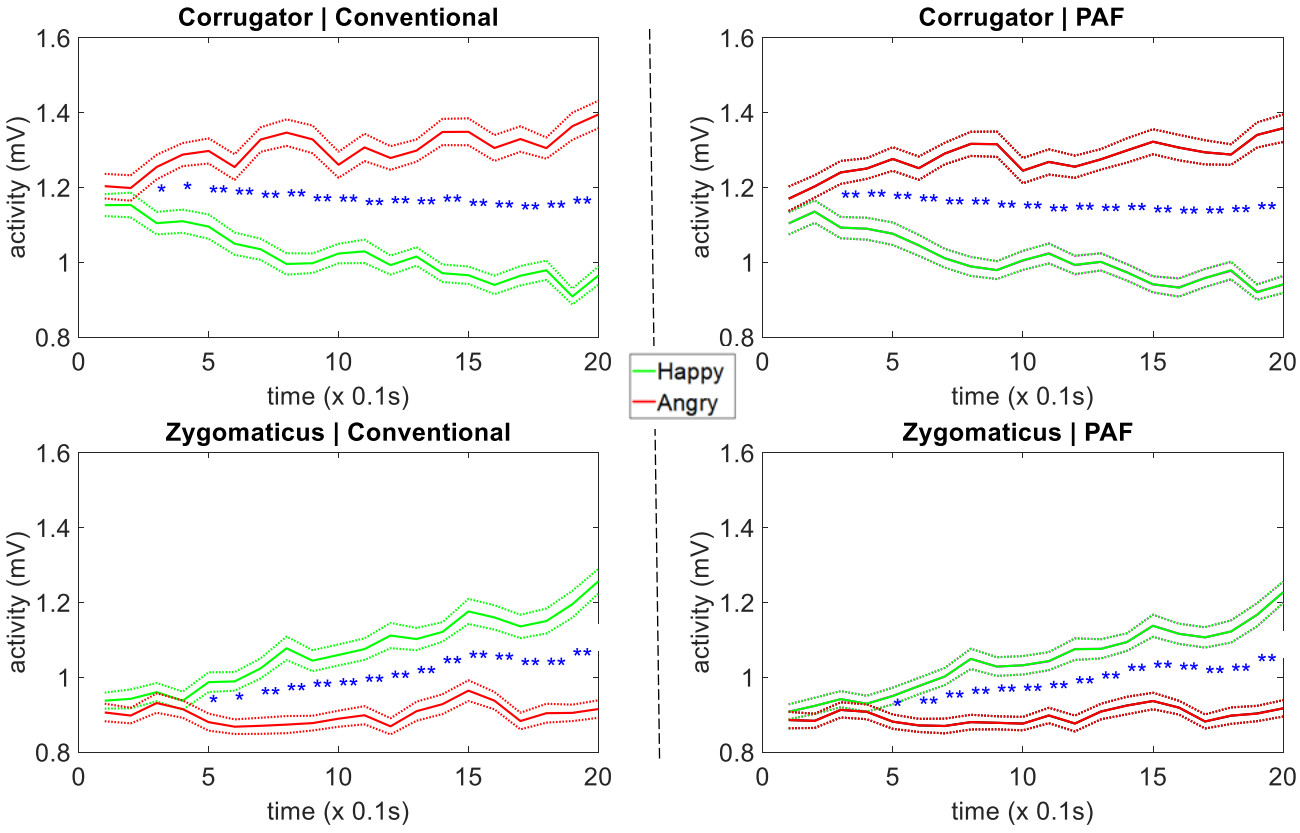


Figure 2.9. Performance comparison of conventional filtering and PAF at noise factor 1. One blue asterisk (*) indicates statistical significance at p<0.05 and two blue asterisks (**) at p<0.001.

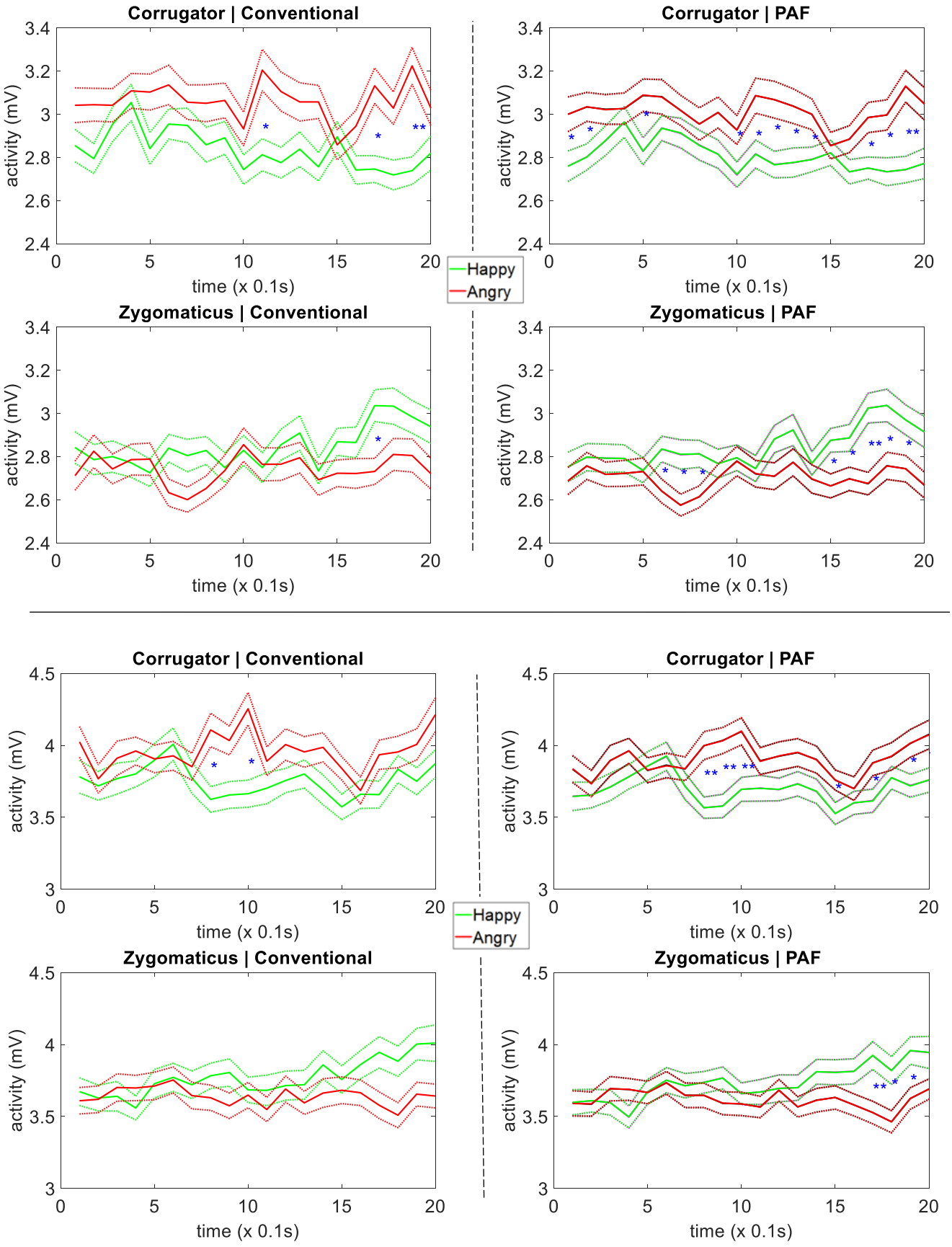


Figure 2.10. Performance comparison of conventional filtering and PAF at noise factors 3 (top) and 4 (bottom). One blue asterisk (*) indicates statistical significance at p<0.05 and two blue asterisks (**) at p<0.001.

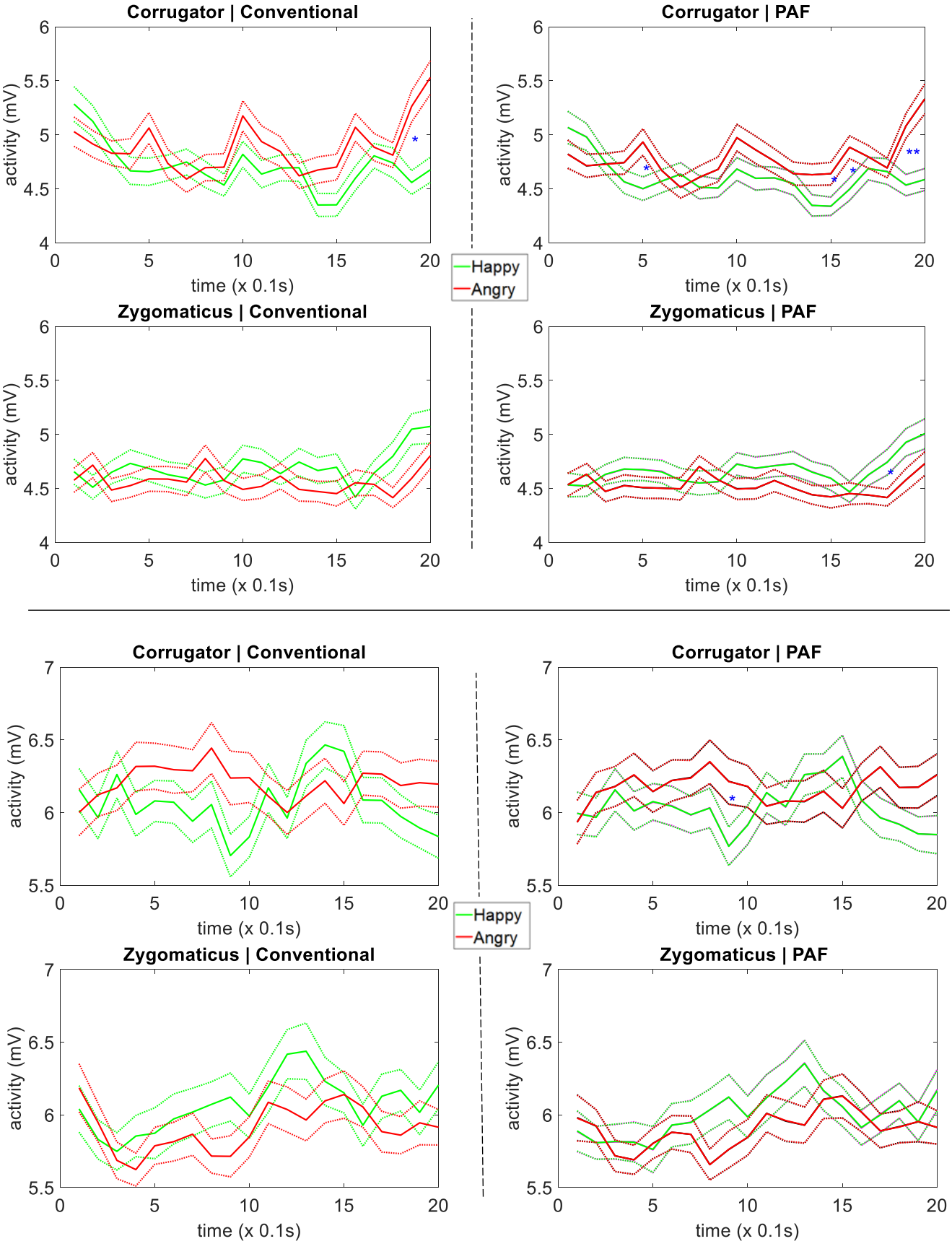


Figure 2.11. Performance comparison of conventional filtering and PAF at noise factors 5 (top) and 6.5 (bottom). One blue asterisk (*) indicates statistical significance at p<0.05 and two blue asterisks (**) at p<0.001.

2.3.3 Validation with Support Vector Machines

As can be seen in Figure 2.12, for both techniques irrespective of muscle sites the best predictive performance was at minimal noise levels with consistent decreases in performance as the noise increased. At the lowest noise level, predictive performance is indistinguishable. At the highest noise levels, predictive performance worsens and both methodologies are similarly indistinguishable at extracting the signal from the noise.

The largest advantage in terms of mean balanced accuracy percentage difference was found at noise level 3.5 for the Corrugator, with a maximum difference of 12.56% and at noise level 3 for the Zygomaticus with a maximum difference of 8.22%. Consistent differences were also found in predictive performance between the Corrugator and Zygomaticus. Specifically, the PAF demonstrated superiority of noise tolerance until level 5.5 for the Corrugator compared to 3.5 for the Zygomaticus.

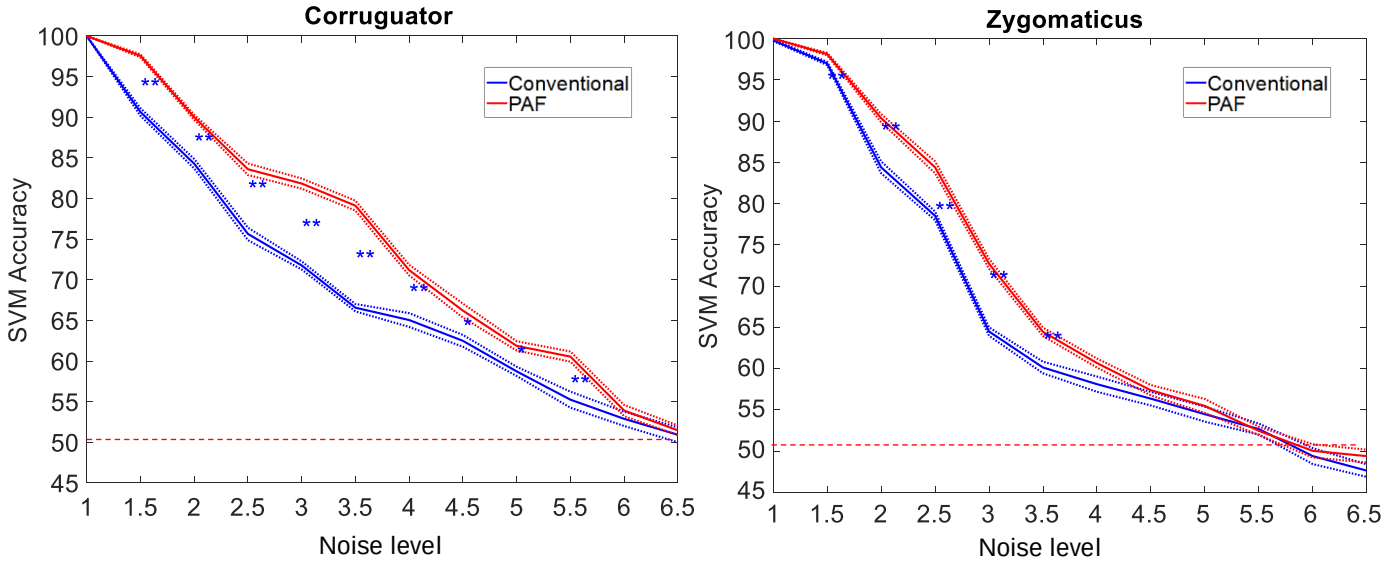


Figure 2.12. Performance comparison of conventional filtering and PAF with SVMs.
The graphs represent mean balanced accuracy percentage values for predicting the correct emotional valence of the signal by each filtering technique based on the simulated data. One blue asterisk (*) indicates two-sample t-test statistical significance at $p<0.05$ and two blue asterisks (**) at $p<0.001$. Thick coloured lines represent different filtering approaches and dotted lines indicate standard error.

2.4 Discussion

2.4.1 Conclusions

The present chapter demonstrated that the PAF is a useful automated approach to extract the facial contagion response to angry and happy facial stimuli as modelled by simulated Corrugator and Zygomaticus EMG activity.

The PAF passed sanity checks indicating that the method was unbiased compared to conventional filtering. The comparison of time-series with increasing simulated noise demonstrated the PAF to be consistently superior to conventional filtering at each muscle site and in response to both conditions. Further testing with SVMs demonstrated the higher efficacy at dealing with noise of the PAF mostly during the medium and high noise levels while performing equally or non-significantly better than conventional filtering at lowest and highest noise simulation levels. This is suggestive of realistic performance as when there are unrealistically low levels of noise, the facial contagion response should be easily discernible and when noise levels are unrealistically high, it is impossible to salvage the signal, irrespective of which filtering method is applied. Overall the results confirm the PAF is a realistic filtering technique; rather than being unaffected by noise, it is more robust than conventional filtering.

Thus, after rigorous and methodologically sound testing the results suggest the PAF is superior to the conventional approach in dealing with simulated noise and revealing the underlying facial contagion response. This demonstrates the soundness of postulating a CRF-driven, model based filtering technique where possible and may circumvent certain points of contention posed by the conventional filtering approach.

2.4.2 Future Directions

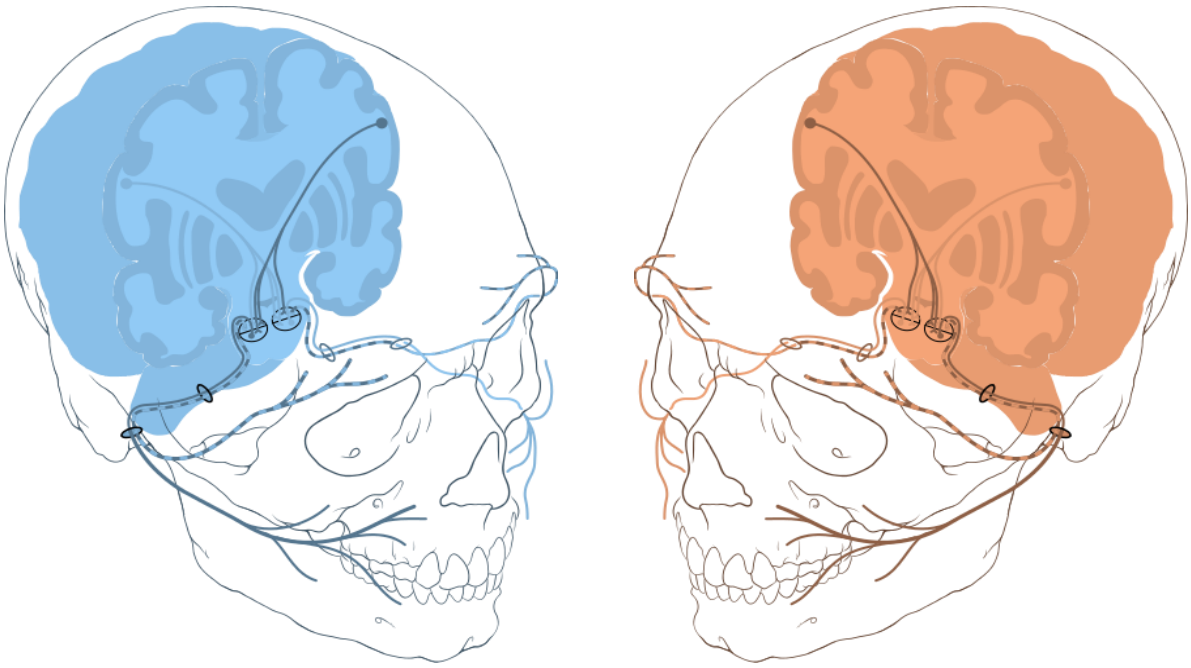
In the current test, the facial contagion response was only modelled in response to two stimuli classes, that is, happy and angry faces and modelled recordings from the two congruent muscle sites (Zygomaticus/Corrugator). The facial contagion response has been established to occur in response to other facial expressions such as disgust (Vrana 1993; de Jong et al 2002; Wolf et al 2005; Oberman et al 2007) and fear (Dimberg 1986; Dimberg and Thell 1988; Oberman et al 2007). Therefore, further investigations could

concentrate on creating site- and emotional-valence-specific Meta-CRFs to explore whether such analyses could potentially benefit from the application of PAF. In the current experiment, other emotions were not parametrised as they are less studied in the literature. This translates to less data available to generate a Meta-CRF which would lead to lower accuracy compared to happy and angry facial contagion responses.

Within the greater context of the PhD the PAF approach fits well as its primary object of study is a facial response. Additionally, it is data-driven as the Meta-CRF comes from historical data and the VMD-based filtering assumes nothing about the data, not even the bandwidth of what is important. It can be considered also high-dimensional, in that the entire waveform is used at the prediction stage to provide further information upon the efficacy of the results. To complete the translational approach in this endeavour, the PAF's demonstrable effectiveness is required.

Since the same Meta-CRFs were used for creation of the data and testing, it will be necessary to test whether the PAF generalises to elucidating a response within data acquired from a human population. This double use of the Meta-CRF was originally necessary to develop the method and demonstrate its efficacy. Although great care was taken in creating a biophysically realistic noisy signal, to demonstrate *effectiveness* the PAF needs to be tested on actual EMG responses elicited spontaneously from a human population. Doing so will further cement the validity of the Meta-CRF as a well-informed signal, descriptive of the facial contagion response and with potential in model-based filtering. This testing is possible and is performed in the next chapter.

Chapter 3: Quantifying emotional dysregulation in Motor Neuron Disease with facial electromyography



3.1 Introduction

3.1.1 Motor Neuron Disease and the problem of diagnosis

More than 1,700 people in the United Kingdom will be diagnosed with a form of MND every year. MND is a term for a range of progressive neurodegenerative disorders whose management remains a major challenge. Whether or not there are one or many aetiological entities still remains a mystery. A marked aspect of the difficulty is that MND is both heterogeneous and diagnosed clinically; it is, properly speaking, a syndrome. In some subtypes there is degeneration of cortical neurons that project from the primary motor cortex to the brainstem and spinal cord through the corticobulbar tract (so called upper motor neuron (UMN) involvement), while in others degeneration may primarily affect motor neurons branching outwards to supply muscles in the rest of the peripheral musculature (lower motor neuron (LMN) involvement).

Detection of UMN involvement is important as it tends to carry a substantially worse prognosis. Patients with suspected UMN involvement may live only 2-3 years compared with those with LMN involvement that may survive indefinitely (Christensen et al 1990; Haverkamp et al 1995; Davenport et al 1996; Turner et al 2002; Scotton et al 2012; Moura et al 2015). As 10-20% of MND patients survive longer than 20-48 months (Chiò et al 2009), resolving this problem of diagnosis could help doctors better inform their prognosis and tailor treatment accordingly. Differentiating between MND subtypes is therefore not a trivial issue. If successful, it would improve categorization which holds a range of potential benefits for recruitment in clinical trials as well as for patients, including earlier administration of best available treatment drugs. For example, small but significant benefits have been demonstrated in MND patients in response to riluzole (Bensimon et al 1994; Miller et al 2007). As riluzole has been reported to only lead to a modest extension of survival (Georgoulopoulou et al 2013), perhaps targeting MND subtypes would lead to improved survival in some subtypes. This differentiation could potentially increase costs of running clinical trials as targeting subtypes might translate to more patients per subgroup needed.

Common forms of the disorder may be characterised by degeneration of UMN and LMN in the absence of alternative causes according to the El Escorial criteria

(Brooks 1994). This is complicated to ascertain as when the disease progresses it may begin affecting both UMN and LMN pathways. Patients whose clinical presentation is deemed to be more characteristic of LMN involvement show evidence of muscle atrophy, weakness, decreased reflexes, and fasciculations.

Patients with primarily UMN involvement generally demonstrate spasticity in the affected territories and when the bulb is involved also often emotional lability. The most common phenotype of the disorder involves simultaneous degeneration of both UMN and LMN and is termed Amyotrophic Lateral Sclerosis (ALS). ALS accounts for over 85% of total cases and although the terms ALS and MND are sometimes used synonymously, it is incorrect. MND encompasses ALS as well as suspected pure UMN involvement (i.e. primary lateral sclerosis), pure LMN signs (aka progressive muscular atrophy), and patients with isolated signs of bulbar dysfunction (aka progressive bulbar palsy).

3.1.2 **Diagnosing Motor Neuron Disease**

The first mention of ALS can be traced back to 1865 when Jean-Martin Charcot (de Paris 1862) described symptoms associated with LMN degeneration in a single patient and published a seminal article on ALS (Charcot 1874), linking LMN signs in the arms and UMN signs in the legs with loss of motor neurons. More than a century later, he is still credited with leading the way towards establishing the modern neurological diagnostic examination for ALS widely used today (Rowland 2001). More recently, electromyographic (EMG) recordings have been incorporated into the diagnostic process. The recommendations require evidence of denervation in a minimum of two muscles in at least two separate anatomical regions innervated by different roots and nerves as measured with needle EMG (de Carvalho et al 2008). Although these changes have resulted in increases in sensitivity to identify LMN involvement in patients with suspected MND, it has not proven equally useful for the detection of UMN dysfunction and may additionally cause discomfort due to the invasiveness of needle EMG (Carvalho and Swash 2009).

UMN degeneration is theoretically rendered visible by the application of diffusion-sensitive MR techniques that can image the corticospinal tract (CST). Nonetheless, identifying UMN signs based on diffusion tensor imaging (DTI) provided mixed results (Ellis et al 1999; Hong et al 2004; Ciccarelli et al 2006; Woo et al 2014) and was most

effective in rare cases of pure UMN or LMN dysfunction (Iwata et al 2011). A limitation of this technique is that it requires expensive equipment and large datasets to achieve sufficient prediction accuracy for individual classification (Mah et al 2014).

Another potential test for UMN involvement has focussed on using transcranial magnetic stimulation (TMS) measures to assess the integrity of the CST. When TMS is applied to M1, this leads to cortical conduction down the CST and activates the target muscle (Barker et al 1985). When recorded from the surface of the muscle with EMG electrodes, the magnitude of the elicited muscle reaction can be measured as a Motor Evoked Potential (MEP) while the time between TMS over M1 and the elicited MEP can be temporally assessed as Central Motor Conduction Time (CMCT). Delays in CMCT or small or absent MEPs have been deemed to be indicatory of UMN involvement (Pringle et al 1992). More recently, beta-band (15-30 Hz) coherence between the primary motor cortex and contralateral muscles was found to be desynchronised in patients affected by a rare form of pure UMN dysfunction (Fisher et al 2012). TMS-based measures, have shown potential in diagnosing pure forms of the disease as well as disease progression (Floyd et al 2009) though not in discriminating cases where UMN involvement is unclear (Ziemann and Eisen 2004), while still requiring expensive equipment and discomfort in an already debilitated clinical population.

Therefore, there are neither clinical features nor laboratory markers that can help us make the distinction between UMN and LMN involvement early on in the course of the illness (Munsat et al 1990; Rosenfeld and Strong 2015). The need for a reliable physiological assessment of UMN function is evident from the limitations of current options. Indeed, though numerous attempts at developing and modifying the formal criteria for diagnosing MND have been made, these efforts are recognized as being more effective at establishing a degree of diagnostic certainty of MND *presence* (e.g. possible, probable, laboratory-supported, or definite), rather than successfully identifying subtypes of the disease or describing a homogeneous group of patients (Rosenfeld and Strong 2015). Thus, the primary purpose of this chapter is to lay the foundations for developing a novel, non-invasive effective test for detecting UMN involvement in patients with suspected MND.

3.1.3 Emotional lability

The facial musculature is directly controlled by corticobulbar nuclei that receive innervation from at least two sets of descending projections. These nuclei receive fibres from the cortex, responsible for voluntary facial expressions, as well as independent fibres from subcortical regions, responsible for largely involuntary, emotional facial expressions (see Figure 3.1). In types of MND where there is UMN involvement, the corticobulbar tract fibres innervating from the cortex are preferentially disrupted, resulting in a pathological predominance of subcortical control (Iwata et al 2011; Floeter et al 2014). Patients so affected report a tendency to express emotion even when the circumstances do not justify it, in a symptomatic disorder termed *emotional lability* (Ironside 1956; Cacioppo et al 1988; Newsom-Davis et al 1999; Palmieri et al 2009; Floeter et al 2014).

Specifically, lability is characterized by episodes of emotional reactions or facial expressions that are out of proportion or unrelated to the eliciting stimuli. Affected patients may laugh uncontrollably at something only mildly amusing or cry at something trivially sad but may also laugh when feeling sad, while usually maintaining lucidity of the inappropriateness or exaggeration of their reaction (Madani et al 2013).

Wortzel et al (2008) found over 25 neurological conditions associated with reports of emotional lability. Incidence of emotional lability in other conditions such as post-traumatic brain injury (Tateno et al 2004), post-stroke (Kim 1997), and Multiple Sclerosis (Feinstein et al 1997) tends to range under 20% while it is highest among patients with ALS. In these patients, as many as 50% may develop symptoms of emotional lability with highest reports in those with bulbar symptoms (Gallagher 1989). Patients affected by these other conditions often report single instances of emotional lability and these may occur only at the early stages of the condition. On the other hand, patients with MND are a large cohort to be likely affected by emotional lability and the symptoms are accentuated with time, adding to the justification of studying them to characterise this disorder.

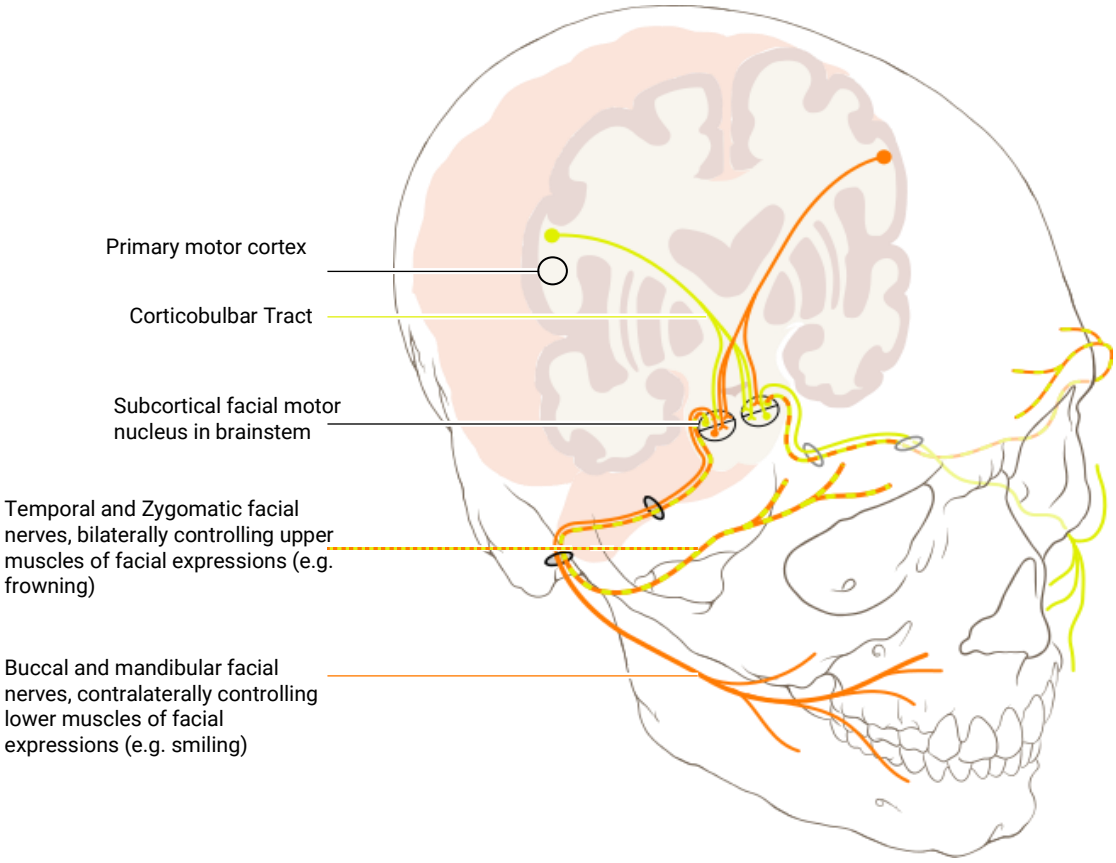


Figure 3.1. Depiction of facial nerve connectivity through corticobulbar tract to cortical and subcortical regions. In UMN involvement the Corticobulbar tract is preferentially damaged leading to subcortical dominance of emotional control and lability. Adapted from Patrick J. Lynch, medical illustrator (Patrick J. Lynch, medical illustrator)

The choice of the term *emotional lability* was selected as most appropriate to refer to this condition although it is not without controversy. Over 20 terms have been proposed in the literature to refer to this state with discordance over whether the feeling expressed is congruent or not, in valence and intensity, with the emotion displayed by the patient (Wortzel et al 2008). These discussions concluded that the original definition of Pathological Laughing and Crying (PLC) proposed by Wilson (1924) captured both these aspects and was therefore the best descriptor. Employing the term PLC to describe this condition however ignores episodes of abnormal smiling which may not progress to overt laughter (Newsom-Davis et al 1999). Accordingly, PLC is too narrow to characterise the full range of the condition and it is therefore recommended that emotional lability be adopted as the preferred diagnostic term for this condition.

3.1.4 Measuring emotional lability

The selection of the appropriate diagnostic terminology is not a trivial matter as reflected by the state of current methods to assess lability through questionnaires in the clinical sector. There are currently two main questionnaires used to diagnose lability in patient populations, the Pathological Laughing and Crying Scale (Robinson et al 1993) and the Emotional Lability Questionnaire (Newsom-Davis et al 1999). The Emotional Lability Scale (ELQ) identified the same limitation (and others) regarding pathological smiling, as the name of the PLC Scale suggests. To overcome this, another dimension was added assessing pathological smiling and provided a more sensitive characterisation of lability.

Moreover, these questionnaires limit their assessments of lability to feelings of sadness and/or happiness and may be further limiting their representativeness of the full breadth of human emotions and ignoring other potential dysregulations (e.g. angry outburst). Also, both rely on patients or carers retrospectively assessing whether there were extraordinary episodes of exaggerated emotional reactions. In many patients the changes are too subtle for their significance to be confidently established on clinical grounds alone and may go unreported. When emotional lability is marked, the diagnosis of upper involvement is more easily made as these episodes may affect patients' everyday life with social anxiety and emotional frailty (Palmieri et al 2009). Since MND is a *progressive* neurodegenerative disease, it is difficult to identify this symptom at early stages of the disease through questionnaires.

As questionnaires are severely limited in their ability to accurately characterise emotional lability and its progression in clinical populations, what is needed is a physiological test that allows us to *quantify* the emotional response to determine measurable diagnostically significant boundaries.

3.1.5 Facial contagion as a physiological test for UMN involvement in MND

I propose to lay the foundations for developing a simple, non-invasive neurophysiological test that can detect the presence of corticobulbar dysfunction as a marker of UMN involvement in suspected MND. The test exploits the dual innervation of the corticobulbar nuclei that control the facial musculature.

A simple physiological test can be derived from facial contagion: the propensity for a person's facial expression to automatically begin assuming that of any face they are viewing. Since it is established that this empathic facial response is not under voluntary control but rather occurs even when the face being viewed is made subliminal by being presented too briefly to register consciously (Dimberg et al 2000; Neumann et al 2014), the response is usually too weak to be externally perceived but the characteristic pattern of muscular contraction can be detected using surface EMG electrodes. The reviewed literature suggests that people naturally respond with a characteristic latency and amplitude that can be estimated reliably with a relatively small number of presentations. This is a clear advantage for creating a diagnostic that is effective and accurate yet minimally invasive, easily administered, and brief. From the perspective of pure scientific enquiry, this will be the first physical characterisation of emotional lability in MND patients.

3.1.6 Scope

In the pathological state presently explored, two predictions can be made to differentiate between UMN and LMN involvement. Since the empathic response is automatic and unconscious, it is plausibly mediated by the subcortical mechanisms that pathologically dominate due to cortical degeneration; it is therefore expected to be abnormally *enhanced* in patients with MND who have UMN involvement relative to those who do not.

Additionally, it is expected that switching from an emotional response of one valence (e.g. happy) to another (e.g. angry) would be faster in patients with high lability compared to those with low lability. This second prediction is due to the dysregulation of the cortical mechanisms that maintain the natural emotional inertia found in patients with suspected UMN in MND (Ironside 1956; Pichon et al 2014), and will be referred to as *emotional momentum*. In this context, the strength of the effect of continuity of the emotional association with the perceived face is measured exploring the effect of preceding trial on ensuing response. If these predictions are correct, they will form the foundations of a quantitative test of emotional *lability* and *sentimentality* respectively that ought to be sensitive to the presence of UMN involvement in MND.

Lastly, the Pompeii Adaptive Filtering (PAF) technique developed in the previous chapter will be tested for the first time on biological signals to uncover the facial contagion response. This will provide an external evaluation of the PAF by reproducing the same comparative analysis with the conventional method on original data collected from a human population. All analyses will be further assessed with machine learning techniques.

3.2 Methods

3.2.1 Design

To explore whether emotional lability could serve as a marker diagnostic for UMN in suspected MND, a noninvasive, cross-sectional study of two groups (patients with suspected UMN involvement in the context of MND and patients without signs of upper motor neuron involvement in the context of suspected MND) was performed in whom the facial contagion response to viewing emotional faces was quantified with surface EMG. This experiment was approved by the NHS Research Ethics Committee under reference 13/LO/1693 in November 2013 and granted an extension of one year by the same body in January 2015.

The following factors were manipulated within participants and comparisons were made at the group level: 1) emotional expression of the viewed face (happy or angry) and 2) emotional congruity with preceding image (congruent or incongruent). The first of these factors was required to elicit a contrast of emotional response on which any difference between the groups could depend. The second factor provided a measure of *emotional momentum* or an index of compatibility of emotions of opposite valence: a feature of emotional lability as assessed clinically. Both factors were therefore expected to increase the sensitivity and specificity of any difference between the groups.

The hypothesis was that the response of patients with suspected UMN involvement and higher ELQ scores would be more pronounced (shorter latency, greater amplitude, and longer duration) and that the effect of emotional momentum in this group would be greater.

3.2.2 Participants

Patients with an established diagnosis of MND were recruited opportunistically from the MND clinic at the National Hospital for Neurology and Neurosurgery in London. Exclusion criteria consisted of a) evidence of cognitive dysfunction sufficient to make either understanding of the task or consent unreliable, b) visual dysfunction sufficient to make perception of large stimuli unreliable, c) significant co-existent cortical pathology

of another cause, d) significant LMN facial weakness, as clinically determined. This last criterion was necessary since the test relied on revealing differences in the input to the brainstem nuclei of the facial nerve. If the facial nerve itself was affected by degeneration, any difference was likely to be attenuated and it was therefore sensible to exclude patients so affected.

3.2.3 Materials

3.2.3.1 Electrophysiological recordings

Facial EMG activity was acquired from ten 24 mm Kendall ARBO disposable electrodes corresponding to five distinct bipolar montages. The electrodes were placed over zygomaticus major and corrugator supercilii muscle regions on both sides of the face with an additional ground montage on the forehead as common reference (Fridlund and Cacioppo 1986). Impedance for all electrodes was kept below 10 kΩ. The raw EMG signals were acquired through a System Plus Evolution EEG system (Micromed SpA, via Giotto, 2 – 31021, Mogliano Veneto, Italy). The raw EMG signals were continuously recorded with a sampling rate of 4096Hz and a bandwidth of 0.5 – 1000 Hz, 16-bit sampling resolution. No notch filter was used.

3.2.3.2 Stimuli

The stimuli were created from validated emotional pictures from the Karolinska Directed Emotional Faces (KDEF) dataset (Lundqvist et al 1998). Front-facing images of men and women exhibiting happy and angry facial expressions were selected. In MATLAB, the images were converted to black and white, aligned, framed, and cropped so that only the faces were visible. This process yielded a total of 100 face stimuli consisting of 50 angry and 50 happy faces of men and women (see Figure 3.2).

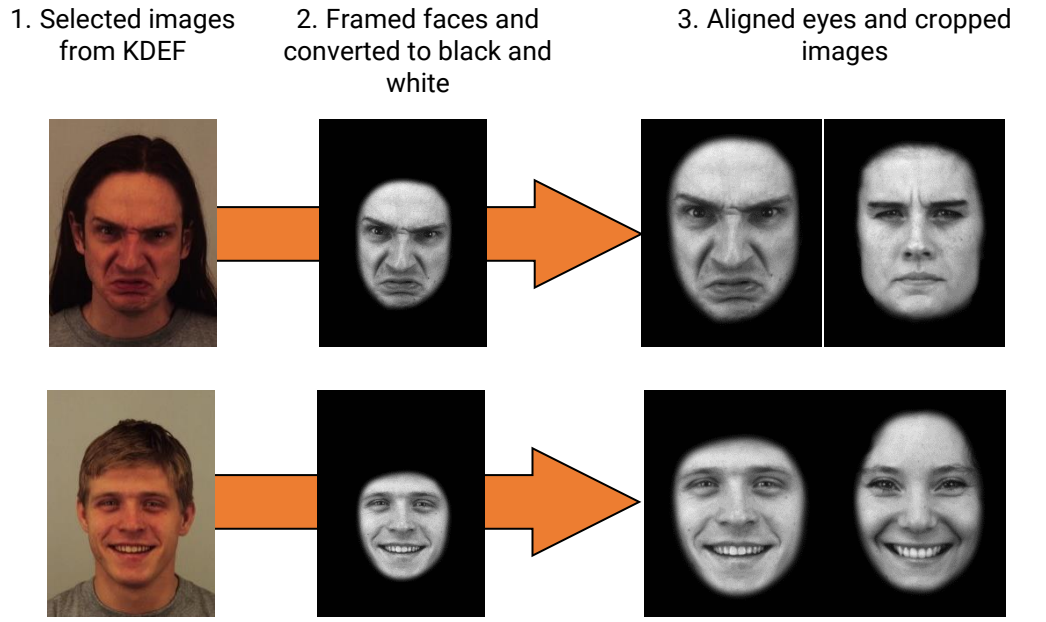


Figure 3.2. Stimuli creation process. 1) Front facing images were selected from the KDEF database, 2) converted into black and white and framed. 3) The images were then aligned by measuring intra-ocular distance and cropped.

3.2.4 Procedure

Prior to inclusion in the study, patients were approached by their practicing clinician during a routine visit at the MND Clinic who explained the study and offered the possibility of volunteering in the study by supplying a Participant Information Sheet (PIS) depending on their potential group allocation (see Appendix A.1). All patients were allowed at least 24 hours to decide and ask further questions prior to inclusion in the study. Recruitment of patients was performed by the clinical team directly involved in patient care.

On the day of testing at the NHNN all participants received the PIS again from a member of the research team and gave written informed consent. Prior to beginning the facial EMG measurements, participants completed the ELQ (Newsom-Davis et al 1999) to quantify the potential degree of exaggerated laughing, crying, or smiling (potential scores on a scale from 0 to 93). This allowed the subdivision of patients into two experimental groups: 1) the high lability group, scoring 8.5 or over on the ELQ and 2) the low lability group, scoring under 8.5 on the ELQ. This cut-off was selected based on Newsom-Davis et al (1999) original consideration that scores 2 SDs above the control

group's mean indicated high ELQ (i.e. ELQ = 8.5+). Since the contrast of interest was confined to within the MND population and specific to the question of lability within it, no normative dataset was necessary as the purpose was to establish whether this measure was indeed useful at differentiating sub-groups rather than compare it to a healthy population.

After being fitted with EMG electrodes, participants, tested individually, sat comfortably in front of a laptop where images of different faces appeared in sequence (see Figure 3.3). They were tasked with identifying the gender of the subjects in the images through button presses to maintain attention on the stream of faces. The experiment ran on Presentation® software (Version 0.70, www.neurobs.com). Equal number of male and female faces were shown for each emotional expression (happy or angry) associated with the muscle responses of interest.

Participants completed 8 blocks of 24 trials lasting approximately 2.4 minutes each and could rest in between.

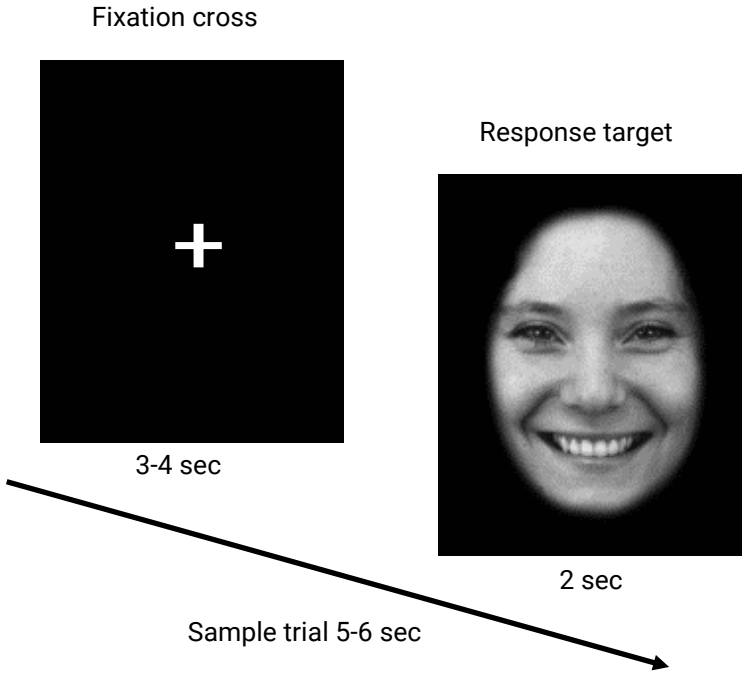


Figure 3.3 Example stimulus presentation. Participants responded to the emotional facial expression by identifying the gender of the target image after viewing a fixation cross.

3.2.5 Analysis

Analysis of button-press accuracy and reaction times were used to determine whether or not participants attended the task. The ELQ data was scored and values used to categorise a patient's data in a particular group. If a patient scored 8.5 or higher on the ELQ, they were considered as high-lability patients whereas the rest were considered to exhibit low lability and grouped accordingly. Where these preliminaries proved satisfactory, event-related EMG responses were compared by parametrising latency, amplitude, and duration across conditions and participants.

The stored EMG raw signals were imported into MATLAB and underwent conventional and PAF filtering analyses (see previous chapter for more details on these techniques) with comparisons drawn at the group level between the two groups. For the conventional analysis, all steps of EMG signal processing detailed in section 2.1.5 were performed. Only step 6 was not performed as impedance levels from muscle sites from the left and right muscles suggested these differed. Instead, left and right muscle sites were analyses separately for each muscle group and for both techniques, the ones with the clearest signal-to-noise ratio (where the facial contagion response was most evident) were selected for inclusion into the group analysis.

Facial contagion was measured as activity in the congruent facial muscles as the ones flexed by the subjects in the images following the presentation of the stimulus (Dimberg and Thunberg 2012). This analysis adhered to standard practice in the field (Weyers et al 2006; Achaibou et al 2008; Krumhuber et al 2014). The magnitude of the EMG response, or muscular strength activity, was calculated by taking the mean activity over 20, 100ms time bins, after the onset of each stimulus and adjusting for baseline at each trial by subtracting the activity 1000ms preceding stimulus presentation.

At the group level, a normalization procedure was followed to allow for comparability of signals at a similar scale in a two-step process. The first step required baselining the mean responses per patient by subtracting the mean of the first three bins (0-300ms) to each individual response across each patient and muscle group. This was followed by *z-scoring* or standardisation involving converting all values to a common scale with an average of zero and standard deviation of one. This was calculated by subtracting the quotient from the mean and the standard deviation of all values from the

actual value, as commonly performed in the field (Moody et al 2007; Hofree et al 2014; Harrison et al 2010).

For conventional filtering methods, two sample t-tests were performed comparing the differential muscle response to happy versus angry faces in high and low lability patient groups, providing a measure of facial contagion as a function of emotional lability. For time series resulting from PAF filtering, 2x2 ANOVAs compared the means between the activity in response to happy and angry faces at each muscle site, looking for the main effect of facial contagion related to expression emotion. All statistical results were Bonferroni corrected for multiple comparisons.

A secondary analysis termed *differential* was performed to compare high and low lability patient responses for each muscle group by subtracting the mean response of the non-congruent emotion to that muscle group (i.e. zygomaticus = happy – angry; corrugator = angry – happy). These differential time series were also fed into an SVM multivariate model so as to assess the predictability of ELQ scores from the non-averaged facial contagion responses which were used to further assess the performance of the filtering methods. Although the calculations to create these data was the same, rather than using mean signals, individual signals from each viable trial were used so as to obtain enough data to train the SVMs. Therefore, the subtraction was performed on the nearest temporal trial of opposite valence rather than on the means.

Similar statistical analyses were performed when grouping the data across temporally congruent and incongruent emotional trials to test the extent of emotional momentum. For this analysis, differential responses were used while the variable of contrast consisted of emotional valence congruency of the image in the immediately preceding trial. For example, when two trials followed each other of the same emotion ([a]smiling/[b]frowning face followed by another [a]smiling/[b]frowning face respectively) it would be a congruent trial while if these were of different emotional valence ([b]frowning/[a]smiling face followed by a [a]smiling/[b]frowning face) it would be an incongruent trial. This was performed so as to assess the potential for the facial contagion response to act as a marker of sentimentality.

3.3 Results

3.3.1 Participants

Although close to 100 patients were assessed for eligibility, 25 MND patients participated in this experiment; however, 3 were excluded due to machine malfunction. The effect size in this group is unknown as previous studies have demonstrated an effect with sample sizes from 13 (Oberman et al., 2009) to several hundreds (e.g. Künecke et al., 2014).

A score of 10 or higher on the ELQ denoted high lability which lead to the sub-division of patients into high lability (*n*=10) and low lability (*n*=12). The low lability group had an average ELQ score of 2.25, ranged in age between 44 and 76 (*M*=59) and had 4 females (8 men). The high lability group had an average ELQ score of 23.9, ranged in age between 46 and 78 (*M*=58) and had 3 females (7 men). The ranges of ELQ scores are presented in the histogram in Figure 3.3.1 with the red dotted line denoting the cut-off between low and high ELQ scoring patients. None of the patients were excluded as no outliers were identified in the data (upper bound = 47.5; lower bound = -28.5).

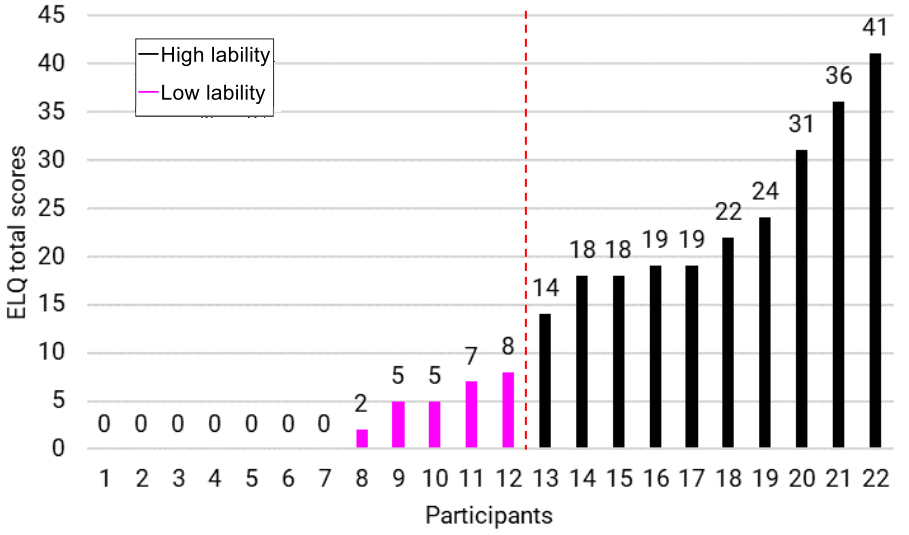


Figure. 3.3.1 Histogram of ELQ scores per patient. The red dotted line denotes the cut-off between high and low lability patients. The scores have been arranged in ascending order and do not reflect acquired order or participation.

3.3.2 **Behavioural Responses**

Overall participants correctly identified the gender of the images in 98% of trials. A two-sample t-test comparing differences in response times between patient groups revealed no significant differences between the high (*M*=1010ms; SD=316) and low lability (M=987ms; SD=317) patients ($t(2706)=1.95$, $p>0.05$).

3.3.3 **Facial contagion**

Conventional and PAF filtering were compared in their performance at elucidating a facial contagion response from unilateral zygomaticus and corrugator activity in response to angry and happy faces at each timepoint (Figure 3.4). Conventional methods revealed a stronger effect of contagion in the corrugator for high lability patients only in response to angry faces. ANOVAs on PAF-filtered time series showed a strong facial contagion response in the zygomaticus for happy faces in low lability patients. PAF filtering suggested that high lability patients demonstrated a weak but consistent stronger facial contagion response to angry faces irrespective of muscle site. Results of the differential analysis echoed these findings (Figure 3.5). No effect of CRF was found in any of the PAF analysis.

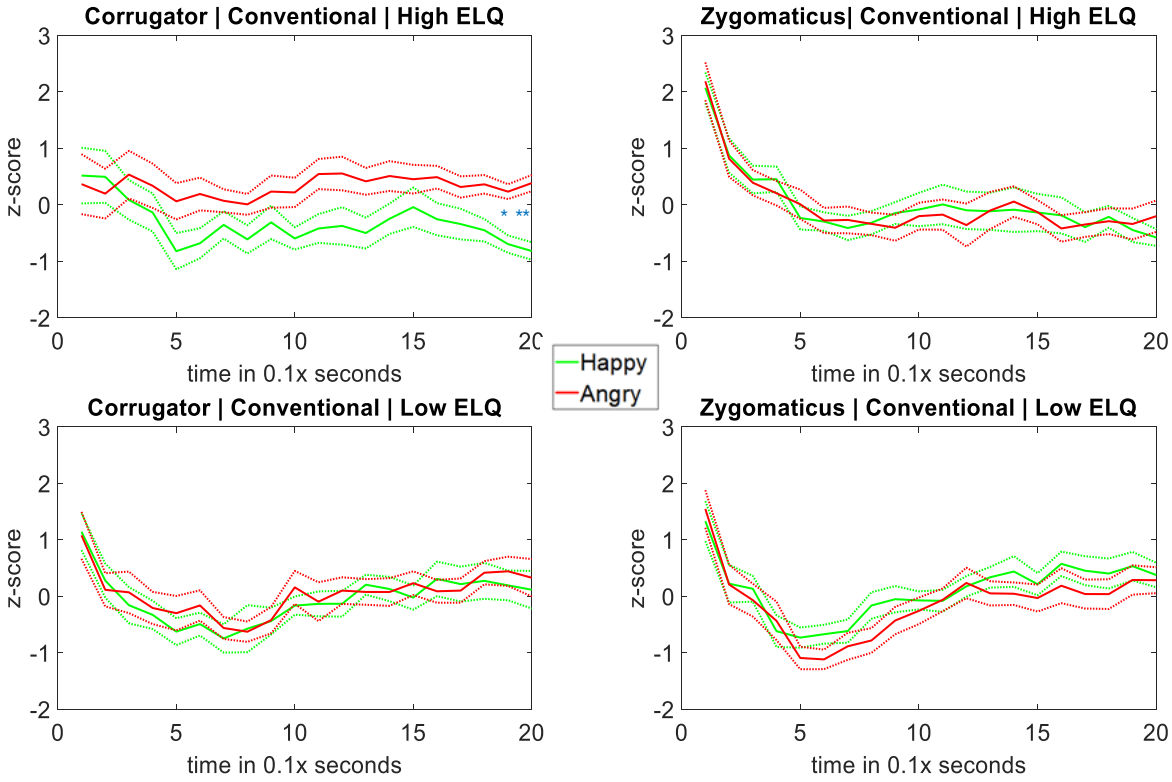
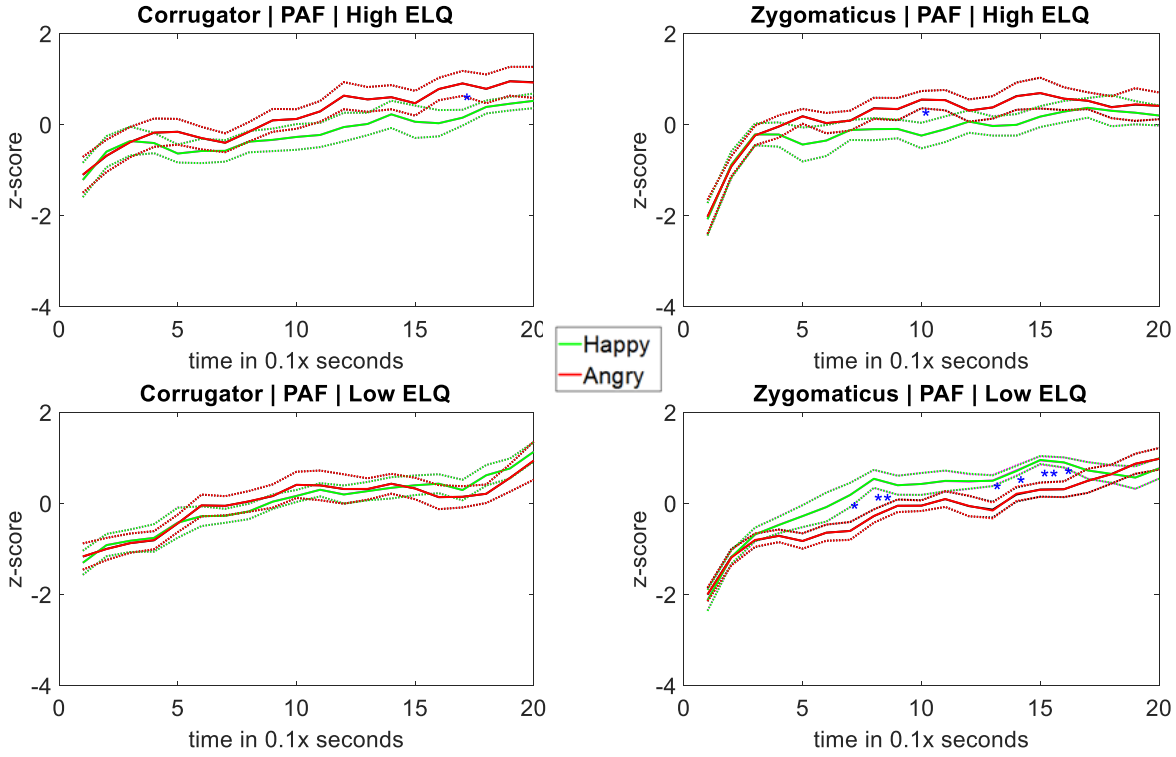


Figure. 3.4 Facial contagion responses with conventional and PAF filtering. The top quadrant is the result of conventional filtering while the bottom four is the result of PAF filtering. For each quadrant, the left columns represent corrugator activity while the right column shows zygomaticus responses. The top row are the aggregate responses of the high lability patients and the low lability patients are in the bottom row while the dotted lines represent standard error. The x-axis shows the change of the signal over time and the y-axis shows the normalized activity as z-scores. One blue asterisk (*) indicates statistical significance at p<0.05 and two blue asterisks (**) at p<0.001 Bonferroni corrected. For PAF filtering, there are 4 plotted lines though only 2 are visible since the other account for the null effect of CRF, hence the slight discoloration.



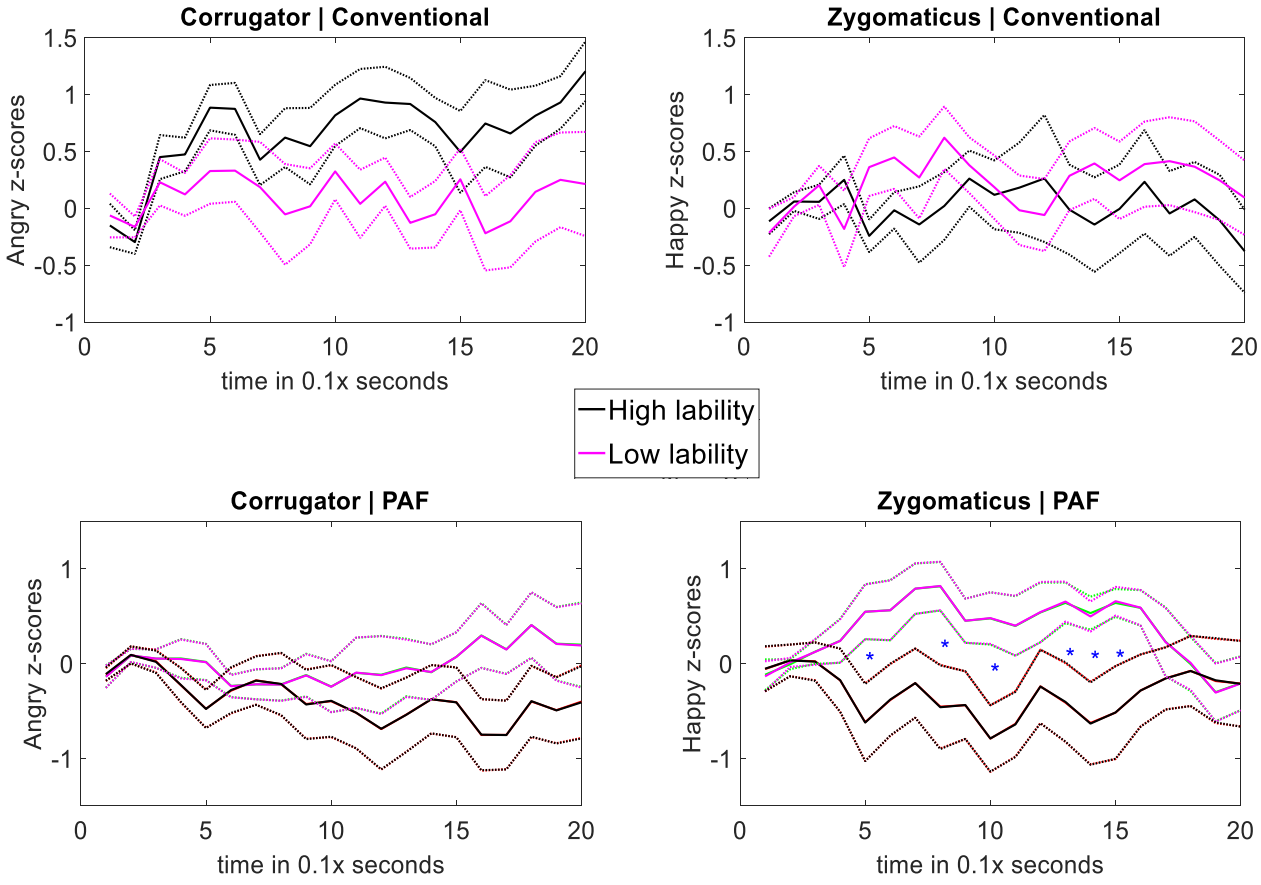


Figure. 3.5 Comparing high and low lability facial contagion responses with conventional and PAF filtering. The top row are the results of the conventional filtering while the bottom are those issuing from the PAF. The aggregate responses of the high lability patients and the low lability patients are shown with the standard error in dotted lines. The left columns represent corrugator activity while the right column shows zygomaticus responses. The x-axis shows the change of the signal over time and the y-axis shows the normalized activity as z-scores. One blue asterisk (*) indicates statistical significance at $p<0.05$ and two blue asterisks (**) at $p<0.001$ Bonferroni corrected.

Further testing of whether the facial contagion response could be better differentiated with the PAF technique was provided by the results of the SVM analysis. The SVM sought to predict whether a particular un-averaged but normalized facial contagion response was from a low or high lability patient. The results ensuing from the binary classification task can be presented by quantifying the performance of the system a correctly classifying a test point into one of the following 4 categories: True Positive (TP) and True Negative (TN) if the system correctly predicts the label or False Positive (FP) if the systems labels it as a positive when it is a negative and False Negative (FN) if labelled as a negative while it is a positive (Table 3.1). These values then allow for the calculation of sensitivity (true positives / positives) and specificity (true negatives / negatives) and may also be used to generate more conservative balanced accuracies ((sensitivity + specificity)/2; for a discussion on the superiority of using balanced accuracies rather than the optimistic *accuracy* measure see Brodersen et al (2010)).

Table 3.1 Confusion Matrices resulting from SVM predictions from Conventional and PAF Filtering

Conventional filtering					
Corrugator			Zygomaticus		
N*= 367	Predicted: No	Predicted: Yes	N = 372	Predicted: No	Predicted: Yes
Actual: No	93	91	Actual: No	91	95
Actual: Yes	87	96	Actual: Yes	86	100
PAF filtering					
Corrugator			Zygomaticus		
N = 366	Predicted: No	Predicted: Yes	N = 376	Predicted: No	Predicted: Yes
Actual: No	90	93	Actual: No	102	86
Actual: Yes	89	94	Actual: Yes	102	86

*N is the total number of predictions made. The rest are outputs of the SVMs resulting from prediction attempts of classifying patient facial contagion EMG responses into high and low lability.

Based on facial contagion responses from the corrugator muscle, the PAF performed at a mean balanced accuracy of 50.4% with 95% confidence intervals (CI) of 49.75 to 51.05 compared to conventional filtering that performed at 51.5% (95% CI 50.83 to 52.17). For the Zygomaticus, the PAF performed at 52.7% (95% CI 52.06 to 53.36) and conventional filtering at 50.3% (95% CI 49.62 to 50.98). Therefore, the PAF filtering performed 2.35% better than conventional methods at correctly classifying high and low lability patients based on their zygomaticus facial contagion response ($t(98)=4.89$, $p<0.0001$) and 1.17% worse if based on corrugator responses ($t(98)=2.45$, $p<0.05$; Figure 3.6). Although the SVM from PAF activity correctly classified patients as high or low lability above chance for the zygomaticus, it did not do so significantly for the corrugator. In contrast, the SVM from the conventional activity predicted lability in patients above chance for the corrugator but not for the zygomaticus.

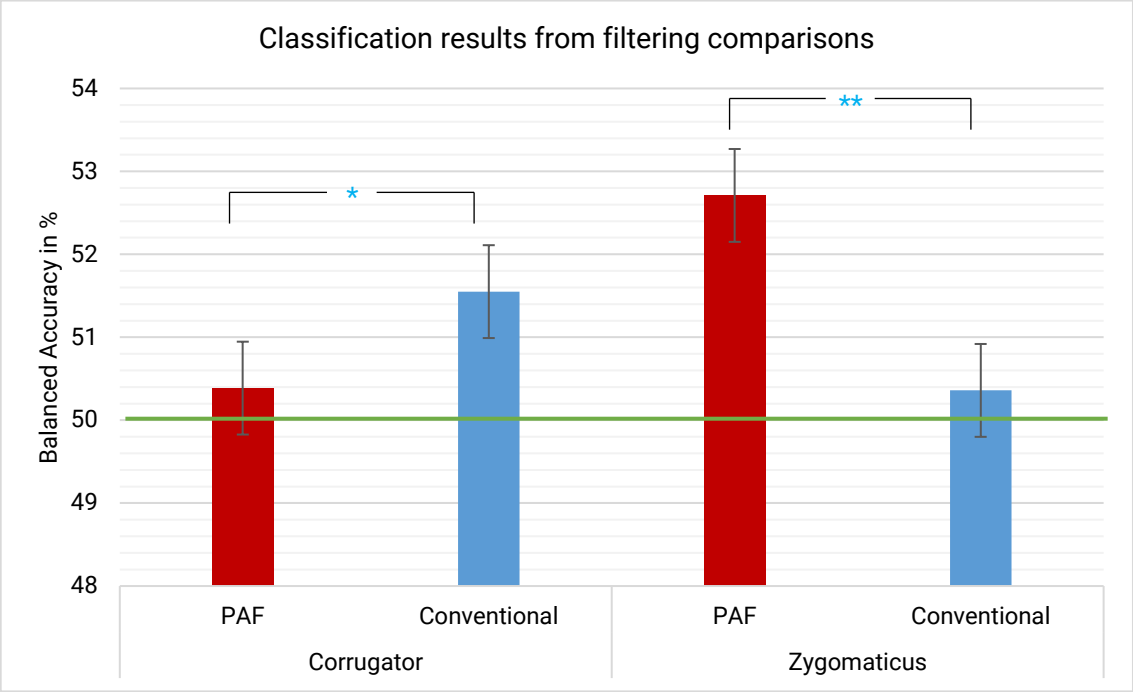


Figure. 3.6 Mean balanced accuracies from SVMs. This shows the success of predicting how each filtering method classifies patients into high and low lability groups based on facial contagion responses. Anything above the green line is above chance performance. One blue asterisk (*) indicates statistical significance at p<0.05 and two blue asterisks (**) at p<0.001.

3.3.4 Emotional Momentum

None of the conventional and PAF filtering EMG responses to congruent or incongruent trials reached significance. A weak trend ($p<0.05$ uncorrected from 1100ms to 1300ms) was found with conventional filtering where high lability patients demonstrated stronger corrugator facial contagion responses to congruent trials (*angry* followed by *angry*). A similar weak trend was found with zygomaticus facial contagion responses in low lability patients for congruent trials (*happy* followed by *happy)* in conventional ($p<0.05$ uncorrected from 1700ms to 1800ms) and in high lability patients for PAF filtering ($p<0.05$ uncorrected from 1600ms to 1700ms; see Figure 3.7 for time series).

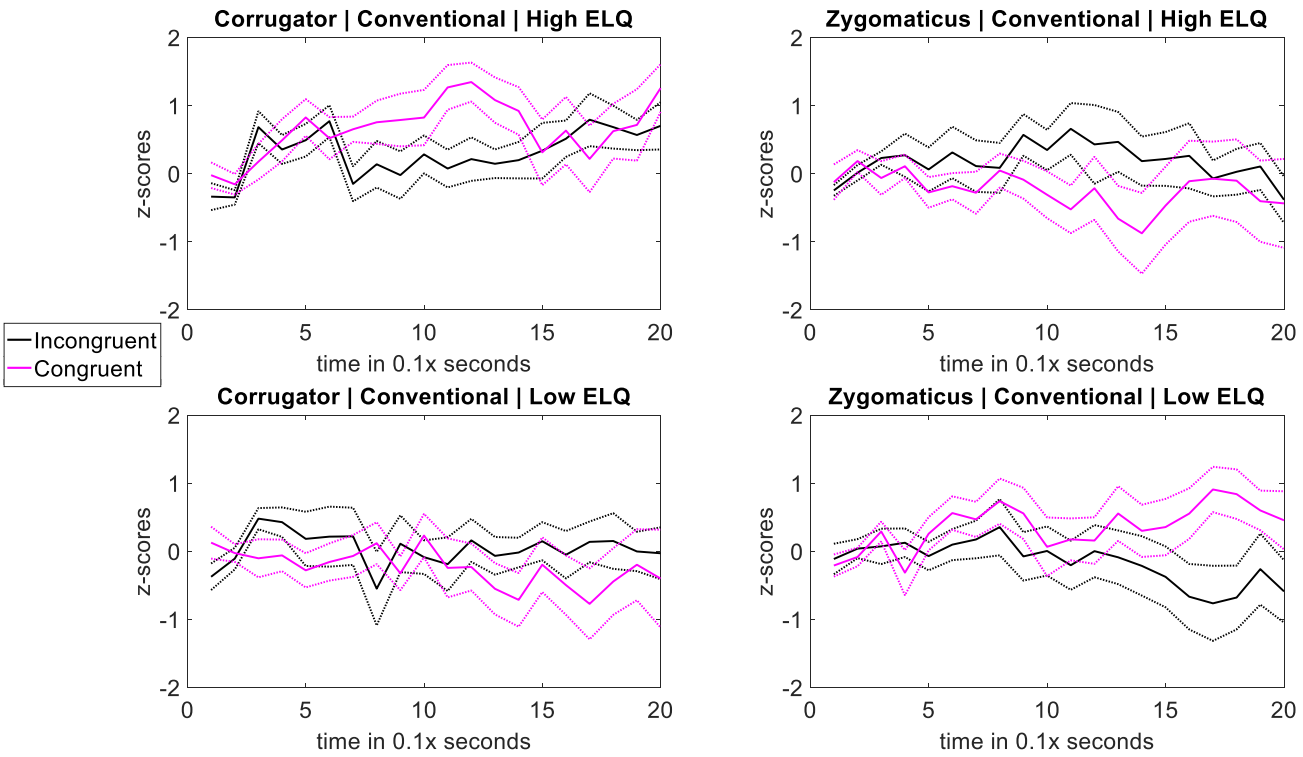
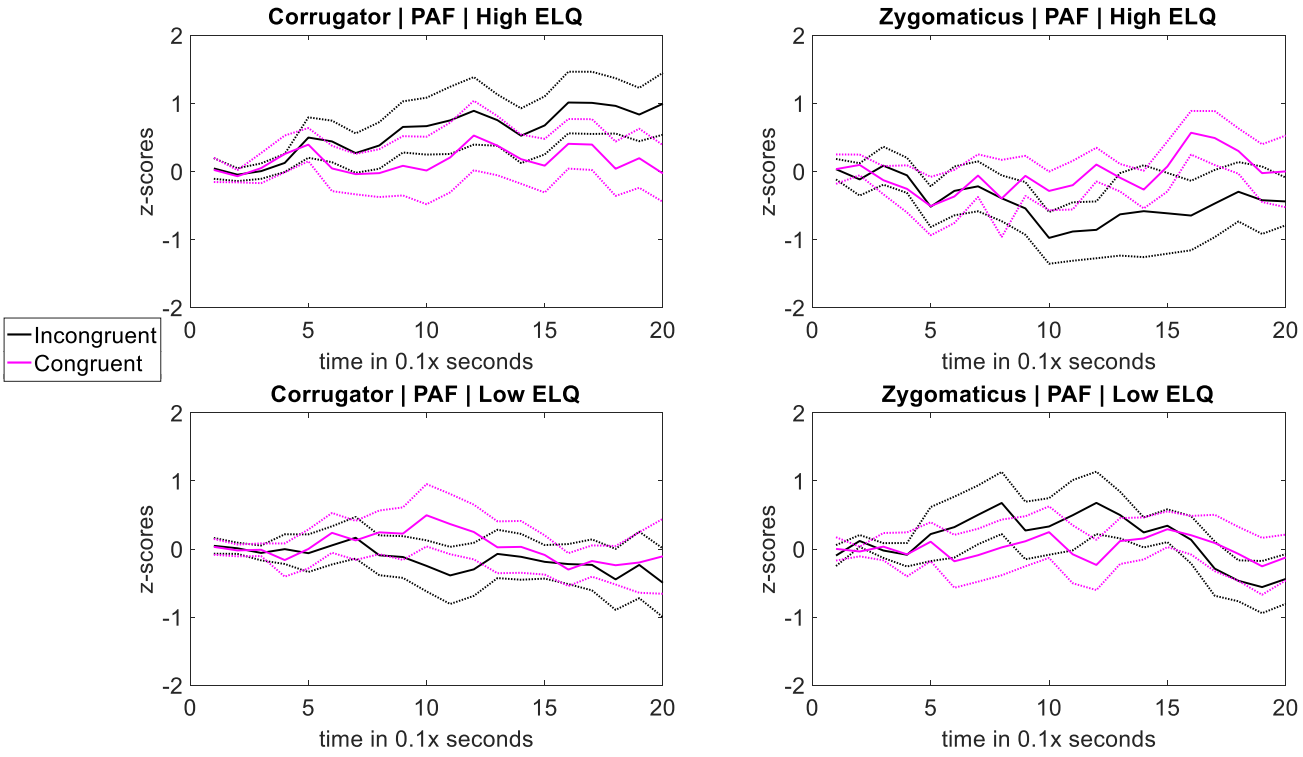


Figure. 3.7 Emotional momentum differential facial contagion responses. The top quadrant is the result of conventional filtering while the bottom four is the result of PAF filtering. For each quadrant, the left columns represent corrugator activity while the right column shows zygomaticus responses. The top row are the aggregate responses of the high lability patients and the low lability patients are in the bottom row while the dotted lines represent standard error. The x-axis shows the change of the signal over time and the y-axis shows the normalized activity as z-scores. None reached statistical significance.



3.4 Discussion

The goal of this study was to assess whether facial contagion responses measured with non-invasive EMG could be used as a diagnostic marker for upper involvement in MND subtypes. The congruent automatic face muscle response was measured from MND patients while they looked at images of smiling and frowning individuals. Patients also reported their level of emotional lability through a self-report questionnaire which allowed the division of the patients into two subgroups for statistical comparisons. I predicted that patients with upper involvement in MND would be more likely to exhibit an exaggerated facial contagion response and emotional momentum response. I also predicted these results would be more easily discerned through the PAF filtering method developed in the previous chapter. This methodological aspect will first be addressed so as to justify interpretation of results with either method.

3.4.1 Comparison of filtering methods

This was the first time the novel filtering method developed in the last chapter was tested on biological data and contrasted with conventional filtering. For facial contagion responses, although conventional filtering worked, the PAF provided more differentiation in terms of statistical power. As with the previous chapter, SVMs were used to further validate the direction of differences between the two methods to assess which filtered signals could best predict high or low lability. The data showed the PAF was better at differentiating the facial contagion response in the zygomaticus but worse in the corrugator. Although the difference in performance between the PAF and conventional filtering reached statistical differentiation, these were too close to chance so neither demonstrated accuracies to denote high confidence or clinical significance.

Since no effect of CRF was found for any condition, echoing the results of testing with artificial modelled data, the step of counterbalancing the cross-correlation of the Meta-CRF with both valences rather than only the congruent one can be skipped. For example, it should be valid to cross-correlate happy facial trials with the Happy Meta-CRF only and angry trials with the Angry Meta-CRF for that muscle site only. This will also allow for faster and more direct comparison of results as the same statistical test will then be run on both conventional and PAF time series, namely a two-sample t-test.

3.4.2 Facial contagion as a marker for upper involvement in MND

This study reports original data measuring the facial contagion response in MND patients with EMG. Consistent with the reviewed literature where the facial contagion response has been previously measured only in healthy controls, MND patients in both groups demonstrated mostly congruent contagion responses. This study was the first to measure the facial contagion response on a sample that is affected by a neurodegenerative disorder affecting the motor muscles.

The shape of the facial contagion response in the current study was consistent with previous studies that did not show a very strong separation of the signals that remained statistically significant (Achaibou et al 2008; Weyers et al 2009; Aguado et al 2013; Heerey and Crossley 2013). Although the corrugator response can be varied between individuals there was no evidence of the so-called 'startle response' at the start of the facial contagion signal, in contrast to past research (Dimberg and Thunberg 1998, 2007; Dimberg et al 2000; Hofree et al 2014). While low lability patients demonstrated a larger facial contagion response in the post 500ms time period, high lability patients demonstrate a delayed response post 1000ms.

The response in the high lability group was also weaker in general as evidenced by the differential comparisons, although significant for both muscle sites. In contrast, the low lability group had strong responses - comparable to healthy responses reported in the literature - in the zygomaticus but this was very weak in the corrugator. It is possible that this was simply due to a difficulty in obtaining a strong signal from the corrugator since it is a considerably smaller muscle.

The corrugator reflected higher contagion to angry than happy faces and the zygomaticus reflected higher contagion to happy than angry faces. There was an exception to this congruent response with the high lability group which demonstrated higher facial contagion responses towards angry faces irrespective of the muscle involved as evidenced by the PAF results. This pattern of exaggerated reactivity to angry frowning faces is evidence of a negativity bias in this patient sub-sample. It is possible that this is due to the generally worse prognosis associated with upper motor neuron involvement in MND which may lead to more negative reactivity. This result is also consistent with the finding that high lability in these patients is due to the subcortical predominance of motor control due to cortical degeneration (Iwata et al 2011; Floeter et al 2014). Therefore, this negative bias could be a shifting of the response towards what are subcortically more dominant emotions such as anger, which are preferentially activated pathways during reduced cortical input (i.e. reduced attention: Vuilleumier et

al (2001; 2003)) or biased subcortical priming (i.e. priming or masking: Morris et al (1998; 2004); Nomura et al 2004).

3.4.3 Emotional Momentum

There is a vast array of studies on how humans react to different emotions and the characteristic intensity of the contagion response for each over time. However, little is known about changing from one emotion to another and how long their effect may last over time. For the first time, this study reported the facial contagion response as a function of congruency between a trial and the one preceding it in a measure called emotional momentum. The analysis revealed no significant differences in congruent and incongruent trials between the groups for either filtering method. It is possible that the lack of effect is due to the small sample size as weak trends were found but did not reach statistical significance, therefore no strong conclusions can be made from the current data about this component of the study.

3.4.4 Limitations

All of these results presented here suffer from a common pitfall which is the relatively small sample size and slight imbalance in the groups *n* with two more patients in the low lability group. All studies finding strong facial contagion responses recruit around 50 healthy participants in their studies (see Table 2.2.1 in previous chapter for full list of studies). The number of recruited patients in each group was expected to be much higher (n=24) however due to time constraints and unexpected difficulties in the recruitment process, under half the number of participants required to elucidate a proper response were obtained.

This is also the first study to perform these measurements on a sample affected by a neuromuscular disorder. Of all tested patients, less than 30% did not report any dysregulation in emotion as measured with the ELQ. It is therefore possible that in this sample the facial contagion response as measured with EMG was more attenuated than in normal controls and would not be distinguished without a larger sample size. Perhaps it could even signify that further segmentation of patients between high, low, and medium ELQ scores could be informational.

It is both possible that should recruitment continue, trends become more accentuated, however these differences could also disappear. Therefore, it is advisable to treat these results with caution and as the basis for future exploration. Importantly,

none of the limitations in our experiment could be explained by demographic differences in the groups studied as they had an equal range of ages and genders and responded to the gender identification with similar reaction times.

Since the onset of the facial contagion response was found to be slightly delayed, it could be that longer ITIs would have permitted a longer timeframe to be studied to individuate a facial contagion response. Although some studies adopt ITIs longer than 20 seconds (Dimberg and Thunberg 1998, 2007, 2012; Dimberg et al 2000; Weyers et al 2006, 2009; Rymarczyk et al 2011), this was opted against as the aim was to maintain the patients engaged in the task and create a short diagnostic test. During testing, the only patient that demonstrated overt lability in response to the emotional faces had a reaction that lasted several seconds into the next trial. It is possible that longer covert reactions were present in other patients and were therefore missed with the current setup.

There were several issues that arose from using the ELQ as a main measure of lability. The cut-off performed at a score of 8.5 to separate high from low lability patients was taken from the pioneering study. If a similar cut-off method would have been applied based on the low lability patients as the "control" sample, it would have resulted in 2 more patients being considered high lability and contributed to further inequality between the groups (as the cut-off with the low lability as control group would have been closer to 4.5 than 8.5). Nevertheless, it is possible that such a comparison could be performed with a more complete recruited participant population. Strangely

Although the ELQ is the most valid measure to the author's knowledge to measure lability in this patient sample, it contains several problems. The ELQ measures changes "over the past 4 weeks" and some patients who have become labile as a result of their condition from before 4 weeks are not classified as labile. Recall that the ELQ is composed of three main questions (i.e. "Have you experienced sudden episodes of laughter/crying/smiling in the past four weeks?" YES/NO) which define whether the patient should answer the questions within the sub-section. In each sub-section there is a question that reads "Does this pattern of laughing/crying/smiling represent a change from that prior to the onset of MND?". This is problematic as the ELQ is meant to diagnose *recent* presence of lability rather than overall state which is what this study was interested in assessing.

The wording of the ELQ is also not precise enough as the two main questions about laughing and crying ask about "sudden" episodes while for the "smiling" section it replaces "sudden" with "unusual". The use of the word "sudden" created outspoken

confusion with patients and the experimenters more than once during the process of the experiment. A better way of assessing lability as a result of the condition should have been to replace the part that reads “in the past 4 weeks” with the wording of the sub-question “change from that prior to the onset of MND?” and ensure all questions contained the same wording with the word “unusual” (i.e. “Have you experienced unusual episodes of laughter/crying/smiling since the onset of MND?” YES/NO). This would have perhaps avoided episodes where a patient referred to as highly labile by the clinicians did not score above 8.5 because they had no “sudden” lability episodes in the past 4 weeks but might have had some “unusual” ones instead since they contracted the disorder.

3.4.5 Future Directions

A possibility to generate a stronger facial contagion response with a small sample size would be to use dynamic facial expressions as stimuli. These would consist of faces morphing from neutral to a particular emotion over several seconds. There are several studies showing the superiority of these stimuli in eliciting a response and it could provide a comparable response with less participants (Hermans et al 2006; Weyers et al 2006; Achaibou et al 2008; Rymarczyk et al 2011; Matzke et al 2013; Ardizzi et al 2014; Künecke et al 2014). This was not implemented within the current study as it was expected that recruitment would be faster and the facial contagion response would be exaggerated (rather than, seemingly, attenuated).

It is possible that since the Meta-CRF was created from healthy individuals, it may not have been applicable to this particular group of patients. Therefore, an Individual Response Function (IRF) could be created for each patient to be cross-correlated to their own data. In a similar methodological study, Bach et al (2013) showed the superiority of modelling an IRF over a CRF to improve predictive validity of skin conductance response. This is therefore a viable possibility for future research and echoes previous findings in other fields that increase specificity of the CRFs to obtain better results when comparing groups with model based filtering methods (e.g. Arichi et al 2012; Steffener et al 2010). To properly validate the filtering methods on biological signals the next step would be to acquire a dataset of normal facial contagion responses and assess whether the PAF method is in fact superior. Although it was interesting to explore the potential of the method here, it was tested on a sample whose facial contagion response could be severely affected by its disorder, so it may be hard to draw strong conclusions prior to testing it on a healthy sample.

In order to test the negativity bias hypothesis about the high lability MND patients, a new analysis could be performed whereby the positive and negative ELQ measurements would be independently scored and used as grouping variables. This would lead to two sets of comparisons between non-labile and positively labile patients and another between non-labile and negatively labile patients. This separation of groups would perhaps help distinguish an effect of specific emotional valence within the high lability group.

Finally, a better grouping of patients based on the proposed criticisms of the ELQ could be devised to accurately categorise patients into MND subgroups for further testing. If found to be effective there is potential to expand the validity of this measure as a test for other patient populations. Patients affected with Multiple Sclerosis are known to also suffer from dysregulated emotional lability (Smith et al 2004). Further elucidating of the exact consequences of lability through the facial contagion response could perhaps inform prognosis or draw further differentiations between patients when self-report is not sensitive enough.

3.4.6 Conclusions

Although the effects found were statistically weak, three main differences were found between high and low lability facial contagion responses which could form the basis for an objective biological marker that differentiates between MND subtypes. This study showed that high lability MND patients demonstrate a 1) delayed onset of the facial contagion response, 2) a weaker overall signal, and 3) a higher reactivity to angry valence stimuli. If it is found that patients with upper involvement in MND consistently express a weaker delayed facial contagion response and that the negativity bias extends to other negative valence stimuli (e.g. disgust, sadness, fear), it could be translated into a marker of clinical utility in informing the diagnostic.

However, due to the discussed limitations the present results can only be said to form the basis for future exploration into the characteristics of the response as a valid marker. The issues discussed exemplify some of the reasons why it is difficult to develop such a diagnostic. This data was the first ever measure of an automatic emotional reaction in the MND population and holds promise as a biomarker for upper involvement in MND.

Chapter 4: Improving affective states through facial contagion

From Mancini et al (2013b).

4.1 Introduction

4.1.1 Depression and the problem of its treatment

According to the World Health Organisation (WHO) depression is “a common mental disorder, characterized by sadness, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, feelings of tiredness, and poor concentration.” (World Health Organisation 2016). This serious and often recurring disorder will affect 25% of the population at least once by the end of adolescence (Kessler et al 2001) and is twice as prevalent in females as in males (Angst et al 2002). It is generally associated with diminished functioning and quality of life, medical morbidity, and mortality (Hsu et al 2015) and the risk of a recurrence can be as high as 50–90% (American Psychiatric Association 2000). Depression is not simply a burden on the afflicted individual but also for employers and society who are burdened by their higher absences and less efficient work and their need of more than twice as much health benefits than their colleagues (Birnbaum et al 2010; Lerner et al 2010; Jain et al 2013; Tomonaga et al 2013). Indeed, depression is projected to be the leading cause of global burden of disease in high- and middle-income countries in 2030 (WHO 2012).

The effectiveness of available treatments for patients with major depression in the long term is generally weak (Robinson et al 1990; Bower et al 2001; Pampallona et al 2004; Hind et al 2014; van Straten et al 2015). Even the most widely used pharmacological treatments are not considered to be significantly beneficial (Kirsch et al 2008) and are sometimes associated with side effects such as weight gain (Vanina et al 2002), lower IQ for children of pregnant patients (Nulman et al 2012), and increased risk of suicidal thoughts (Nischal et al 2012). It is possible that these issues account for the under-usage of available treatments, as only 50% of afflicted individuals are estimated to seek help (Angst et al 2002). Nevertheless, barriers to accessing and initiating treatment may also include geography, socioeconomic status, national healthcare system capacity and waiting times, economic costs, cultural beliefs, lack of knowledge about mental health disorders, stigma, and adherence to treatments (Rost et al 1998; Wahl 1999; Nutting et al 2002; Cooper et al 2003; Simon et al 2004; Neighbors et al 2008; Thomas et al 2009; Mojtabai et al 2011).

4.1.2 Digitalising therapies

The undesired consequences and ineffectiveness of these therapies have led to searching for nonpharmacological-based treatments that could function alone or in addition to a drug regimen. Interventions administered through the internet or mobile phones hold great potential to address many of the issues and close the treatment gap for mild depression as well as treat individuals in remote or poor populations where drug treatments are unavailable or unaffordable.

Mood defines dispositions towards one kind of emotion more than another. Someone in an irritable mood is predisposed to outbursts of irritation, someone in a depressed mood to periods of sadness, and so on. Mood can be induced independently of any factual reality in contrast to our rational thinking processes. This is obvious from the many forms of theatrical and cinematic entertainment which induce a particular emotion and the general mood in the audience. While the change in mood may last for several minutes or hours after the event ended, the viewers never rationally believe that it is more than special effects and actors. This is because mood is largely set by emotional influences: exposure to a particular emotion, induced in whatever way, which can change mood, at least for a while. This can be observed in daily interactions where fictional jokes can induce transient feelings of joy. The effectiveness of interventions that capitalise on this mechanism to combat depression is evidenced by art- and drama-based therapies (Jeong et al 2005; Slayton et al 2010; Blomdahl et al 2013). Evidently these interventions may be more relevant when dealing with 'reactive' rather than 'endogenous' depression, the latter being a long lasting deep seated one, that may have a different biological basis. By combining these established aspects of how mood works, one could theoretically change it by inducing emotions without changing any factual aspect of people's lives (e.g. raising their salary). Thus, delivering therapy via mobile devices, in an artificial way, need not make it less potent and is based on a sound theoretical foundation.

Indeed, computer-based psychological treatments for depression are becoming more common due to their evidenced effectiveness (Richards and Richardson 2012). Mental health interventions that are generally less resource intensive have been found to be similarly effective as more intense face-to-face treatments (Bower et al 2001). If effective, app-based therapies could be more widely prescribed to reduce load on healthcare systems and hold potential for remote interventions and especially in hard-to-reach populations. There is evidence suggesting that people's general attitudes towards mobile self-help treatments for health problems are positive (Katz and Rice

2009). Other advantages common to mobile interventions may include free and immediate access to information, assistance and treatment, seemless assimilation with patients' everyday lives, reduced costs, and having a direct access to communicate with experts. Smartphones are generally close to their owners throughout the day and provides continuous access like no other device except perhaps wearable technology which may include specialised sensors. Also since people use smartphones to interact with many aspects of their lives (e.g. calendar, events, social media, communication) it allows researchers to intervene within the stream of life without disturbing the user. This transparency to lifestyle is essential for compliance while delivering a relatively high dose of treatment throughout the day. For the current purposes, the high degree of customizability of mobile applications also provides an unprecedented low-cost means to ecologically monitor the temporal dynamics of mood.

Creative uses of mobile phones and sensors have successfully digitalised existing treatments for depression and low mood (such as Cognitive Behavioural Therapy (CBT)) and delivered these through apps with promising results (Burns et al 2011; Kauer et al 2012; Watts et al 2013; Ly et al 2014). Nonetheless, none of these innovations proposed a novel treatment but rather implemented existing ones through more cost-effective and immediate means enabled by the technology.

Here I lay the foundations for the development of a series of novel mood and emotion-enhancing app-based interventions developing research from established findings in neuroscience into a practical intervention, in accordance with the translational approach of this PhD.

4.1.3 Emotional contagion as a solution

As established in Chapter 1, studies show that people spontaneously mimic a variety of behaviours, among them emotional facial expressions. Indeed, facial EMG indicate that facial expressions elicit spontaneous facial muscular activity congruent with presented emotional expressions. Facial expressions are not just a manifestation of emotion but rather activating the appropriate face muscles may lead to a modification of the psychological experience of that emotion. Therefore, these series of interventions aim to enhance emotional state by inducing facial contagion - smiling - reflective of happiness.

There is evidence suggesting that smiling faces are effective stimuli and good candidates to non-invasively induce positive emotion changes in viewers. The functions

and characteristics of a smile have been extensively studied (Segerstråle and Molnár 1997). The ability to recognize a smiling face and produce a smile is long established to develop early in infancy (Jones 1926; Bühler 1931). Happy smiling faces have been consistently found to be the fastest most easily recognised facial emotional expressions (Lydon and Nixon 2014), even at great distances (Smith and Schyns 2009; Du and Martinez 2011) and when backwards masked and presented at a subconscious temporal level for other emotions (Neumann et al 2014). Further studies have shown smiles have important impacts on the behaviour and choices of others (Scharlemann et al 2001). Indeed, smiles are powerful for both those that observe the expression and those that adopt it. For example, waitresses will receive larger tips from men when smiling (Tidd and Lockard 1978), judges will dispense more lenient sentences to people who are smiling (LaFrance and Hecht 1995), and if we are smiled at by a stranger we are more likely to help another person subsequently (Gueguen and De Gail 2003).

Importantly, the beneficial effects of exhibiting genuine smiles have also been documented. A genuine or Duchenne smile (Ekman 1989) is characterised by the pulling upwards of the corner of the lips by the zygomaticus major muscle along with the orbicularis oculi muscle flexing around the eyes creating “crow’s feet”. Duchenne (1862) wrote that “the [zygomaticus] obeys the will but the [orbicularis oculi] is only put in play by the sweet emotions of the soul; the fake joy, the deceitful laugh, cannot provoke the contraction of this latter muscle”. For example, Harker and Keltner (2001) found that the more intensely a subject smiled in college yearbook photos, the more likely the subjects were to be satisfied with their lives and still married after 30 years. This finding was later replicated (Hertenstein et al 2009) and the association with smile intensity extended to predict longevity as gleaned from 1950’s baseball card pictures (Abel and Kruger 2010) and life satisfaction from Facebook profile pictures (Seder and Oishi 2012).

Moreover, participants passively observing images of a Duchenne smiling face experience unconscious congruent facial muscle reaction as measured with EMG (facial contagion) and report higher ratings of the felt emotion (emotional contagion; Surakka and Hietanen (1998)). The special status of smiles in facial contagion also generalises to real-world social interactions whereby people spontaneously mimic smiles more than frowns (Hinsz and Tomhave 1991) and is an expression that will be mimicked for both in-group and out-group members (Hess and Fischer 2014). Indeed, the beneficial effects of simply flexing the congruent muscles involved in smiling have been found to grant physiological and psychological benefits (Dimberg 1987; Schhnall and Laird 2003).

This is also true when participants are unaware that they are portraying the positive expression. For example, Strack et al (1988)[4] asked participants to rate the humorousness of cartoons while holding a pen in their mouths vertically by the tip (inhibiting flexion of the zygomaticus) or horizontally across the lips (flexing the zygomaticus and inducing a smile). Participants rated the cartoons as funnier when presented under the smiling condition than under the inhibiting condition while being unaware of the manipulation. These results were extended by Soussignan (2002) who adopted a similar procedure and in addition measured EMG activity from the Zygomaticus major and orbicularis oculi to further explore the effect of Duchenne smiles. They confirmed that the unconscious facilitation of human smiles reliably affected the rating of emotional experience and found that Duchenne smiling (involving simultaneous contraction of both zygomaticus and orbicularis oculi muscles) was key in reporting increased congruent emotional contagion.

Similarly, Kraft and Pressman (2012) asked participants to hold chopsticks in their mouths in a manner that induced a Duchenne smile, a posed smile, or a neutral expression, and induced stress under pretence of a multi-tasking task. They found that all smiling participants had lower heart rates during stress recovery than the neutral group regardless of their awareness of the induced expression (with a slight advantage for the Duchenne condition).

Furthermore, when people were instructed to frown by lowering their brow muscles, Lewis (2012) found that it was sufficient for participants to report significant negative changes to their mood. This study also found similar effects with relation to raising the eyebrows and feeling more surprised in response to facts and finding odours more unpleasant when instructed to wrinkle their nostrils. Another series of studies found that Botulinum toxin injections to the frown may reduce depressive symptoms (Murry et al 1994; Weber et al 2005; Finzi and Wasserman 2006; Finzi and Rosenthal 2014). This is an interesting finding suggesting that being unable to make a sad face has a direct effect on felt valence of negative emotions.

As it is established that inducing a smile naturally may increase a happiness response and selectively blocking this process (e.g. Botulinum toxin to corrugator) may attenuate the corresponding emotion (e.g. anger, sadness, and other negative emotions requiring a frowning response), there is a sound theoretical basis for developing a mood-

[4] Although these findings were recently called into question (Wagenmakers et al 2016), Strack (2016) provided a concise critique of this replication report along with 70 studies demonstrating the effect generalises in different ways.

enhancing intervention based on facial and emotional contagion. Though promising and shown to work even without the awareness of participants, existing methods require high degrees of invasiveness (e.g. physical injections or holding pens in one's mouth). Therefore, although the effect desired – stimulating the smiling muscles – was a simple one, it required the integration of behavioural economics principles to explore best practices in modifying behaviour with minimal disruption to daily life.

4.1.4 Designing the intervention with 'nudge theory'

Behavioural economics research has shown that it is possible to induce behavioural changes *en masse* by implementing simple changes to people's environment. Thaler and Sunstein (2008) considered 'nudges' as voluntary changes to the environment that can influence actions to achieve a desired behaviour more often. In the UK, the Behavioural Insights Team (BIT) was created in 2010 within the Cabinet Office to apply such insights to help achieve real world policy problems (Halpern 2015). In just a few years, their successes included increasing the donor register by 100,000 people (Perry et al 2015), encouraging more ethnic minorities to successfully apply to police officer positions (Linos et al 2015), and creating programs that increased rates of adherence to pharmacological treatments (Behavioural Insights Team 2015). Of course, the literature on implementing nudge interventions to increase health outcomes is large (Hollands et al 2013; Arno and Thomas 2016; Hansen et al 2016) and reviewing it in its entirety is beyond the scope of this section. However, the BIT consolidated decades of behavioural economics literature and their own experience into best-practice guidance to implement nudge interventions in their EAST framework (Behavioural Insights Team 2014).

The framework proposes that a nudge that is Easy, Attractive, Social and Timely (EAST), will increase the likelihood of a desired behaviour occurring. This acronym describes the four principles that should be part of the development process of any intervention wishing to enact behavioural change in a population. As the EAST framework is very recent, it has not yet been rigorously tested in an academic setting. Nevertheless, it provides useful guidance to develop a nudge intervention for the present purposes. Making it *Easy*, requires reducing the effort in performing the desired behaviour, making the action simple, and if a choice is required, it should be the default option. To make it *Attractive* the BIT advises designing reward or sanction mechanisms for engaging in the behaviour and making it aesthetically appealing so that it leads to a pleasant interaction. By making it *Social*, the intervention should leverage social networks, encourage people to make commitments to others or publicly, and show that

many are engaging in the desired behaviour. Finally, to make it *Timely* means to target people when they are most likely to be receptive to the nudge and making the time between the nudge and the intervention as immediate as possible.

Some of these points would have been difficult to implement ethically in this project (e.g. sanctions for lack of participation and public commitments raise anonymity issues) however many of these points were integrated into the development of the current affective state modification tool to test its potential as a successful nudge intervention. To make it *Easy*, the intervention ran on mobile devices: either tablets or participants own mobile phones. This meant everywhere they went, the nudge (i.e. smiling face) was always easily accessible to participants with minimal intrusion in their daily life. Indeed, as it masked as an application or a game, it did not require any additional effort than participants would generally engage in. This also made it *Timely* as the desired behavioural outcome (smiling) was achieved through the automatic effects of facial contagion and participants retained control of when the nudge (the smiles) would appear on their screens. The applications were all developed with user interface designers to make them *Attractive* in appearance and when being used. Lastly, the nudge itself contained an element of the *Social* framework since it was a smiling human face however a presence on social media was also maintained to encourage users to engage with the app.

4.1.5 Exploring the question in three experiments

The question of whether seeing smiling faces has a positive effect on emotions and mood was explored in three experiments that constitute the next sections of this chapter. The first experiment explored the extent to which smiling faces may influence mood over a 6-month period. The second experiment explored a similar question but operationalised it over a shorter period of 20 days. It also included the addition of a control condition where seeing smiling faces versus landscapes was compared over 10 days in each condition. Both experiments ran on smartphone applications. The third experiment compared the effects of seeing smiling faces or landscapes on transient emotions in a gamified tablet application where participants were unware that there were two viewing conditions.

While the first two experiments explored the interventional effectiveness of facial contagion and emotional contagion as nudge interventions on dispositional mood (Hacker 2009), the last operated in the realm of shorter emotional perturbations and without awareness of participants. The aim of these experiments was therefore to

address the question as fully as possible over different periods of time and different stimuli to answer the questions: can facial contagion induced emotional contagion lead to lasting positive changes in emotional states? Can these neuroscientific principles be translated into useful nudge interventions for low mood in the general population?

Experiment 1: A longitudinal pilot of mood enhancement through emotional contagion

4.2 Scope

The purpose of this iPhone based research was to pilot a large scale single arm observational study assessing whether seeing smiling faces several times a day could enhance dispositional mood as measured by standardized self-report questionnaires. This was the first presentation of a non-invasive attempt to modify self-reported mood over the long term with foundations in neuroscientific evidence that operated at a subconscious level and delivered through smartphones.

4.3 Methods

4.3.1 Study design

To assess the efficacy of seeing smiling faces several times a day in raising baseline mood in a non-clinical population, a large-scale observational study was performed through an iPhone based application (app) named Pocket Smile. The experiment was approved by a UCL Research Ethics Committee as study number 4746/001.

4.3.2 Materials

4.3.2.1 Mood questionnaires

Mood questionnaires were presented prior to commencing stimuli presentation (T0), after 10 days of viewing the smiling faces (T1), and at 1 month intervals subsequently until participants left the study freely (i.e. T2 = T1+1 month; T3 = T1+2 months; T4 = T1+3 months, etc). In-app assessments of mood were selected largely for their brievity and formed a total of 14 questions at each timepoint. These consisted of a negative mood measure, a happiness questionnaire, and a control question for potential external confounds detailed below.

The Patient Health Questionnaire 9 (PHQ-9; (Kroenke et al 2001) was instrumentalised within Pocket Smile for the self-assessment of low mood severity as "*Mood Questions*" (see Appendix B.4)[5]. This specific version was selected as it performs

[5] Although the term *depression score* will be used in reference to PHQ scores we do not mean to imply that the participant population studied are indeed depressed as no clinical in-person assessment was performed.

more accurately than ultra-brief versions (Gilbody et al 2007), comparably to longer versions such as the Beck Depression Inventory-II (Kung et al 2013), and is generally well-validated (Löwe et al 2004). In addition, it has been successfully operationalised in app-based studies for treating depression (Watts et al 2013) and is potentially more sensitive to capture certain aspects of depression than the traditional version of the PHQ-9 (Torous et al 2015).

The Subjective Happiness Scale (SHS; Lyubomirsky and Lepper (1999), a four-item, global self-assessment of happiness was implemented as "*Recent Events*" (Appendix B.5). The SHS fulfilled similar brevity criteria as those of the PHQ-9 and demonstrated cross-cultural validity in the general population (Shimai et al 2004; Swami et al 2009). An additional question was added to assess the presence of major life events to account for skewed responses.

For both questionnaires, the timescale of the questions requiring retrospection was modified to cover the time periods in between the times of interest.

4.3.2.2 Stimuli

A set of 200 genuinely smiling faces was created to ensure participants would not see the same face twice before the first mood questionnaires were presented again at T1 (i.e. T0 + 10 days) should they select the maximum setting of 20 faces a day.

A general requirement for the face stimuli was ensuring anonymity of the smiling subjects to avoid any participant encountering a person they knew. This requirement narrowed stimuli possibilities to virtual avatar faces or averages of real faces. Avatars are an ideal stimulus because they elicit equivalent facial contagion responses as real faces when measured with fEMG (Weyers et al 2006) and can be created rapidly with commercially available software such as FaceGen (http://facegen.com/). However, participants must find the faces *Attractive* to keep participating in the experiment and not to be disturbed or distracted by the stimuli. To assess preferences to virtual or average faces an online poll was conducted comparing virtual and composite sample stimuli and found a 92.68% preference for real faces (N=41; Figure 4.1). Therefore, the images used in the present experiment were derived from photographs in the public domain processed to make them unidentifiable by aligning and merging disparate features across the dataset. Though identifiable as a face, each image was therefore not the face of anyone real.

To construct the stimuli, 110 images of women and men with genuine smiles were selected from the internet, representative of a diverse range of adult ages and ethnicities, with minimal skin imperfections and face obstructions (e.g. no glasses, no hair across forehead, front facing). Both faces depicting genuine "Duchenne" smiles and fake smiles elicit similar facial contagion responses as measured with EMG, however Duchenne smiles induce stronger facial contagion responses (Krumhuber et al 2013) leading us to select images exhibiting Duchenne smiles to form the basis for the stimuli. The images were then imported into Abrasoft FaceMixer 3 (http://www.abrosoft.com/facemixer/) to perform the parametrisation and averaging of the faces[6]. This allowed automatic parametrisation and more meticulous control over averaged face composition. All faces were created from 2% of all faces of the same gender and 20% from three different faces per iteration. This ensured the creation of 100 images of each gender with distinct individual differences. A contour of the average of each image was created and turned into a short mask in MIPAV (http://mipav.cit.nih.gov/). This black-and-white background was then converted to grey (RGB: 185,184,180) and multiplied with the average faces in MATLAB after applying a 0.01 low pass Gaussian blur so that the faces were roughly contoured by a smooth, neutral grey background rather than the average of the pictures' original backgrounds which varied immensely at times and could be distracting. These images were imported into Photoshop CS6 where the background was blurred further to maximise attention to the smiling face and cropped to fit the screens of the iPhone 5 (640 x 960) and previous versions (320 x 480). This process yielded 200 images which were presented in a full pseudorandom cycle.

[6] The reader may find it useful to refer to the visual depiction of the image creation process provided in Appendix B.11 to accompany the remainder of the paragraph.

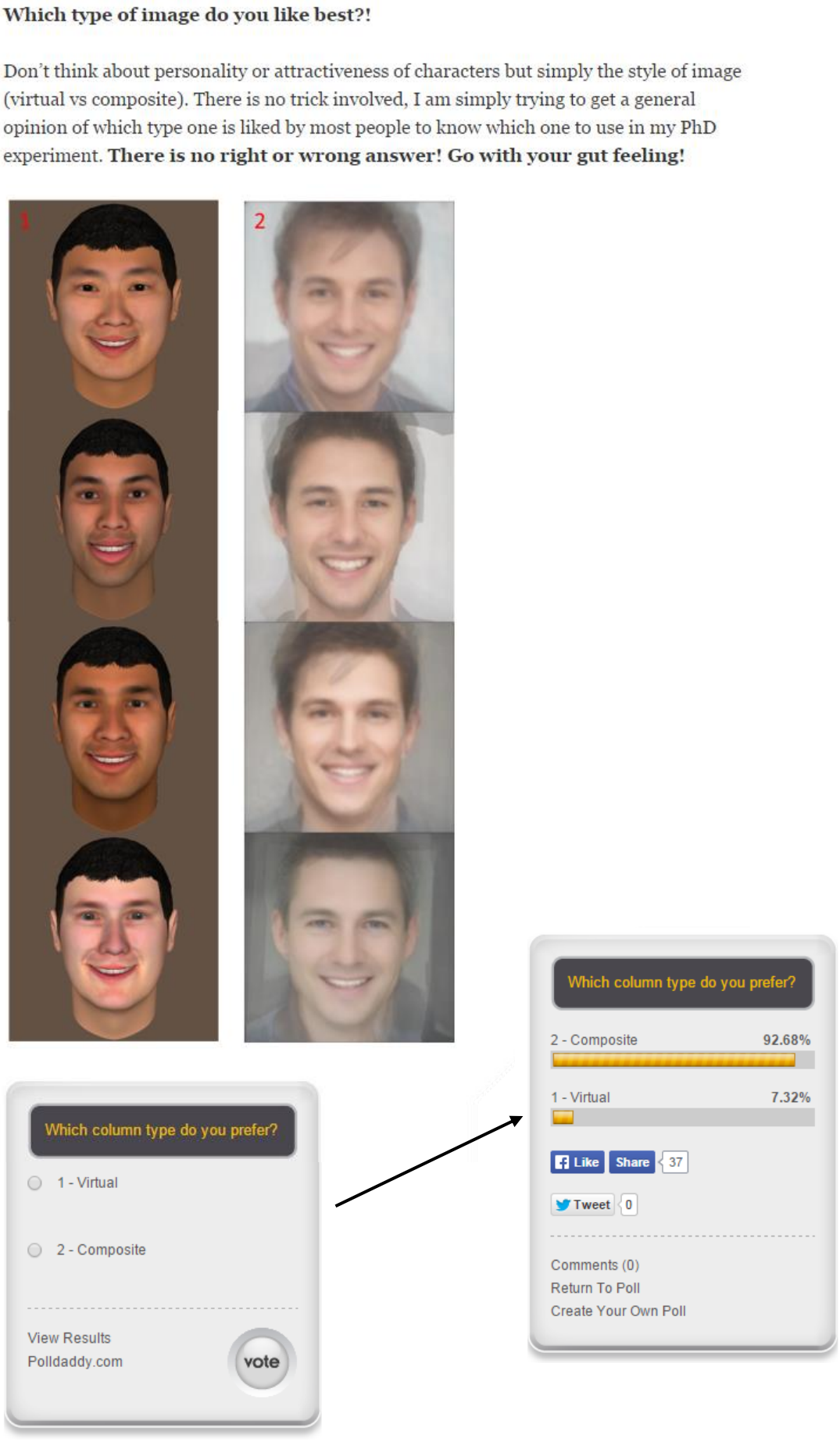


Figure 4.1. Stimulus type selection. This figure shows the online poll used to assess preferences to avatar and composites prior to creating the full dataset of face stimuli. Participants were instructed to submit a simple aesthetic preference and composite stimuli were widely preferred (N=41).

4.3.3 Participants

Participants were recruited opportunistically by freely downloading the app from the iTunes App Store (see Appendix B.8). Though the app was only available in English, the description on the App Store was translated into French, Spanish, and Italian for increased exposure to international participants. To maximise recruitment, an online presence on social media was maintained on Facebook and Twitter (Appendix B.10) that redirected participants to a dedicated website with information about the project optimized for mobile and desktop devices (Appendix B.9). Following research showing that attitudes towards electronic self-help treatment options for mental health are subject to the amount and quality of information available to the user (Musiat et al 2014), efforts were made to provide participants with sufficient information to make an informed choice during sign up and on the website as well as operating a dedicated email contact address (contact-pocketsmile@ucl.ac.uk). Significant promotional activity was undertaken to promote Pocket Smile including three public engagement events on "The Science of Smiles" and networking with mental health charities and the press. Social networking platforms also served as sources for the latest information on the project and as accessible platforms for contact to current and prospective participants which were also embedded into the official website.

Although other mobile-based studies related to mood have previously incentivised participation (e.g. LiKamWa et al (2011), no offer of payment or compensation was offered beyond telling participants they would be helping the potential development of a drug-free intervention to enhance mood. Since usage of the app is not onerous, the incentive of a monetary or other reward could have led to using the app as a lottery-style system and not looking at the faces at all. Also, considering participants were recruited worldwide, a bank transfer or address would have been necessary to award and certify the prize that would have complicated ethics significantly.

4.3.4 The Pocket Smile infrastructure

A back-end developer and a designer were hired from Tapparium ltd to create the app named *Pocket Smile* (source code in Appendix B.14). Over a 3-months reiterative process with the research team, their work yielded an app compatible with iPhone, iPad, and iPod touch running iOS 5.1 or later, a dedicated website and contact email address, and a secure storage database running on a local server at the UCL Institute of Cognitive Neuroscience.

Extensive effort was undertaken to develop an attractive app with a coherent design and user-friendly interface that would perform successfully as a scientific experiment. The logo and name were conceived to portray the function of the app intuitively, by portraying a smiling emoticon in a denim-style pocket (Fig 4.2).

Figure 4.2. The Pocket Smile Logo.

4.3.4.1 App implementation

The first time the app was launched, participants landed on a 'welcome' menu with the first of four steps enabled for completion required for taking part in the study (see Appendix B.1). They were presented with a combined PIS and consent form (see Appendix B.2), where they self-assessed eligibility to participate and tapped an 'I agree' button to give consent. The user also had the option to contact us by email or using other contact details provided prior to giving consent. If they did not consent to take part in the research, they were unable to use the remaining functionalities of the app and participate in the study. Participants who gave consent were asked to provide basic demographic information in the subsequent section (Appendix B.3). None of this data constituted 'identifying information' as classified by the Data Protection Act 1998. Participants were then presented with an initial 14 question assessment of low mood and happiness which provided baseline scores for further comparisons. These steps completed the registration process.

Subsequently, participants were automatically taken to the 'Home' tab of the in-app menus (see Appendix B.6) where a countdown informed them of the time until the next face was ready for viewing. Pocket Smile also contained 3 other bottom tabs consisting of 1) an 'About' section, with a link to the project website, options for contacting the team, references to the questionnaires, and credits to Tapparium developers, 2) a 'Settings' section consisting of 3 sliders to regulate the frequency and timing of

appearance of the stimuli, and 3) a 'Help' section with frequently asked questions, access to the PIS, and contact details for the research team.

Once participants entered their preferred settings, the app sent a push-notification when a face was available for viewing accompanied by the text: "a new smile awaits you!" (see Appendix B.7 for a sample notification depiction and all stimuli). Participants who acknowledged the notification by either sliding the notification or selecting the app icon would be presented with one of the created stimuli until they left the study or set their settings to '0 faces per day'. They could then return to the 'Home' tab by tapping a black arrow at the bottom right corner of the image and or exit the app by pressing the iPhone's 'Home' button.

4.3.4.2 The server database

The data from Pocket Smile was sent over an industry standard encrypted secure connection (Secure Sockets Layer; SSL) and stored in a firewalled Apache HTTP Server at the UCL Institute of Cognitive Neuroscience. The server ran a MySQL database management system and used InnoDB as its storage engine. A PHP script created an API endpoint reference on the server to which Pocket Smile could send an HTTP POST request with the relevant data to upload. The request consisted of an HTTP Content Body to be included in a JavaScript Object Notation (JSON). The JSON was then saved to a single column in a row that could be browsed through phpMyAdmin and exported in a Comma-Separated Value file (CSV) to form a spreadsheet layout.

The initial demographics, time zone, and mood questionnaire data, were sent to the server after sign-up or when participants next had network coverage. Face appearances and settings configuration were stored in the app and sent to the server on a weekly basis henceforth. Mood questionnaire responses were uploaded after participants responded. All uploaded data was fully anonymised by use of unique device identifiers, which were never attached to personally identifying data.

A Countly infrastructure provided online mobile data analytics to the research team through a password protected web-interface. This platform was mostly consulted for general tracking and monitoring purposes and to have a quick overview of usage analytics (Appendix B.12).

4.3.5 Data analysis

The CSV file containing the data was exported from the database into MATLAB where it was re-organised to score the questionnaires.

Since the PHQ-9 scores each of the DSM-IV criteria for depression from "0" (not at all) to "3" (nearly every day) and yields a total score of depression severity of: 0-4 none, 5-9 mild, 10-14 moderate, 15-19 moderately severe, and 20-27 severe, a total score per participant at each timepoint was computed. SHS responses were scored as likert type scores from 0 to 7 and the last item was reversed score. The total of each score per participant were averaged yielding an SHS score indicating 0-2 low, 3-5 average, and 6-7 high happiness.

These were then inserted into a large table with one row per participant consisting of a randomly generated username and codified responses for gender, age, life satsifaction, overall health, relationship status, employment status, and time zone, followed by PHQ, SHS, and major life event scores for each timepoint.

Overall changes in PHQ and SHS scores were then explored with separate mixed effects repeated-measures ANOVAs in SPSS with timepoints as within participants factors. A mixed-effects model was opted to be more appropriate than classic repeated-measues ANOVA as it corrects for intra-participant correlation of scores at subsequent timepoints.

The effect of faces seen on participants was explored running linear regressions in SPSS comparing changes in scores as defined by *test of interest at timepoint of interest* minus *baseline score* (e.g. PHQ score at T1 minus PHQ score at T0) with regressor variables of interest (i.e. T0 score (baseline), average faces seen between timepoints of interest, gender, age, presence of a major life event). Since there was no control group, the presence of a dose response was explored between changes in mood questionnaire scores and average faces seen. A linear relationship was explored whereby higher averages were expected to drive larger changes in scores.

4.4.1 Descriptive statistics

Pocket Smile was downloaded 1742 times from the App Store and presented over 112,742 smiling faces to participants. Complete registration data were obtained from 1293 participants mainly from English speaking countries (92%; Figure 4.3).

Country	New Users
United Kingdom	868
United States	238
France	132
Unknown	54
Australia	47
Italy	41
Canada	34
Mexico	31
Germany	20
Spain	19

Figure 4.3. Top 10 participant nationalities as ranked by Countly

As can be seen in Table 4.1, at baseline most participants were female and around 30 years of age, though 47% were between 19 and 25 years. Overall, participants reported slightly above average life satisfaction (see Appendix B.3 for questions), "Good" average health, and most reported being 'Single' (35%), 'In a relationship' (32%), or 'Married' (34%), and 'Employed' (33%) or 'Studying' (50%).

Attrition analyses at baseline demonstrated that individuals who did not respond at follow-up were older, male, no longer employed, reported poor overall health at baseline, and tended to be in Africa, Australia, India, or the Pacific Islands. It is important to note that the studied population is not stationary but rather reflects a snapshot of asynchronous moving windows. Since there was no fixed starting date there were participants in earlier timepoints who remained active in the study, thus missing data could signify incomplete data because the user had recently joined the study or dropout due to earlier improvement. Time constraints did not allow this consideration to be included in the attrition analyses but can be performed in future work to explore whether those leaving the study are not more depressed at baseline and therefore introducing a bias in the sample studied.

Table 4.1 – Demographic Characteristics of Participants in the Pocket Smile Study and Attrition Comparison from Baseline Key Variables

	Sample (*n*)	1293	394	169	104	62	39	31	25
	Timepoint	T0	T1	T2	T3	T4	T5	T6	T7
Gender, *n* (%)									
Male		461 (35.7)	124 (31.5)	56 (33.1)	32 (30.8)	19 (30.6)	11 (28.2)	8 (25)	5 (20)
Female		832 (64.3)	270 (68.5)	113 (66.9)	72 (69.2)	43 (69.4)	28 (71.8)	24 (75)	20 (80)
Age group, *n* (%)									
18-34		992 (77)	276 (70)	112 (66.3)	65 (62.5)	35 (56.5)	22 (56.4)	17 (55)	16 (64)
35-51		228 (18)	91 (23)	45 (26.6)	31 (29.8)	22 (35.5)	15 (38.4)	12 (39)	8 (32)
52-68		69 (5.7)	26 (6.6)	12 (7)	8 (8)	5 (8)	2 (5)	2 (6)	1 (4)
69+		4 (0.3)	1 (0.4)	0	0	0	0	0	0
Mean age, *continuous* (SD)		29 (10.7)	31 (11.1)	32 (11.1)	33 (11.4)	34 (11.6)	34 (10.4)	34 (10.0)	33 (9.2)
Employment, *n* (%)									
Studying		651 (50.3)	200 (46.2)	63 (37.3)	38 (36.5)	21 (33.9)	12 (30.8)	9 (29)	8 (32)
Employed		432 (33.4)	163 (41.4)	81 (47.9)	51 (49)	39 (62.9)	25 (64.1)	21 (67.7)	16 (64)
Retired		10 (0.08)	8 (3)	6 (3.5)	4 (3.8)	0	0	0	0
Unemployed		49 (3.8)	22 (0.6)	10 (0.6)	5 (4.8)	1 (1.6)	0	0	0
Skipped		28 (2.2)	1 (.2)	9 (0.5)	6 (5.8)	1 (1.6)	2 (5.1)	1 (3.2)	1 (4)
Life Satisfaction, *n* (%)									
Low (0-3)		134 (10.1)	39 (9.9)	17 (10)	10 (9.6)	8 (12.9)	7 (18.0)	7 (21.9)	6 (24)
Medium (4-6)		490 (37.3)	142 (36)	68 (40)	48 (47)	31 (50)	17 (43.5)	12 (38.7)	9 (36)
High (7-10)		669 (51.7)	208 (52.1)	84 (50)	46 (44.4)	23 (37.1)	15 (38.5)	13 (41.9)	10 (40)

Continued

Table 4.1 - Continued

	Sample (*n*)	1293	394	169	104	62	39	31	25
	Timepoint	T0	T1	T2	T3	T4	T5	T6	T7
Overall Health, *n* (%)									
Poor		86 (6.8)	20 (5.2)	12 (0.6)	0	0	0	0	0
Fair		226 (17.4)	70 (17.7)	30 (18)	25 (24)	17 (27.4)	12 (30.8)	10 (32.3)	9 (36)
Good		621 (48.0)	194 (49.2)	86 (51)	50 (48.1)	29 (46.8)	20 (51.3)	17 (53.1)	11 (44)
Very Good		360 (27.8)	110 (27.9)	41 (24.4)	29 (27.9)	16 (25.8)	7 (17.9)	4 (12.9)	5 (20)
Marital status, *n* (%)									
Single		457 (35.3)	126 (32)	39 (23.1)	27 (26)	18 (29)	13 (33.3)	12 (37.5)	11 (44)
In a Relationship		425 (32.9)	135 (34.3)	57 (33.7)	30 (28.8)	17 (27.4)	9 (23.1)	7 (21.9)	5 (20)
Married		321 (24.8)	112 (28.4)	60 (35.5)	44 (42.3)	26 (41.9)	16 (41)	12 (37.5)	9 (36)
Widowed		45 (3.5)	10 (2.5)	8 (4.7)	3 (2.9)	1 (1.6)	1 (2.6)	1 (3.1)	0
Divorced		28 (2.2)	7 (1.8)	2 (1.2)	0	0	0	0	0
Skipped		17 (1.3)	4 (1.0)	3 (1.8)	0	0	0	0	0
Geographical location, *n* (%)									
Africa		10 (0.8)	0	0	0	0	0	0	0
North and South America		252 (19.5)	76 (19.3)	27 (16)	15 (14.4)	10 (16.1)	6 (15.4)	5 (15.6)	5 (20)
Asia		53 (4.1)	8 (2)	7 (4.1)	4 (3.8)	3 (4.8)	2 (5.1)	2 (6.3)	1 (4)
Australia		36 (2.8)	13 (3.3)	5 (3)	2 (1.9)	1 (1.6)	1 (2.6)	1 (3.1)	0
Europe		928 (71.9)	295 (74.9)	129 (76.3)	83 (79.8)	48 (77.4)	30 (76.9)	24 (75)	19 (76)
India and Pacific Islands		12 (0.9)	2 (0.5)	1 (0.6)	0	0	0	0	0

4.4.2 Mood questionnaire scores

The mean depression PHQ scores at baseline for the total sample were characteristic of *mild depression* as measured with the PHQ-9 (*M* = 8.83, *SD* = 4.92). Mean happiness scores were below the median value of 4.75 as measured with the SHS (*M* = 3.93, *SD* = 1.5) and indicative of a state of unhappiness (Lyubomirsky and Lepper 1999). Both results are unsurprising since participants self-selected into the study and did so most likely because of their low mood. Most participants reported they had not experienced any major life event (61%). Participants who experienced a major life event generally reported no effect on their mood (26.2%), while a minority reported these events resulting in significant negative (7.2%) and positive (5.6%) changes to their mood. These individuals were discarded from further analyses.

Mixed-effects repeated-measures showed that participants reported a quick initial improvement in mood followed by a stabilisation of mood that suggested higher happiness than at baseline, but no continued improvement subsequently (see Fig. 4.4)[7].

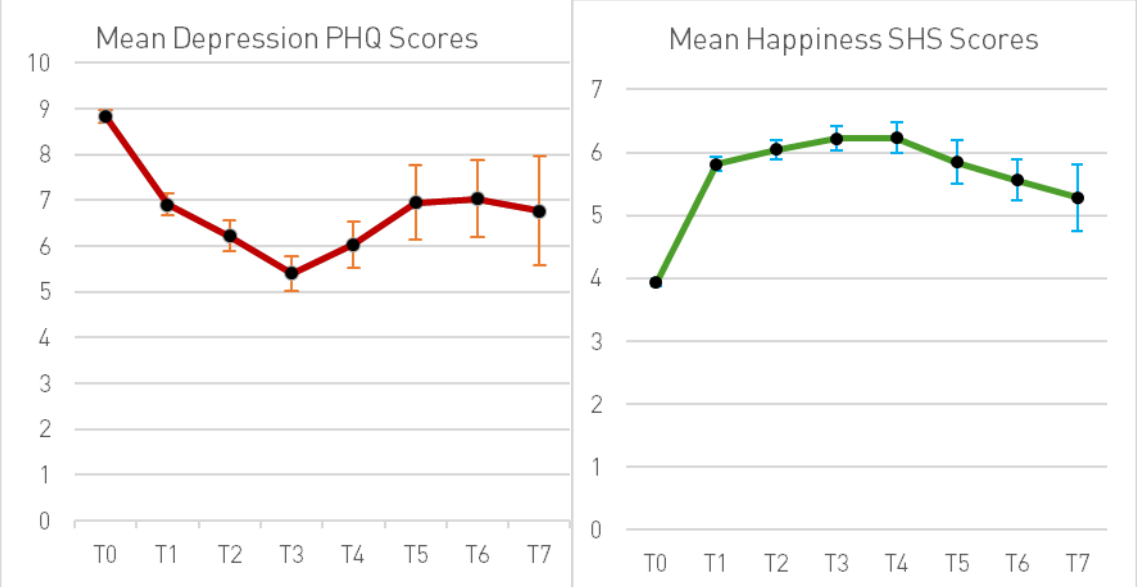


Figure 4.4. Mean depression and happiness scores at baseline (T0), at 10 days (T1), and at 1 month subsequently for the first six months of the study, plotted with standard error bars. Higher PHQ scores indicate higher depression and higher SHS scores indicate higher happiness.

[7] see Appendix B.13 for mixed effects repeated measures results

4.4.3 Evaluating the intervention

Since no changes in scores were observed following initial improvement after 10 days, the intervention was assessed at T1 and at the last timepoint (i.e. T7; 6 months). This method therefore limited results to the timepoints of most interest: when there were the most participants (T1) and at the last gathered measurement point (T7), ensuring conservative analyses and yielding less likelihood of type-2 errors related to multiple comparisons.

Visual inspection of the data at the timepoints of interest showed weak trends in line with the hypothesised direction of treatment effectiveness at T1 and stronger trends at T7 (Figure 4.5).

Multiple linear regressions were used to evaluate the effect of *average faces seen* on changes in mood for both mood questionnaires at T1 and T7. Change in mood was computed as the mood questionnaire score at baseline subtracted from that at a latter timepoint (i.e. T1/7 minus T0) therefore adjusting mood levels prior to enrolment. Univariate linear regressions were performed in SPSS to explore which baseline characteristics were related to changes in score. The variables consisted of dynamic variables (average faces seen during time period and presence of a major life event at each timepoint) and unchanging demographic baseline variables (gender, age, life satisfaction, overall health, geographical location, marital and employment status).

At T1, unadjusted linear regressions only returned a significant model for *Life Satisfaction* for PHQ (F (1, 391) = 10.33, p < .001, $R2$ = .026) and SHS (F (1, 391) = 11.22, p < .001, $R2$ = .028) measures (Figure 4.5). At T7, adjusted linear regressions only returned a significant model for *Average Faces Seen* for the SHS (F (1, 22) = 4.34, p < .05, $R2$ = .165) measure (Figure 4.6).

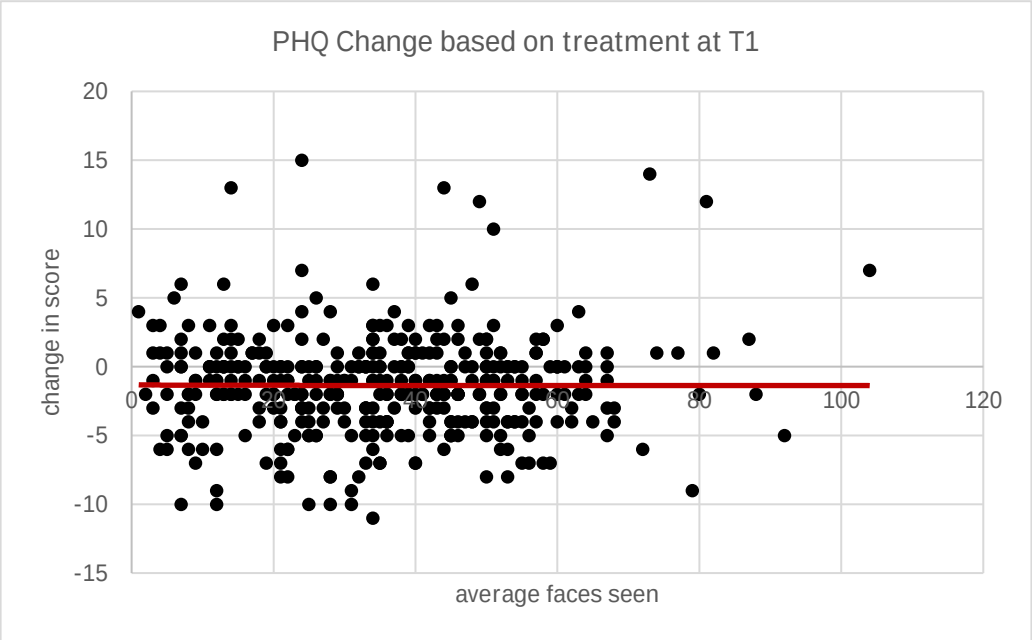
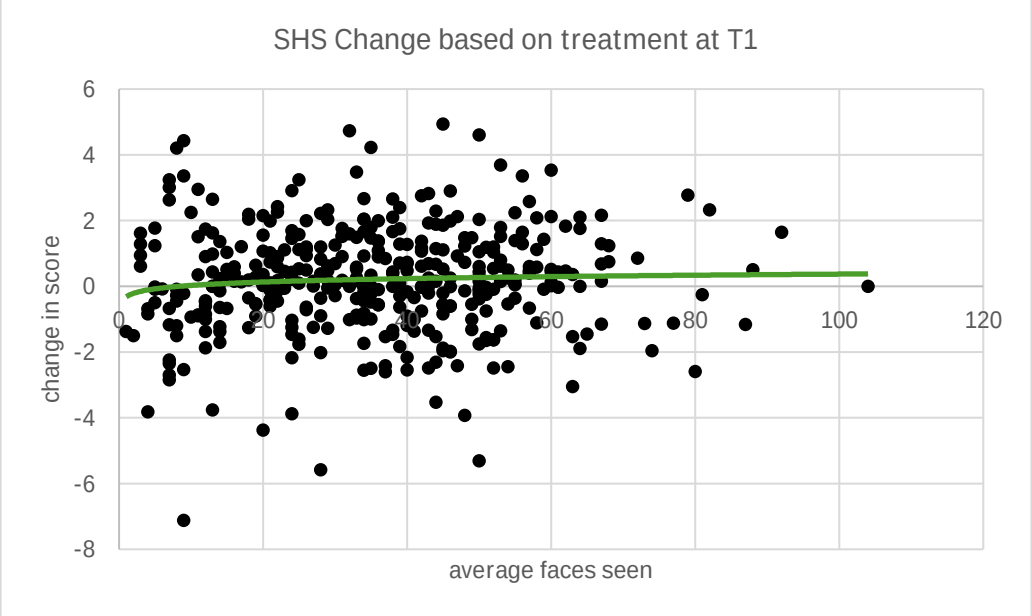


Figure 4.5. Changes in scores for all participants at T1 (10 Days) for depression (top) and happiness (bottom) scores plotted with average faces seen and fitted with a logarithmic trend line.



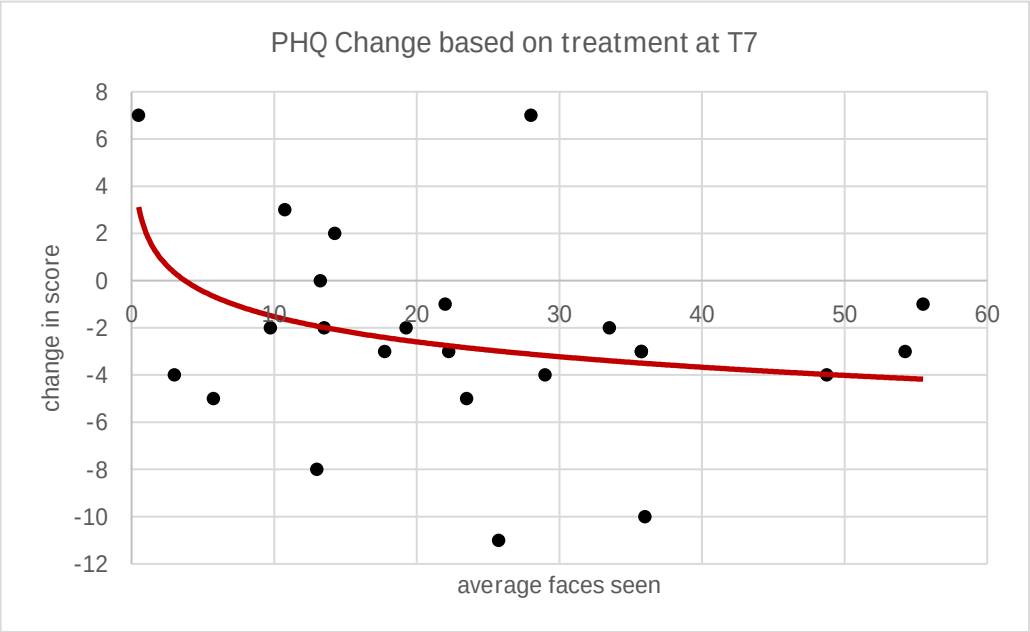
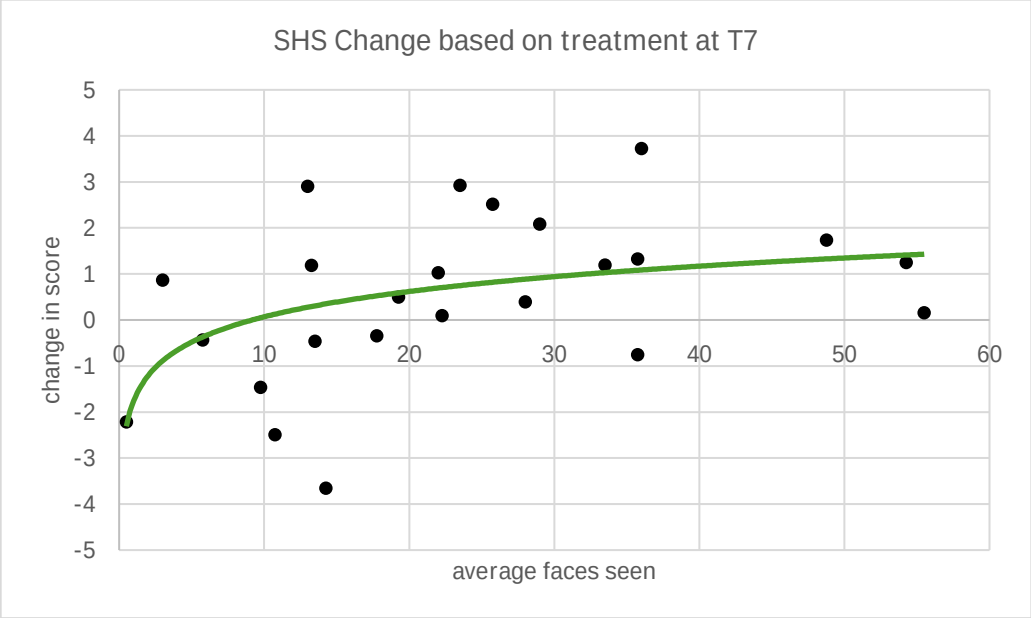


Figure 4.6. Changes in scores for all participants at T7 (6 months) for depression (top) and happiness (bottom) scores plotted with average faces seen and fitted with a logarithmic trend line.



4.5 Discussion

4.5.1 Conclusions

This section presented the analyses of the long-term version of Pocket Smile; the first fully anonymised app-based intervention for low mood to demonstrate general improvements in mood. The results indicated that Pocket Smile participants reported reduced depressive symptoms as measured with the PHQ-9 and SHS when comparing pre- to post-intervention scores at baseline and up to 6 months subsequently. Since the main goal of this study was to test feasibility of conducting such research with mobile devices it was also evidence of support for an intervention founded on facial contagion and delivered through smartphones.

Changes in low-mood and happiness scores indicated a clear improvement in mood as demonstrated by independent self-report measures. PHQ-9 scores decreased an average of almost 4 points at peak difference, indicative of a 39% decrease in PHQ-9 assessed depression from 'moderate' to 'mild'. These results show stronger effects of change on a participant population with mild depression than other app-based studies looking at similar populations as measured by change of PHQ-9 score (Watts et al 2013). For example, once participants in the Watts et al (2013) study had reached medium to low depression as measured with the PHQ-9 questionnaire, their scores descended an average of 2.5 less points compared to Pocket Smile participants. Compared to Ly et al (2014), Pocket Smile participants exhibited an improvement of 2 PHQ-9 points less however this study also included up to 20 minutes of therapist contact per week and CBT delivered through the app. Thus, the strong changes seen in Pocket Smile with the comparatively minor requirement of the intervention suggest it was effective intervention.

Accordingly, SHS scores increased on average by 2.3 points at peak difference, indicative of a 37% increase in happiness, from 'below average' to 'slightly above average' happiness. Participants reported improvements in mood that were then robustly maintained, suggesting that these are promising proof-of-concept observations.

An exploration of the strength of the intervention showed that after continued usage of the app, participants who saw more smiling faces between mood questionnaires also reported the most benefits. However, this statistical support for the effectiveness of Pocket Smile was only found for the Happiness measure at 6 months,

while life satisfaction prior to treatment was the best predictor for score changes after 10 days. These results suggest that Pocket Smile had a beneficial effect on general mood though the mechanisms interacting with the treatment may be multifaceted.

This is the first comparison of the PHQ, an egocentric subjective measure of depression, and the SHS, an allocentric measure of happiness, administered within the same English-speaking population. It is therefore possible that, though they demonstrate negatively correlated results as expected, the aspects of mood they measure are differently affected by average faces seen. Thus, Pocket Smile's positive effect on certain aspects of mood may become independent of the high number of faces seen as the simple reminder of seeing a few faces every day could be enough to maintain the individual at a higher mood level than prior to enrolment. It is also possible that other variables such as motivation or expectation that the intervention will work could lead to higher adherence to the study. This would be in line with the finding of higher life satisfaction predicting improvement in mood at 10 days, as previous studies find participants with optimistic expectations to work harder and longer at following the instructions of voluntary mood enhancing treatments (Seligman et al 2005; Lyubomirsky et al 2011).

Lastly, a principal goal of this pilot was establishing feasibility and addressing concerns that dropout rates might be too high to include a voluntary control condition. It was interesting that most people who dropped out were unemployed older than average (31+ years of age) males with poor overall health and from non-European or North American countries. Perhaps this is due to the smiling faces looking quite young overall and could provide reason to not identify easily with these stimuli, an issue addressed in the next experiment.

4.5.2 Limitations

As a main purpose of this experiment was to establish the feasibility of this approach no control condition was included to assess the effect of the faces. A way to circumvent this issue would have been to include a group naïve to the experiment in a control viewing condition where they would see, for example, landscapes. Nevertheless, it was a general impossibility to operationalise a control condition for a publicly listed app-based study on mood as participation was anonymous and proper debriefing would not be possible. Also, there would be no way to certify participants remained naïve to the manipulation after answering questions about their mood on a monthly basis. Still this could mean

that the improvement in mood seen represents spontaneous recovery of mood, independent of Pocket Smile usage.

Another shortcoming of the present study is that participants self-selected into Pocket Smile. The large studied sample was mostly representative of the general population in Europe and was effectively attracted to the free app-based treatment. Many participants continued using the intervention voluntarily for an extended period (over 6 months). As the app was freely available, the sample population did not necessarily engage with the treatment in the same way as a patient population might have, had it been recommended by a doctor as a supplementary therapy. In comparison to similar studies (Watts et al 2013; Ly et al 2014), the participant population reported less severe depressive symptoms at baseline as measured by the PHQ-9 (i.e. up to 6 and 4 points less respectively on average). To account for this factor and explore the effect of the intervention in the target population rather than on casual app users -who might particpate out of curiosity rather than need- future analyses will exclude those participants that do not exhibit at least mild depression at baseline as defined by the measures (e.g. by removing all participants who scored under 5 in the PHQ-9).

Although the app was available for download worldwide, the study sample is more representative of the general population in Europe. This allows comparison of the results with Western European and American populations, but no claim can be made as to the representativeness of countries with higher cultural heterogeneity as there are lower rates of participants located in Africa and Southeast Asia.

Additionally, the current design did not permit the analysis of change in mood to be compared with *total* faces seen between mood questionnaires. This was due to the significant time discrepancy between T0 and T1 (10 Days) and subsequent timepoints (1 month). Although utilising average faces seen between the timepoints was considered to be a valid measure, changes to the future experimental designs could be made to allow this analysis.

Ideally a measure of empathy should have been administered at baseline, since this could affect participants' reactivity to facial contagion (Sonnby–Borgström 2002) and thus introduces a potential confound on the intervention as it may have been less beneficial for those with lower empathy scores. As the shortest validated empathy measure requires responding to 16 items (Spreng* et al 2009), this might have proven too onerous on the registration process and was therefore opted against.

4.5.3 Future directions

This version of Pocket Smile showed that it was possible to influence changes in mood in a large population through an app with results indicating comparable or more potent effects for moderate to mildly depressed populations than existing app-based depression therapies. It therefore holds potential to be further studied as a candidate therapy to help afflicted individuals. If found to be effective only for those with mild depression, this would still be beneficial for society as they remain significantly burdened as a patient population (Jain et al 2013).

Since this is the first demonstration of the effect that facial contagion can have on mood modification through a smartphone, it is important to take this phenomenon into account when designing future app-based mental health treatments. For example, the *Happy Taps* function of the *Mood Mint* app for iPhone claims to re-direct attentional bias from negative to positive valence stimuli and "reduce stress, anxiety and depression". The app requires participants to find and tap a smiling face among a mosaic of four faces where one is smiling and the others exhibit negative expressions such as anger, sadness, or fear. As the effects of facial contagion occur at a subconscious level it is possible that the app is unintentionally inducing negative emotions in participants due to the three times higher presentation of negatively valenced facial expressions. This is concerning considering the high degree of comorbidity between anxiety and depression (Alloy et al 1990; Brady and Kendall 1992) which could result in depressed patients accessing this intervention which has not been properly validated yet. Indeed, as no peer review or scientific publications are necessary to publish an app and label it as a treatment for mental health disorders, I advise prospective users to only consider app-based treatments with prior scientific or clinical validation. I also encourage colleagues to make their apps available as treatments only following rigorous scientific process or have it designated as a medical device where appropriate (Medicines and Healthcare products Regulatory Agency 2016).

Experiment 2: Assessing the specificity of changes in mood to seeing smiling faces

4.6 Scope

The purpose of this experiment was to explore the mechanism behind the effects of the first Pocket Smile app experiment while integrating the feedback received by participants. Since the goal of the Pocket Smile project remained creating a nudge-based intervention that exploits facial contagion to improve low mood, this version tested the validity of the face stimuli rather than the theoretical explanations posited beforehand as the main driver behind the Pocket Smile effect. The main changes with the previous experiment consisted of changes to the 1) stimuli conditions and 2) the timings of the questionnaires, to transform Pocket Smile from a feasibility study into a double blinded randomized-controlled trial (RCT).

4.7 Methods

4.7.1 Study design

Whereas previously the experiment had only one stimulus condition (smiling faces), this version compared the effect of viewing landscapes and smiling faces. Landscapes were used as control stimuli to determine whether the effect observed was specific to viewing smiling faces and whether seeing smiling faces several times a day could enhance mood more than seeing landscapes, a type of neutral non-social valence stimulus. Viewing conditions were setup in a Latin-square design whereby participants were randomly allocated to view either landscapes or smiling faces for the first 10 days after the initial mood assessment. Subsequently, their mood was assessed again, and participants were switched to the other image condition for the next 10 days before the final mood assessment. For example, a participant would begin the experiment by answering the baseline initial questionnaire and randomly allocated to viewing landscapes. After 10 days of viewing landscapes they were given the questionnaire again, and once completed the app showed smiling faces for 10 additional days. After 20 days, the mood questionnaire was presented again. These changes allowed for more data points per participant to be obtained in less time, thereby improving the reliability of the data. This experiment was ethically approved by the UCL REC as an extension from the previous experiment with minor amendments.

4.7.2 Materials

4.7.2.1 Mood questionnaires

Feedback from the first experiment suggested that the questionnaire timings were not onerous on participants however the greatest dropout was found after 1 month. Therefore, efforts were made to shorten the intervals to 10 days and a maximum of 20 days total participation was required. The same mood questionnaires (PHQ-9 and SHS) were used as those in the first Pocket Smile experiment with the retrospective time adapted to cover the 10 days prior to responding. These were administered initally, after signing up, to provide a baseline measure and at day 10 of participation, during the middle of the experiment and prior to switching viewing conditions, and at day 20, after completing the experiment. Since participants were randomly allocated to a stimulus condition at baseline and switched to the other after 10 days, after 20 days they would have experienced both conditions which is when official data collection ceased.

4.7.2.2 Stimuli

A set of 94 genuinely smiling faces of both genders and 94 landscapes were created to ensure participants did not see the same image more than once during the experiment. The same considerations were taken to create and select these images as in the previous experiment.

A variation in the software used to create the averages was introduced by importing the images into a web version of Psychomorph (www.WebMorph.org) where 189 points were placed on facial features manually. This configuration was used as a reference for averaging the faces. To create the composite averages, 3 images were averaged and so forth, incrementally, with the aim of yielding maximum genetic variation across generations with minimum repetition. The faces were then aligned and imported into Photoshop CS6 where each face contour was traced manually to include facial expression and leave out hair, ears, and neck. The edge of this contour was feathered at a 10-pixels radius and a Gaussian blur at 7-pixels was applied to the inverse selection (i.e. the background). The images were then cut and resized, yielding 94 images which were each uniquely presented during the 20 days (see Figure 4.7). This created a new range of higher quality smiling average faces that responded to previous user feedback (less homogenous, more ethnic and age differences, more realistic).

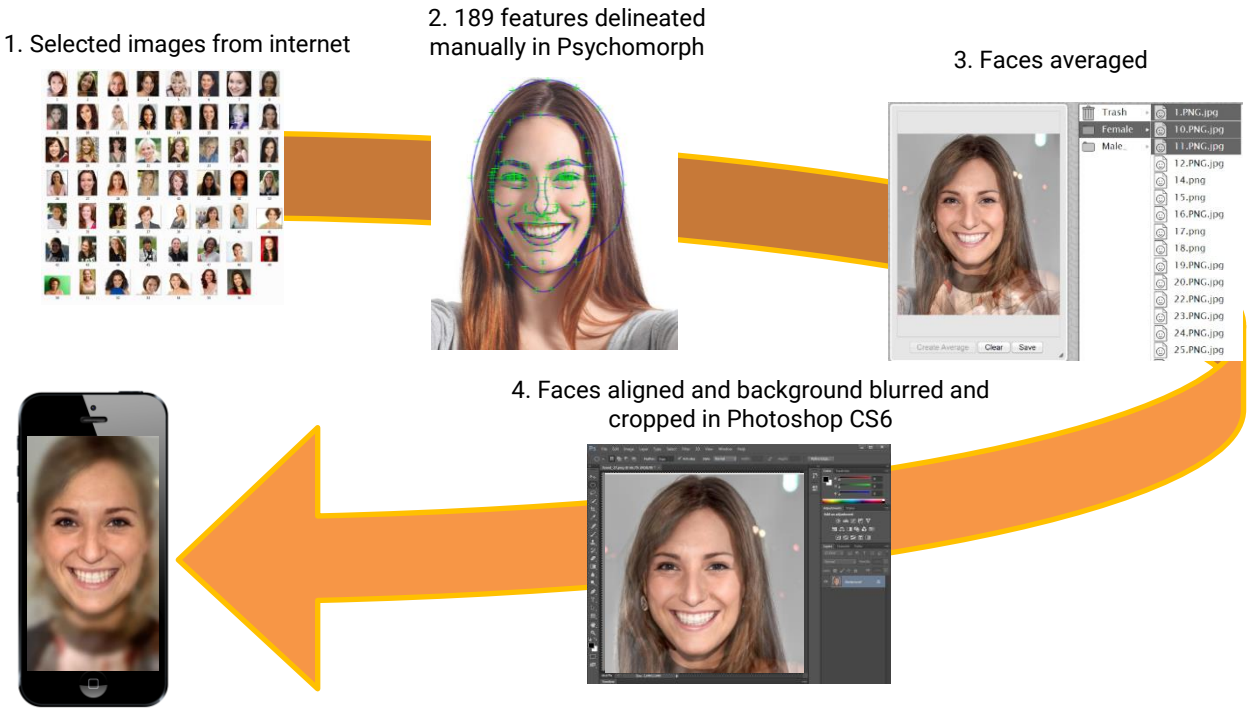


Figure 4.7 Image creation process for face stimuli. 1) Images were selected from the internet exhibiting genuine Duchenne smiles. 2) Facial features were delineated through 189 points used for reference during 3) averaging. 4) Resulting faces were modified to hide unreferenced artefactual background features and used as images in the second app experiment.

For the landscapes, a Google search was performed for scenic landscapes of beaches, lakes, mountains, meadows, hills, and deserts. Images free of copyright restrictions without social stimuli (i.e. including depictions of people or major human constructions) labelled for non-commercial re-use with modifications were selected and cropped to the appropriate size (see Figure 4.8).

Figure 4.8 Sample notifications and all stimuli with two example image options.

4.7.3 Participants

Participants were recruited through similar channels as the Pocket Smile experiment however a small budget was also allocated to advertise through Facebook. Three digital adverts were created to promote downloads of the app targeting iPhone and Android users globally (Figure 4.9). This allowed unrestricted global access to 1.79 Billion potential participants across the world while maintaining the same social media and digital presence as in the previous experiment.

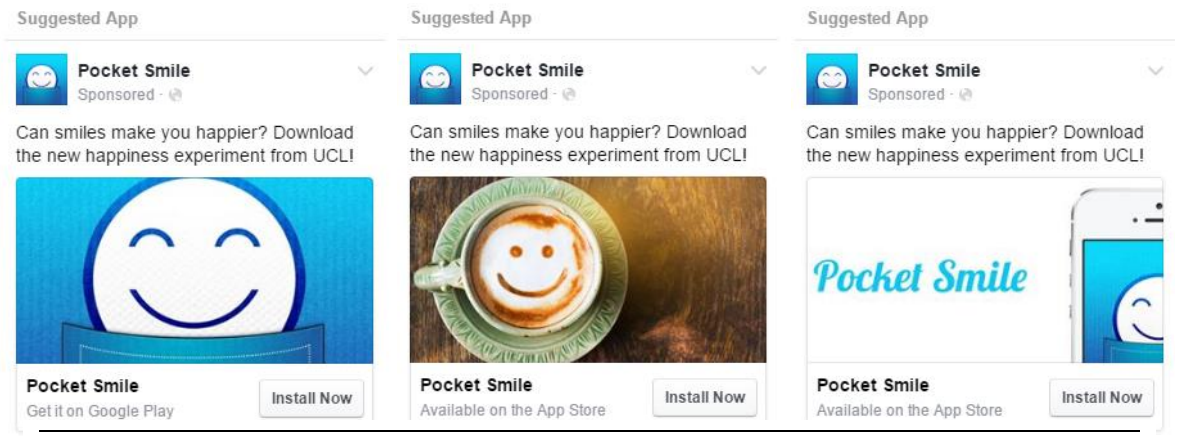


Figure 4.9. Facebook advertisements used to reach users globally in the second Pocket Smile experiment.

4.7.4 The app

The current app was identical in many ways including in style and design to the first version of the app. Milo Creative developers were hired to perform the backend changes required to the app infrastructure and release the app on iOS and Android (source code in Appendix B.15). This involved several tasks including coding the new landscape condition, the new timings and switching after 10 days, randomly allocating participants to either conditions at the start, changing the landing page from the home page to the settings page, including a new timer about when the next questionnaires were going to appear, allowing participants to choose what images to see upon completion of the experiment, notification to display a stimulus-neutral message “a new image awaits you” rather than “a new *smile* awaits you”, and all the appropriate changes to the database.

The same questions and data were gathered as in the first app except for the Countly analytics since these were not compatible with React Native. A new database

was created to house this new experiment in the Institute of Cognitive Neuroscience at UCL.

4.7.5 Data analysis

Two samples t-tests were used to compare baseline mood questionnaire scores and changes in scores over time, and paired samples t-tests to explore within-subject effects in changes in scores.

For each mood questionnaire, changes in scores were computed by subtracting the second score from the previous one so a change was computed for 0 to 10 days and 10 to 20 days. A repeated measures two-way ANOVA was performed on the changes in scores while factoring for the type of image seen and whether this change was after 10 or 20 days to elucidate general effects.

Furthermore, a differential mood questionnaire score was computed which allowed to parametrise how much of the changes in scores were due to seeing smiling faces as attenuated by the effect of seeing landscapes[8]. To calculate this differential score, the change in score after seeing landscapes was subtracted from the change in score after seeing smiling faces for each participant, irrespective of when each participant saw a particular image type.

These analyses were performed separately on each questionnaire type with multiple robust linear regressions while including baseline static measures as confounders variables (device used, geographical location, age, gender, overall life satisfaction, relationship status, overall health status, and baseline questionnaire score). To discern whether a dose response was associated with changes in scores *total images seen* of each image type were also inputted as regressors. A least squares estimates of each model compared the effects of exploring interactions to explore whether this better accounted for the data with a likelihood ratio test. An ANOVA further explored the added value obtained from adding additional variables to the regression models.

[8] Differential Score = (change in score after seeing faces) - (change in score after seeing landscapes)

4.8 Results

4.8.1 Descriptive statistics

This version of Pocket Smile was downloaded a total of 1,240 times on iOS and about 2,000 times on Android between the release of the app in December 2016 and data analysis on February 10th 2017. In total, 1922 people completed the sign up and first mood questionnaire, 509 completed the first 10 days, and 327 completed all 20 days of the experiment. Due to different app malfunctions (e.g. database duplicates or upload bugs), a total of 67 people were removed from the final studied sample. The analysis was therefore performed on 260 individuals (see Table 4.2).

Overall, there were equal numbers of males and females that started participating in both conditions. All users were above 18 however the vast majority were middle aged and based in Europe or the USA. Most participants were working and were either married or had been married at some point in their lives (widowed or divorced). In general, the majority reported being in good health, were just satisfied with their lives, and most users used iPhones.

Table 4.2 Demographics of Pocket Smile 2 participants grouped by image seen in first 10 days.

	Smiling Faces (*n*= 137)	Landscapes (*n*= 123)
Gender		
Male	58 (55%)	47 (45%)
Female	79 (51%)	76 (49%)
Age		
Mean years	45	46
Location		
Europe	82 (53%)	73 (47%)
Americas	33 (52%)	32 (48%)
Asia	17 (50%)	17 (50%)
Australia/Pacific	5 (71%)	2 (29%)
Africa	0 (0%)	1 (100%)
Life Satisfaction		
Mean out of 10	6.7	6.3
Health		
Very good	26 (52%)	24 (48%)
Good	72 (53%)	63 (47%)
Fair	34 (52%)	32 (48%)
Poor	5 (56%)	4 (44%)
Relationship Status		
Single	37 (51%)	36 (49%)
In a relationship	30 (63%)	18 (37%)
Ever-married	70 (50%)	69 (50%)
Occupation		
Student	17 (71%)	7 (30%)
Employed	89 (51%)	87 (49%)
Retired	11 (46%)	13 (34%)
Unemployed	16 (59%)	11 (41%)
Device		
Android	38 (55%)	31 (45%)
iPhone	92 (51%)	88 (49%)
iPad	7 (64%)	4 (36%)

4.8.2 Overall depression and happiness score changes

Irrespective of image seen, at the beginning of the experiment participants reported a mean baseline PHQ-9 score of 10.2 which decreased by 2.4 points after 10 days ($t(258) = -9.6$, $p<0.001$) by 0.9 points ($t(258) = -3.7$, $p<0.001$) after 20 days on a scale of 27 (Figure 4.10). Similarly, across all viewing conditions participants reported a mean baseline SHS score of 3.2 which increased by 0.2 points after 10 days ($t(258) = 3.7$, $p<0.001$) and by 0.3 points after 20 days ($t(258) = 4.3$, $p<0.001$) on a scale of 7.

For those who saw smiling faces for the first 10 days, the mean change in PHQ-9 score was -2.5 compared to -2.4 with those who started seeing landscapes. A two-sample t-test showed there was no significant difference between the two groups ($t(2,260) = 0.02$, $p>0.05$). Within those who began seeing smiling faces, the mean change in PHQ-9 score from day 10 to day 20 was -0.8 compared to -1.0 with those who started seeing landscapes. A two-sample t-test showed there was no significant difference between the two groups ($t(1,261) = 0.27$, $p>0.05$).

A repeated measures two-way ANOVA showed changes in PHQ-9 scores to be greater during the first 10 days ($F(1,271) = 7.94$, $p<0.05$) though there was no effect of image type seen ($F(1,271) = 0.8$, $p>0.05$) or interactions between these variables ($F(1,271) = 0.6$, $p>0.05$).

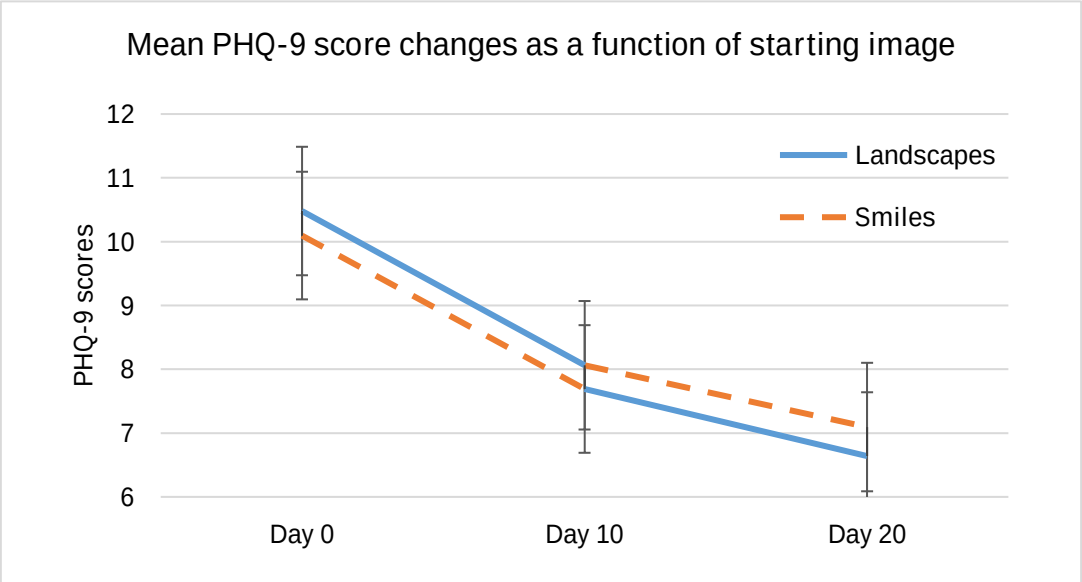


Figure 4.10 Mean PHQ-9 scores plotted with standard error bars showing decreases in time.

A two-sample t-test showed there was no evidence ($t(2, 260) = 0.35$, $p>0.05$) that there was a difference between increases in SHS scores between day 0 and 10 of those

who began the experiment seeing smiling faces (mean change in SHS score of 0.19) or seeing landscapes (0.24). For those who began seeing smiling faces the mean change in SHS score from day 10 to day 20 was 0.17 compared to 0.38 with those who started seeing landscapes. Although this suggests a higher change in score for seeing smiling faces, a two-sample t-test showed there was no difference between these two measures ($t(1, 261) = 1.56$, $p>0.05$; figure 4.11).

Paired samples t-tests revealed that participants had greater decreases in PHQ-9 scores from day 0 to day 10, than day 10 to day 20 irrespective of image type (landscapes: $t(125) = 2.37$, $p<0.05$; smiling faces: $t(137) = 2.38$, $p<0.05$). This effect of time was not found in response to SHS scores where both timepoints demonstrated similar increases in happiness level (landscapes: $t(125) = 0.8$, $p=0.42$; smiling faces: $t(138) = 0.1$, $p=0.91$).

A repeated measures two-way ANOVA showed changes in SHS scores to be equally changed after 10 and 20 days ($F(1,269) = 0.53$, $p>0.05$) and there was also no effect of image type seen ($F(1,269) = 0.06$, $p>0.05$) or interactions between these variables ($F(1,269) = 2.24$, $p>0.05$).

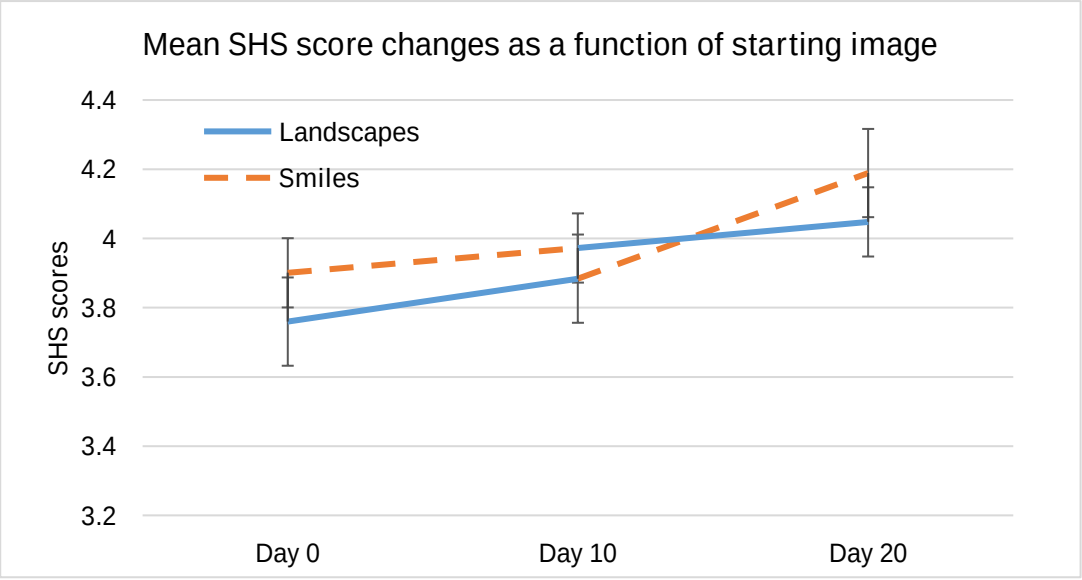


Figure 4.11 Mean SHS scores plotted with standard error bars showing changes in time.

4.8.3 Effect of seeing smiling faces on depression scores

A multiple linear regression was fitted to the differential PHQ-9 score adjusting for total landscapes seen, total smiling faces seen, and the order in which these images were seen (Figure 4.12.A). The model suggested that seeing more smiling faces was associated with a significant decrease in PHQ-9 score ($F(1,255) = 4.12$ $p<0.05$). Specifically, for every 10 smiling faces seen, PHQ-9 score decreased by 0.3 on the 27-point scale. There was no evidence of interactions between the order that the images were seen and the number of faces and landscapes shown ($chi^2(1) = 1.41$, $p= 0.50$).

After controlling for static baseline variables, the model showed that an increase in number of smiling faces seen was associated with reductions in differential PHQ-9 scores, so that PHQ-9 scores decreased -0.18 points for every 10 smiling faces seen. However, this result did not reach statistical significance ($F(1,220) = 1.08$, $p=0.30$). The model also echoed the paired samples t-tests showing a greater decrease in differential PHQ-9 scores after 10 days than after 20 days ($F(1,248) = 13.76$, $p<0.05$). The model also showed that participants with higher PHQ-9 scores at day 0 benefitted less from seeing smiling faces than landscapes ($F(1,222) = 5.10$, $p<0.05$) overall. None of the other covariates reached statistical significance and are therefore not reported. The likelihood ratios test comparing the adjusted and unadjusted models showed both equally explained the data (chi^2 $2(27) = 36.71$, $p=0.10$).

4.8.4 Effect of seeing smiling faces on happiness scores

Similar unadjusted and adjusted multiple linear regressions were fitted to differential SHS scores as with PHQ-9 scores (Figure 4.12.B). The unadjusted model suggested a trend that seeing more faces was associated with an increase in SHS score ($F(1,256) =2.71$, $p=0.10$). Specifically, for every 10 smiling faces seen, SHS scores increased by 0.06 points on the 7-point SHS scale. There was no evidence of interaction between the order of images were seen and the number of faces and landscapes shown ($chi^2(1) = 2.07$, $p=0.36$). When adjusting for the static baseline variables a change of 0.06 points was associated with seeing a dose of 10 smiling faces however it did not reach statistical significance ($F(1,221)=2.52$, $p=0.11$). None of the other covariates reached statistical significance. The likelihood ratios test comparing the adjusted and unadjusted models showed both equally explained the data ($chi^2(1) = 24.24$, $p=0.56$).

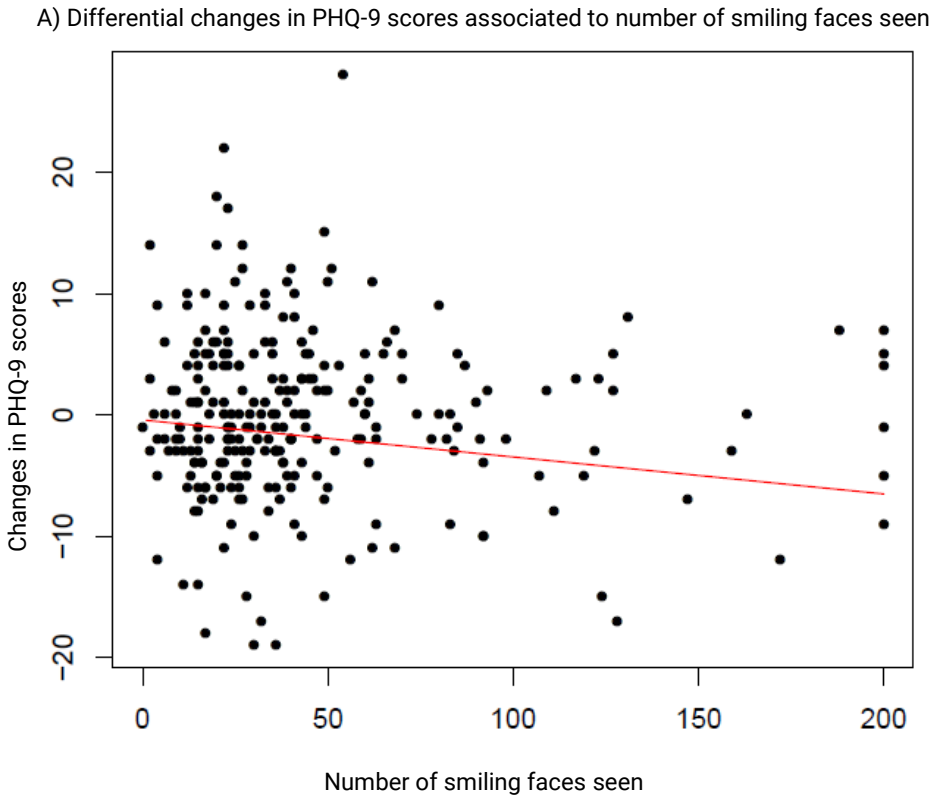
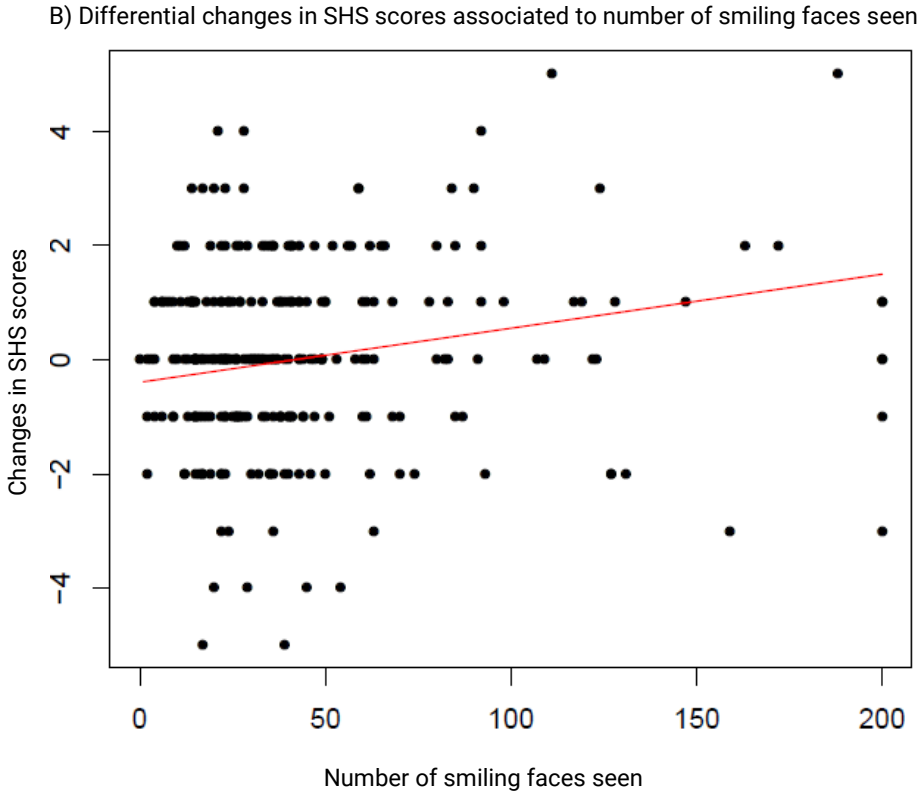


Figure. 4.12. Multiple Regression results from Pocket Smile 2 with least squares lines. (A) Top graph showing differential PHQ-9 scores changes as a function of smiling faces seen. The negative slope of the regression line suggests seeing more smiling faces was associated with decreases in PHQ-9 scores. (B) Bottom graph showing differential SHS scores changing as a function of smiling faces seen. The increasing slope of the regression line suggests seeing more smiling faces was associated with increases in SHS scores.



4.9 Discussion

4.9.1 Implications

The last experiment in this chapter studied the effects of seeing landscapes and smiling faces on mood via smartphones in alternation over 20 days. An independent measure of depression (PHQ-9) and another of happiness (SHS) were obtained at the beginning, before switching viewing conditions at 10 days, and at the end of the 20 days.

Participation in the experiment was associated with a reduction in low mood from a state of *moderate* to *mild* depression as measured with the PHQ-9 (Kroenke et al 2001). This decrease was equal for both image types and found to be most effective over the first 10 days. When looking at the specific contribution of the amount of each image type towards these decreases there was evidence suggesting that smiling faces were effective at lowering depression scores. When also considering the many baseline measures, the effect was still present but less robust and participants who were most depressed reported greater improvements in response to landscapes than smiling faces. With respect to the happiness measure, the population demonstrated well below average SHS scores at the beginning of the experiment (Lyubomirsky and Lepper 1999) and small improvements were recorded. Nevertheless, participation did not influence this measure categorically as participants remained generally below the mean for their age group. Although there was an effect of seeing smiling faces on increasing happiness levels, this was very small and was not robust enough to pass conventional statistical thresholds of significance.

The results therefore suggested that the app was effective at driving positive mood changes and these were associated with the total amount of smiling faces seen. The beneficial effects of seeing smiling faces were demonstrated to be superior to landscapes at decreasing low mood. The scale of the effect was small and statistically not significant. A so-called dose response, where there is a clear relationship between number of faces seen and changes in score was not found, so the exact process to further tailor this as an intervention remains unclear. Nevertheless, these are promising results in the development of a nudge-based intervention based on the neuroscience of facial and emotional contagion for low mood.

The finding that those scoring highest on the PHQ-9 did not react in the same manner as those who reported most benefits to the smiling faces was suggestive of a modulated effect of emotional contagion through facial contagion. It has been shown

that patients with major depressive disorder manifest attentional biases whereby they will look less at happy faces compared to other emotional expressions (Duque and Vázquez 2015). In addition, those scoring higher on depression measures such as Beck's Depression Inventory have been found to demonstrate normal facial contagion for sad expressions but none in response to happy expressions (Sloan et al 2002). The present results suggest that the studied population reacted in line with the more severe spectrum of symptoms. Therefore, those reporting relatively more severe PHQ-9 scores could perhaps have also been affected by more serious depressive symptoms which would lower their sensitivity to the smiling face stimuli.

The Pocket Smile app was conceived to improve low mood which is in line with a mild to moderate score on the PHQ-9 questionnaire. This specific effect is interpreted as further validation for Pocket Smile as a potential intervention for this segment of the depressed population. Indeed, inducing emotional contagion in this low-cost manner may attract patients suffering from severe depressive symptoms and it will be important to ensure targeting of those with *mild* to *moderate* scores to ensure safe use of the app. This finding could help focus the scope of applicability of the app and inform targeting of populations to ensure maximum benefit is gained from usage.

4.9.2 Limitations

Having two image viewing conditions made it evident that the landscapes were a control condition and should, theoretically have less of a positive effect on participants' mood. This generated a problem in terms of retention as demonstrated by the many emails of people complaining to have been allocated to the landscape condition. For example, one user reported: "*I seem to be in a landscapes control group. I have no interest in being excluded from the actual test group.*" Potentially, this could have induced a confound whereby a person's mood was modulated, in line with expectations of the effects of the type of image seen. This is not a major concern as several people also emailed stating a preference for landscape images counter to their own expectations. For example, one participant emailed "*I love your landscapes - so calming and I feel as though I'm there, much better than the faces. They enable me to take a deep breath.*". It is therefore possible that landscapes had a calming effect. This would indicate that landscapes were positive stimuli as well while not relying on the facial contagion effect to induce mood changes. This is suggestive of the mechanism whereby the landscapes induced a positive change in mood and accounts for the small effect size found which was necessary to include to ensure some retention rates.

Despite the many efforts at making the smiling faces more realistic and varied in ethnicities and ages, inevitably, there were a few complaints nonetheless. For example, one participant simply requested to “please make the photos at least somewhat higher quality” while another commented: “So far I've been sent pictures of beautiful people with perfect big smiles and perfect white teeth. […] Nice one. How about pictures of real people?”. Others were more understanding of the experimental constraints but made requests nevertheless: “I know you're trying to get the viewer to focus on the person's facial features, but I don't think blurring out their hair is necessary and just makes it a bit creepy looking instead. It distracts instead of focusing the attention.” App users are generally accustomed to very high standards with this category of products and it should be known it is difficult to satisfy large populations of users when embarking on this kind of experiment.

With respect to representativeness of the population studied, it must be noted that participants originated largely from the European and North American continents. The social media advertising aimed to widen this reach globally and succeeded in recruiting participants mainly from the Pakistan region. However, most participants were recruited following press coverage in *The Telegraph* and on German radio which contributed to this geographical imbalance.

There were also a host of technical issues that delayed the project and contributed to the smaller sample size than expected within the time frame of the study. For example, opting for multi-OS compatibility led to React Native being selected as the most cost-effective platform by the developers. Although React Native allows developing the application once to be subsequently released on multiple platforms across OSs, it also carries more risks of bugs and issues since it is not developed for any device in particular. This led to many more bug reports than if developed only on iOS, multiple database bugs, and lower completion rates on Android as well as significant delays during testing. This could be important in terms of considerations for future app experiments as it might be worth sacrificing sample size for a more reliable user experience and data quality.

4.9.3 Future Directions

The findings that seeing smiling faces can drive positive mood changes was in line with hypothesised expectations. The Pocket Smile app was specifically conceived to improve low mood and was never pitched as a replacement for conventional therapy or applicable to serious cases.

The attenuation of the benefits in the most affected population further validated the findings and will guide the application of this intervention as a potential therapy. This experiment quantified data from peoples' mood as supporting evidence of the potential benefits of this method and in the future, I hope to also be able to assess more qualitative reporting of mood as user feedback was highly insightful.

Experiment 3: A gamified version of Pocket Smile at the Science Museum

4.10 Scope

This experiment expanded upon the previous two experiments in three main ways. First, it tested whether the mechanism that was implemented over the long term could be effective within a shorter time frame. Second, rather than attempting to influence mood, this experiment operated at the emotional affect level – emotional perturbation - an important temporal distinction to identify generalisability of this effect across emotion types. Third, since it was conceived to last under five minutes, debriefing participants became possible. This allowed the introduction of deception so that the effect of awareness of the stimulus could be studied. In other words, whereas previously participants could not be debriefed and were therefore required to know that the manipulation of interest would be on *mood* prior to enrolling into the experiment, in this study, a more general approach could be taken, shifting attention away from mood and only explaining the particular interest in mood once the experiment concluded.

4.11 Methods

4.11.1 Study design

To assess the efficacy of seeing smiling faces in having a positive effect on emotions, a digital video game called *Can Playing Video Games Change You?* was designed and implemented between November 14th 2016 and December 18th 2016 in the Live Science area of the Wellcome Trust sponsored *Who Am I?* gallery of the London Science Museum (Exhibition Rd, London SW7 2DD). The experiment was opened every day of the week except for Thursdays, between 11:00 and 17:30 with a one-hour break at 13:30. The experiment received ethical approval by the UCL Research Ethics Committee under study number 4746/002. This study was a single-blind RCT where participants played a simple game with either smiling faces or landscapes and reported their positive and negative affect before and after playing the game. The experiment was presented as a videogame on brain functions so participants were unaware that there were two categories of images they could play with and that the experiment was about influencing their emotional state.

4.11.2 Materials

4.11.2.1 Emotion Questionnaire

The questionnaire used to assess emotional state was the International Positive and Negative Affect Schedule Shortened Form (I-PANAS-SF; Thompson (2007)). The questionnaire was presented as a list of 10 words that described positive (active, determined, attentive, inspired, alert) and negative (afraid, nervous, upset, hostile, ashamed) affect. Participants were requested to indicate the extent to which each item described the way they felt by answering a Likert-type scale consisting of 5 options ranging from 1) very slightly or not at all, to 2) a little, 3) moderately, 4) quite a bit, and 5) extremely. Since the recruited population at the Science Museum was likely to be international, highly varied in age and in English proficiency, and would be short on time, the I-PANAS-SF was selected as the best option to measure changes in positive and negative affect.

The I-PANAS-SF has been used as a measure of emotion for over a decade in which time it was validated across gender and across 16 countries and different cultures (Karim et al 2011; Agbo 2016). Also, the I-PANAS-SF was validated with the question "How do you feel this way right now, that is, at the present moment" which was important since this experiment lasted only a few minutes. Thus, it is appropriate for brief inter-assessment-intervals as the one presently required (Thompson 2007; Kuesten et al 2014; Agbo 2016). To the best of the author's knowledge this was the first time the I-PANAS-SF was used to assess affective states in a population that included under 18's.

4.11.2.2 Stimuli

A set of 10 genuinely smiling faces of both genders (20 total) and 20 landscapes were selected from the previously created stimuli to ensure participants did not see the same image more than once during the experiment. The same considerations were taken to create and select these images as in the second experiment with the only variation being the cropping and resizing to fit the current experimental parameters.

4.11.3 Participants

Participants were recruited as they walked by the *Who Am I?* gallery of the Science Museum. Advertisements were placed across the Science Museum under the category of "Live Science Demonstrations" with the opening days and times (see Appendix C.1). A total of 4116 healthy participants voluntarily participated in this experiment. All participants were randomly assigned to each of the two stimuli condition (smiling faces or landscapes). No upper age limit was imposed and the UCL REC granted permission to treat individuals aged 12 and above as adults with regards to giving informed consent (i.e. without obtaining prior guardian approval). This was very useful in gathering a large amount of responses as teenagers are usually visiting the museum on their own or can roam freely within the museum space and participation was not onerous.

4.11.4 The game infrastructure

To gather the highest amount of responses the experiment was gamified and ran on 10 Acer Iconia One tablets using an Android operating system.

The tablet application was developed and designed by Fillippo Aiello (http://www.filippoaiello.it/) over the course of 6 months (source code available in Appendix C.3). It was developed as an Adobe Air application using the Starling framework (http://gamua.com/starling/). The final product was a clean and well-designed functional application that stored results in a secure storage database running on a local server at the UCL Institute of Cognitive Neuroscience.

Additionally, a call was made within UCL to recruit volunteers to collect data and engage visitors of the Science Museum in the experiment. After receiving over 70 applications, 11 were interviewed and selected from across post- and under-graduate programs at UCL to form the volunteer team. A team leader was also tasked with supervising the daily activities, scheduling, and management of the volunteers. All members of the experimenter team underwent a Disclosure and Barring Service (DBS)

check and public engagement training prior to the beginning of the experiment as required by all Science Museum staff (see Figure 4.13).

Figure 4.13 The experimenter team at the Science Museum with a tablet.

4.11.5 Procedure[9]

Visitors of the Science Museum that looked age-appropriate were approached by a member of the volunteer team and asked whether they would like to play an experiment that was a short video game. If participants asked more questions, they were told this was an opportunity to engage in scientific research while learning about an aspect of themselves and their brain functions in line with the theme of the *Who am I?* gallery. Participants were therefore naïve to the manipulation that there were two possible sets of images they could be viewing (smiling faces or landscapes) and that the purpose of the experiment was to test the effect of viewing these images on their affective state. This mild deception was necessary to ensure participants' potential changes in mood were not due to expectations that they were in the "experimental" or "control" condition should the manipulation be revealed. If visitors expressed interest in participating, they were given a tablet with a joint information sheet and consent form.

Participants had the opportunity to ask questions and those that wished to participate agreed to the consent form by tapping on the 'I agree' button. Only visitors who confirmed their age to be above 12 completed the joint information sheet and consent form. Those aged between 6 and 12 required a guardian or parent to give

[9] Appendix C.2 provides a screen-by-screen storyboard of the experiment and may be a useful reference for the reader to consult to accompany the next paragraph.

informed consent in their place. Visitors then provided some basic non-identifiable demographic information (gender, age, general health status, relationship status, and occupation) and completed a digital version of the I-PANAS-SF to assess their current emotional state. This served as the baseline emotional assessment.

The rules of the game through which participants were to be exposed to either smiling faces or landscapes was then explained. It was presented as a card game known as “concentration” or “memory” in which a set of cards are placed face down and the goal was to find matching pairs by flipping two at a time. Through trial-and-error, players flipped cards that tended to not be pairs and were forced to remember where each face card is located so that when they find the duplicates, they recalled where the other card was located. Once a pair was found, they remained face up and so on, until all pairs were matched. Visitors then randomly played either with cards that showed smiling faces or landscapes. There were 3 consecutive levels, each increasing in difficulty, from easy (i.e. cards set up in a 4 x 2), to medium (4x3) and hard (6x3). After completing the three levels, participants completed an identical version of the I-PANAS-SF.

After completion of the second questionnaire, a debrief was presented revealing the manipulation and a verbal debrief was provided. Special care was taken to ensure adults were included in a conversation when debriefing children to ensure they understood the implications of the research fully and because this was part of a public engagement effort to bring visitors closer to research. If any errors were experienced during gameplay or participants reported realising the true purpose of the manipulation, their data could be identified and removed from the server.

For both conditions, participants unlocked the possibility to play in a bonus round. Although the rules were the same, the cards displayed both faces and landscapes, making the game easier to play. This allowed participants to learn about the different effects that landscapes and smiling faces could have on their face muscles and try to detect the effects of emotional contagion first-hand. Between 2 and 4 volunteer experimenters were on-site at all times to answer any questions.

4.11.6 Server database

The data from the experiment were saved locally in an XML file on every tablet. In addition, the data were sent over a secure internet connection using PHP 5.6 and stored in a firewalled Apache HTTP Server at the UCL Institute of Cognitive Neuroscience. This solution allowed the data to be saved even when the tablet could not access the internet

and update the database in real time. The data collected in the MySQL database was accessed through phpMyAdmin.

4.11.7 Data Analysis

The XML file containing the data was downloaded from the database into MATLAB where it was analysed to quantify any changes in affective state dependent on stimuli seen.

The main experimental effect involved assessing the effect of *image type* to *change in emotional state*. Scoring of the I-PANAS-SF involved generating a score of all positive affect (PA) items and negative affect (NA) items separately by adding each item scored. For both affect valences, higher scores represented higher levels of the affect in question ranging from 5 to 25 (as there are 5 items for each valence ranging between 1 and 5). Overall this yielded two separate measures of affect (PA and NA).

Since recruitment was performed opportunistically, intial assessments explored whether certain assumptions were met prior to proceeding. These included whether there were equal numbers of participants in both conditions (landscapes and smiling faces), any differences in completion time between the conditions, and whether groups had similar baseline scores. Where these assumptions were upheld the distribution of the data were explored to guide the correct selection of subsequent statistical tests.

Overall changes in I-PANAS-SF scores were explored with a 6-way ANOVA including demographic variables (age, gender, relationship status, health, occupation) and image type as covariates. If main effects proved significant, potential interactions were studied. As the PA and NA scores have 5 distinct dimensions, the granularity within these was explored further with Canonical Correlation Analysis (CCA). CCA finds the optimal linear combination of how different but interrelated attributes relate to each other and can yield insights into how the population changes throughout the experiment (Thompson 2005). All baseline attributes were contrasted to changes in scores and the relationship between these attributes was explored.

4.12 Results

4.12.1 Descriptive statistics

A total of 4116 visitors to the museum completed the experiment. No dropouts were tracked since the application only collected data from completed questionnaires, so it is possible that the total participation may have been higher. Users who were reported by the team to respond randomly were removed from the sample (n=12). Also, participants who were deemed to have responded randomly to the questionnaires as evidenced by all responses being the same across both positive and negative questions of I-PANAS-SF were removed (n=112). The analysis was therefore performed on 3992 individuals (see Table 4.3).

Overall, there were equal numbers of males and females (53.6%) that participated in both conditions. The majority were between 12 and 29 (65%) with most participants being over 18 (69%). Most participants were either working or studying (96%), single or in a relationship (82%), and generally in good or very good health (91%).

Table 4.3 Demographics of Science Museum Participants

	Smiling Faces	Landscapes
Gender		
Male	975 (49%)	935 (47%)
Female	1024 (51%)	1058 (53%)
Age		
6-11	256 (13%)	270 (14%)
12-17	347 (17%)	353 (18%)
18-24	597 (30%)	610 (31%)
25-29	313 (16%)	327 (16%)
30-34	154 (8%)	144 (7%)
35-39	108 (5%)	81 (4%)
40-44	76 (4%)	78 (4%)
45-49	66 (3%)	61 (3%)
50-59	42 (2%)	44 (2%)
60+	40 (2%)	25 (1%)
Health		
Very good	919 (46%)	956 (48%)
Good	887 (44%)	869 (44%)
Fair	162 (8%)	143 (7%)
Poor	17 (1%)	14 (1%)
Prefer not to say	14 (1%)	11 (1%)
Relationship Status		
Single	915 (46%)	926 (46%)
In a relationship	698 (35%)	723 (36%)
Married	327 (16%)	276 (14%)
Widowed	3 (0%)	5 (0%)
Divorced	12 (1%)	13 (1%)
Prefer not to say	44 (2%)	50 (3%)
Occupation		
Student	965 (48%)	991 (50%)
Employed	895 (45%)	858 (43%)
Retired	28 (1%)	35 (2%)
Unemployed	73 (4%)	70 (4%)
Prefer not to say	38 (2%)	39 (2%)

4.12.2 Playing the game

No differences were found in how quickly participants completed the games irrespective of image type as average exposure to images was 2.7 minutes with equal completion times for both smiling faces (*M* = 160.2 seconds; *SD* = 47.2) and landscapes (*M* = 160.6 seconds; *SD* = 47.8); t(3990)= 0.21, *p*= 0.83).

Irrespective of type of image, a two-sample t-test on changes in scores (*post* minus *pre*) to assess the effect of playing the game showed I-PANAS-SF attributes *inspired* (*M* = -0.1), *afraid* (*M* = -0.1), and *nervous* (*M* = -0.1) all decrease significantly (p<0.01) after playing the game. The rest of the I-PANAS-SF items all showed positive change (ranging in 0 to 0.3 in mean of effect size; p<0.05).

4.12.3 Changes in Positive Affect (PA) Scores

Upon inspection, PA scores at baseline appeared normally distributed (Fig 4.14) so a two-sample t-test was used to compare if there were any differences between those that played with different image types. No difference in baseline PA scores were found between participants who played with smiling faces or landscapes, so PA scores were deemed valid to explore further (*M* = 14.9; *SD* = 4.1; t(3990)=1.54 ,*p*> 0.05).

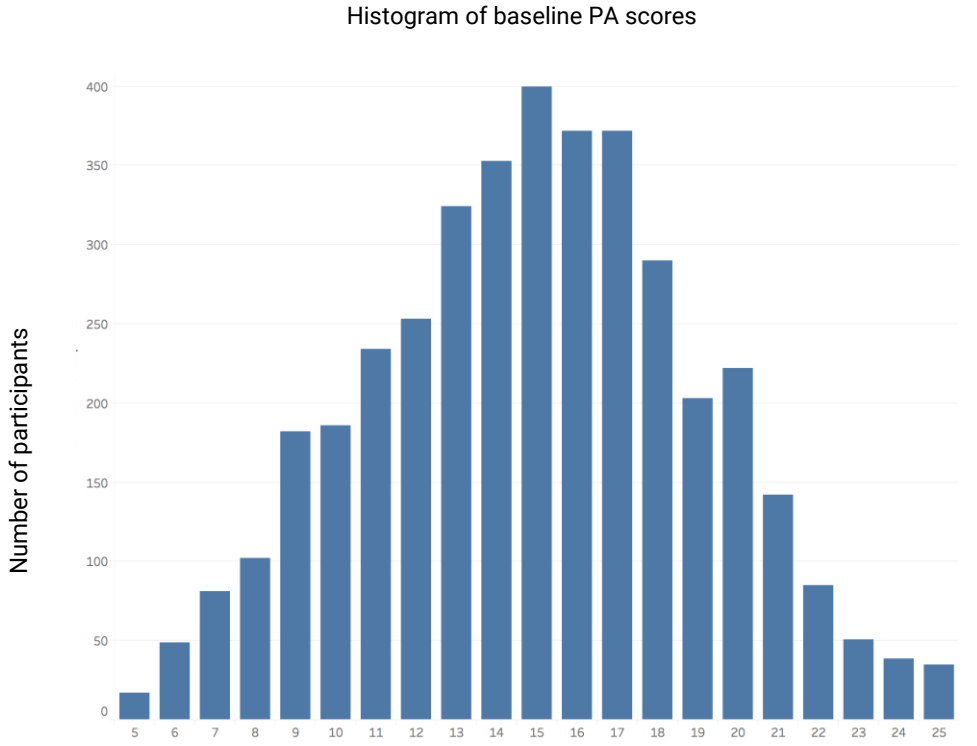


Figure. 4.14 Baseline PA scores of all participants. The scores (x-axis) range from 5 (lowest PA score possible) to 25 (highest PA score) and are normally distributed.

A 6-way ANOVA examining the effect of changes in PA scores (calculated as *post- experiment* minus *pre-experiment* PA scores) was performed with 6 covariates of interest (image type, gender, age, health, occupation, relationship status). This showed that playing the game differently affected PA scores only for image type, gender, and age ($p<0.05$; see Table 4.4). Interactions were explored on the covariates that reached significance on the ANOVA, but none reached statistical significance ($p>0.05$).

Table 4.4 Six-way ANOVA for change in Positive Affect (PA) Scores

Variables	Sum of Squares	Degrees of Freedom	Mean Square	F	Probability values
Image type	39.9	1	39.9	3.92	0.048*
Gender	51.6	1	51.6	5.07	0.024*
Age	199	9	22.1	2.17	0.021*
Health	51.7	4	12.9	1.27	0.27
Relationship status	48.6	5	9.7	0.95	0.44
Occupation	39.1	4	9.8	0.96	0.43
Error	40381.8	3967	10.2		
Total	40878.6	3991			

*significant at $p<0.05$

Exploring significant effects further, showed an increase of 0.93 points on the I-PANAS-SF scale (ranging from 5 to 25) for landscapes and a 0.73 points increase for smiling faces in terms of PA. Further exploration of *Gender* effects showed female participants had higher increase in PA scores to both smiling faces and landscapes than men and that landscapes were more effective at driving positive changes in PA score (Fig 4.15).

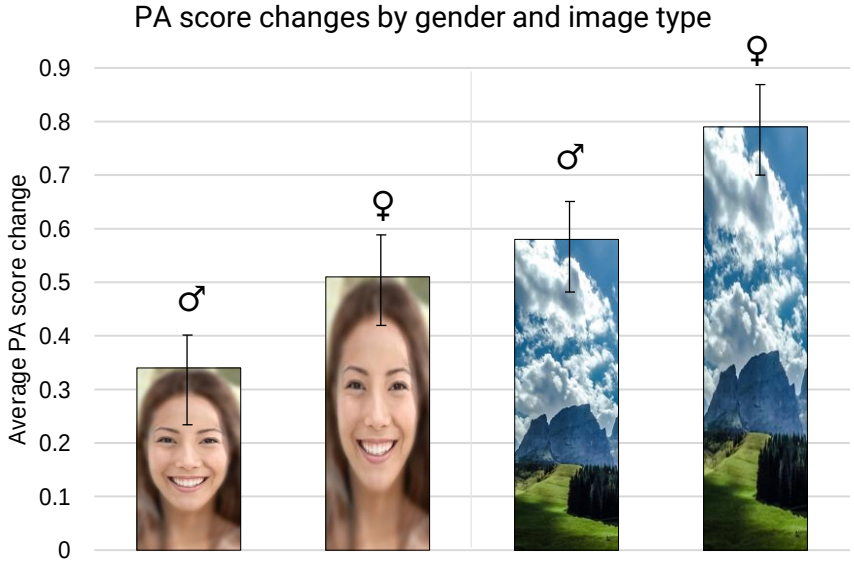


Figure. 4.15 Effect of Gender on changes in PA scores. The images in the bars exemplify the image type and the Mars and Venus symbols denote gender. The figure shows a greater increase in PA scores generally when playing with landscapes and for women.

Segmenting *Age* further showed a general trend of increased effect of the game with older age as bigger changes in scores tended to be associated with increased age (Fig. 4.16). This should be interpreted with caution since there is no strong effect and over 70% of participants were under 30. The only groups to have responded more positively to smiling faces than landscapes were those aged between 25-29 and 50-59.

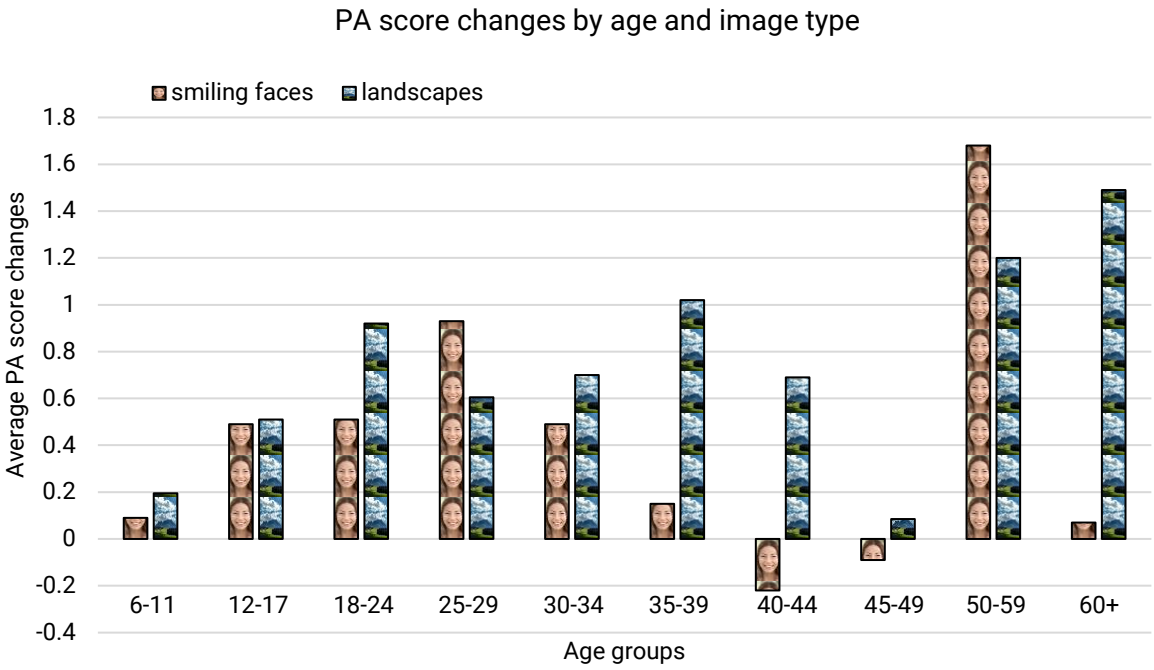


Figure. 4.16 Effect of Age on changes in PA scores. This shows a greater increase in PA scores generally with age although there is no clear trend since most participants were under 30.

A targeted exploration of a sub-sample of the population was performed on those who might benefit the most from the intervention as defined by a low PA score (<=10). This segmentation yielded a sub-sample of 617 individuals with no differences in completion time and equal numbers of participants in the two image type conditions which allowed for further analysis. A 6-way ANOVA with the same covariates as above on changes in PA scores on this sub-sample population showed that none reached statistical significance ($p>0.05$).

In the interest of elucidating the major effects output from the CCA analysis, only the top three components for each linear combination of attributes comparing baseline and change PA score values that reached significance ($p<0.001$) were explored (Fig. 4.17). Interpreting these highest correlated scores tells of different effects present within the data dependent on starting PA scores. For example, component 1 showed that beginning the experiment with low *determined* attribute scores but scoring highly in *alert*

and *inspired* are likely to become more *determined*, and less *inspired* and *alert*. The second component showed that those beginning with high scores in *attentive* but low on *alert* and *active* are likely to score lower after participating in *attentive* but higher on *alert* and *active*. The third most highly correlated and significant component showed beginning lowly *inspired* but very *active* and attentive will tend to lead to higher *inspired* and *determined* scores but lower on *active*.

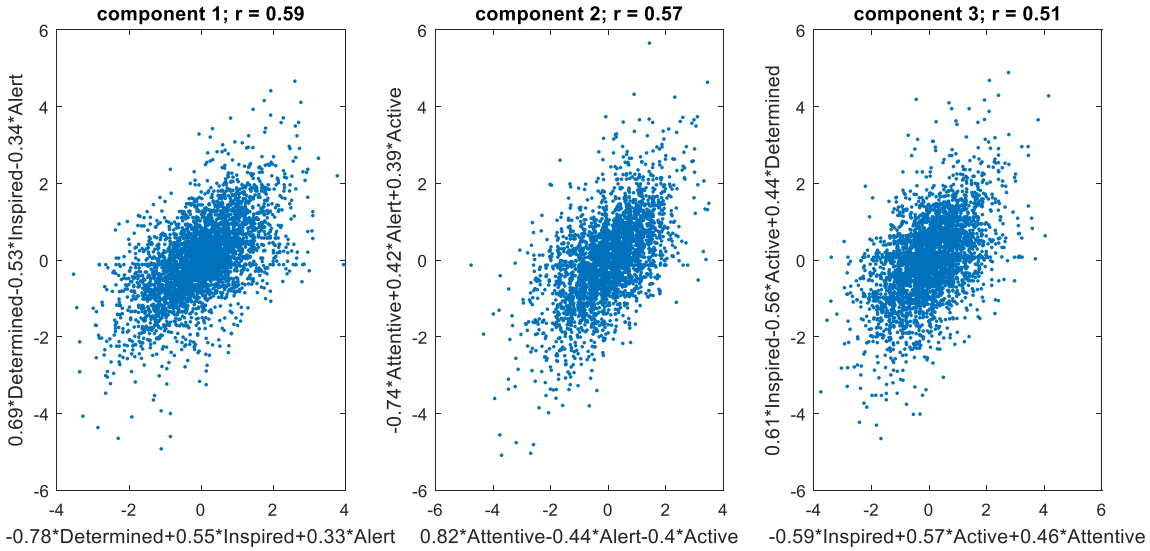


Figure. 4.17 Top 3 components of CCA on PA scores. The top three components (from left to right at p<0.001) from the CCA analysis comparing baseline to changes in PA scores are plotted in this figure. The correlation (*r*) for each component is included above each plot. The x-axis shows the outcomes (change in scores) most likely predicted by the y-axis (baseline) attributes.

4.12.4 Changes in Negative Affect (NA) Scores

Baseline NA scores from the I-PANAS-SF for both conditions were compared prior to proceeding with analyses. NA scores at baseline did not appear normally distributed upon inspection (Fig 4.18) so a Wilcoxon rank sum test was used. This showed no significant difference in baseline scores prior to playing the game between smiling faces and landscapes which allowed the data to be further investigated (*median* = 6; *Inter Quartile Range* = 2); Z = 1.39; *p*= 0.16.

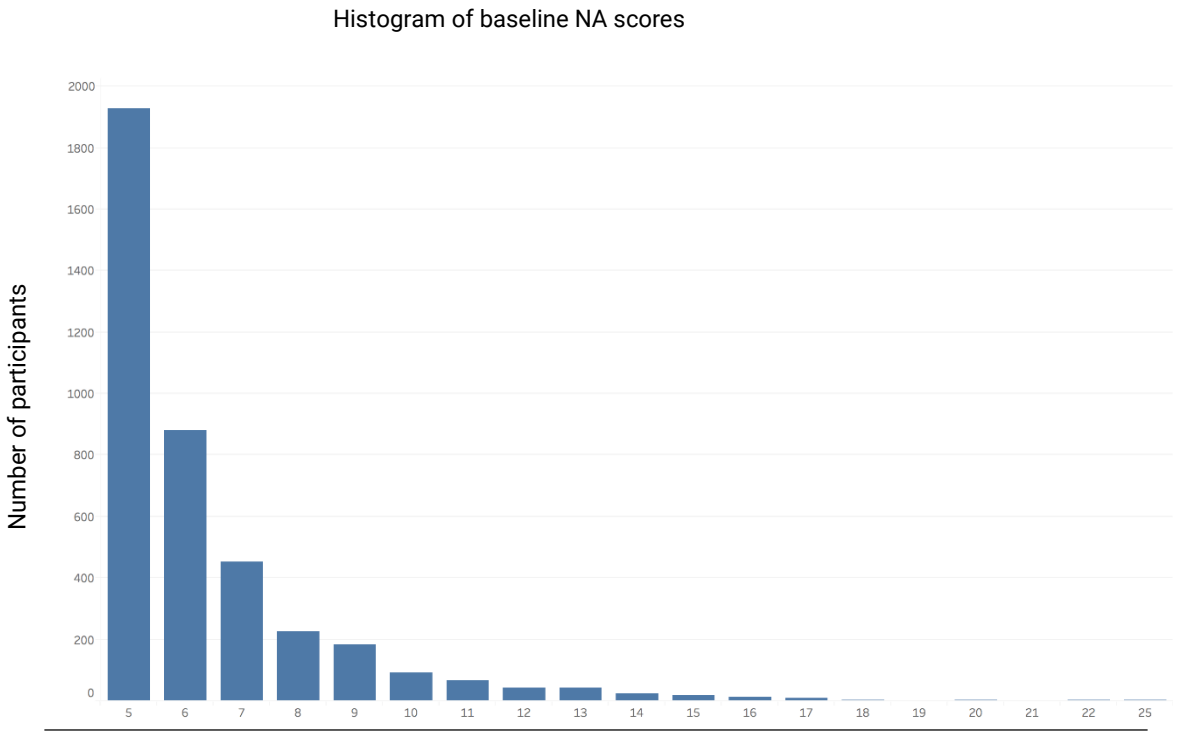


Figure. 4.18 Baseline NA scores of all participants. The scores (x-axis) range from 5 (lowest NA score possible) to 25 (highest NA score) and are skewed in distribution towards the lowest values.

A 6-way ANOVA examining the effect of changes in NA scores (calculated as *post- experiment* minus *pre-experiment* NA scores) was performed with the same 6 covariates of interest as when analysing PA scores (image type, gender, age, health, occupation, relationship status). This showed that playing the game did not differently affect NA scores ($p>0.05$; see Table 4.5). Since no covariates reached significance during these initial analyses no interaction effects were explored.

Table 4.5 Six-way ANOVA for change in Negative Affect (NA) Scores

Variables	Sum of Squares	Degrees of Freedom	Mean Square	F	Probability values
Image type	0	1	0	0	0.99
Gender	0.6	1	0.55	0.15	0.70
Age	63.6	9	1.07	1.87	0.08
Health	30.7	4	7.68	2.03	0.09
Relationship status	27.7	5	5.4	1.47	0.19
Occupation	10.3	4	2.6	0.68	0.61
Error	14978.1	3967	3.8		
Total	15138.7	3991			

As over 60% scored *a little* or *not at all* in the overall NA score a targeted exploration of those who might benefit the most from the intervention was performed. This was defined by a high NA score (>=10) which yielded a sub-sample of 321 individuals. No differences in scores were found between number of participants in each of the two image conditions and between completion time which allowed further investigations. Studying this sub-sample with a 6-way ANOVA with the same covariates as above on changes in NA scores showed none reached statistical significance ($p>0.05$).

Experiment 3: A gamified version of Pocket Smile at the Science Museum

A CCA analysis was performed in a similar fashion as that on PA scores and the top three components for each linear combination of attributes comparing baseline and changes in NA scores ($p<0.001$) were explored (Fig. 4.19). Component 1 showed that beginning with high scores in *afraid* and *ashamed* but low *nervous* has a good chance ($r=0.68$) of leading to feeling more *nervous* but less *afraid* and *ashamed*. The second component showed that those that begin less *afraid* but feeling more *ashamed* and *hostile* will likely feel more *afraid* and a bit less *ashamed* and *nervous*. The third component of the CCA showed that beginning with high scores on *ashamed* but very low in *upset* and low in *nervous* will lead to a significant reduction after playing the game in feeling *ashamed* and *afraid* but higher levels of *upset*.

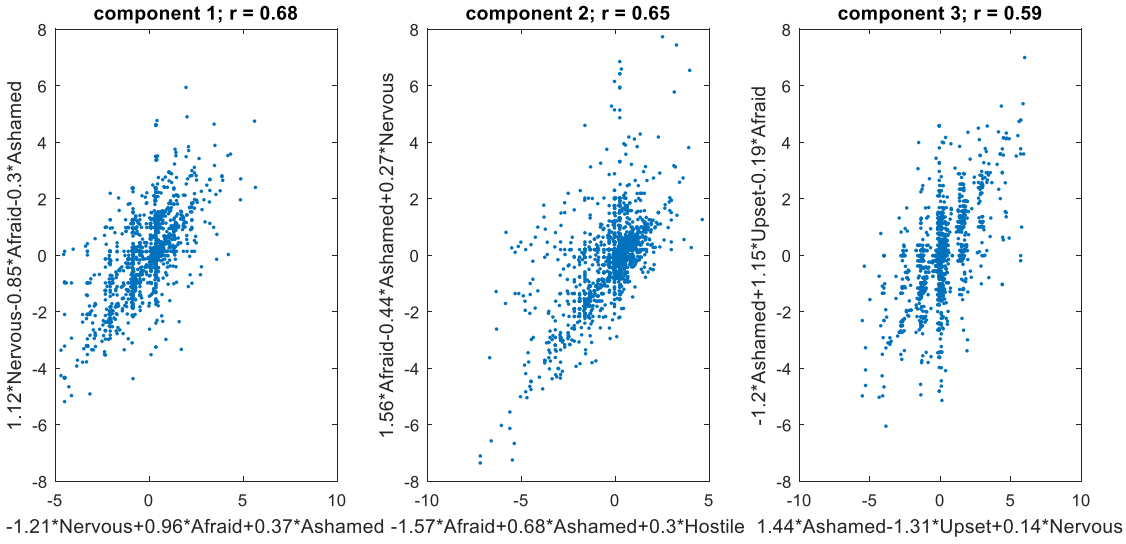


Figure. 4.19 Top 3 components of CCA on NA scores. The top three components (from left to right at $p<0.001$) from the CCA analysis comparing baseline to changes in NA scores are plotted in this figure. The correlation (*r*) for each component is included above each plot. The y-axis shows the outcomes (change in scores) most likely predicted by the x-axis (baseline) attributes.

4.13 Discussion

4.13.1 Conclusions

This experiment explored the effects of seeing smiling faces and landscapes on affective state while playing a short video-game. Changes in affect measured with the I-PANAS-SF showed that after playing the game there was an improvement in PA and no changes in NA irrespective of image type. This was an original demonstration of a short intervention (160 seconds) to have a measurable positive effect on emotional state. Contrary to the translational hypothesis based on the literature of facial mimicry and emotional contagion landscapes seemed to drive changes in PA scores more than smiling faces. The effect itself was very small as the change was less than 4% increase in PA on the I-PANAS-SF scale and the differences between conditions was equivalent to 0.8% (0.2 on a scale of 25). Therefore, although landscapes were more effective at increasing PA than smiling faces this difference is likely not clinically relevant and the same can be stated about the overall effect.

There was also no consensus among anecdotal accounts of experimenters who reported participant impressions on difficulty between playing with smiling faces and landscapes. Statistically there were no differences in completion times between the games and the groups were equally distributed in terms of demographic variables including age and gender. Indeed, though the effect sizes of the PA changes were statistically significant they were very small which demonstrated the requirement of conducting experiments that require big datasets to elucidate small effects. The design of the experimental platform and randomised assignment of participants was therefore sound and supports the validity of the results.

Since participants were visiting the Science Museum, one of London's most popular attractions, many unsurprisingly began the experiment with little to no NA as measured by the I-PANAS-SF. Considering the small effect sizes there was no room for improvement in this population with this scale in terms of NA. Nevertheless, the targeted analysis on sub-samples that were profiled as most likely to benefit from this intervention (i.e. high NA and low PA) did not produce any significant effects of change. It is possible this is due to the small effect size found with thousands of participants to not be revealed when reducing the sample to the hundreds. However, it is also possible that the intervention was not effective beyond positively affecting individuals in the average emotional spectrum of a healthy population. Indeed, as the experiment did not

formally exclude or measure the presence of serious mental and emotional disorders and some participants did exhibit disabilities it is possible these participants scored higher in the I-PANAS-SF and were unaffected by the intervention to the same extent as others.

This was the first time the I-PANAS-SF was used to assess affective states in a population that included children. Although very young children required parents, guardians, or the experimenter to explain the terms and close supervision older children and teens found it very easy to complete. There was no difference between adults and children in general and the trend of an increase in PA score with age was also very weak. Interestingly, those between 25 and 29 and 50 and 59 responded more positively to smiling faces than landscapes and could potentially constitute a target group for future research into this intervention. Since on average those between 40 and 49 demonstrated a decrease in PA after seeing smiling faces it would be wise explore in more detail the differences between such age groups. The exact reasons behind these unclear changes linked to the age of participants however remain mysterious and could perhaps best be elucidated with a post-experiment qualitative interview.

Another interesting finding was that female participants tended to respond better than men irrespective of image type. In Western culture it is an established and pervasive stereotype that women tend to be more emotionally expressive than men in general and in particular to happiness (Fischer 1993; Hess et al 2000). Although large differences are reported between men and women in the actual amount they are observed to *express* negative emotions (Plant et al 2000; Durik et al 2006), a recent meta-analysis found no gender differences in the degree to which men and women report *feeling* these emotions (Else-Quest et al 2012). This finding was supported by research in which participants record their emotions in a daily diary and found men and women report feeling the same types and amount of emotion as one another (Barrett et al 1998). Another noteworthy study measured the degree of facial and emotional contagion experienced by men and women in response to watching jubilant sport winners or heartbroken losers and also found equal degree of contagion susceptibility between genders (Arakawa 2012). Therefore, despite stereotypes, the current literature suggests men and women do not differ in the extent they experience emotions but rather just in the extent to which they express these outwardly. In other words, if women are more emotionally expressive than men it is possible they may also be more likely to change or report more changes in affect and could potentially be more receptive to interventions such as the present one.

The CCA analysis provided interesting insights into the different changes that occurred in PA and NA following playing the game. Some of the results can be attributable as reflecting regression to the mean as starting with particularly low or high scores at baseline will tend to shift scores towards the opposite direction. This is most evident in all changes in attributes in components 1 and 2 for PA and component 1 for NA.

For example, the third component of the PA CCA analysis showed that those starting out very *attentive* and *active* but with low *inspiration* would decrease in *active* while having higher *inspired* and *determined* scores after playing the game. This could be indicative of the different kind of experience that the experiment provided to the rest of the exhibit. Compared to the rest of the *Who am I?* gallery, this was the only one where visitors could sit rather than stay standing (linked to a decrease in *active*) but were provided with a game which many people took as a challenge as it progressed in difficulty (higher *determined*) while feeling like they had achieved something when completed (higher *inspired*).

For the NA CCA analysis the second component suggested that some participants started out feeling more *hostile* and then felt less *nervous* by the end. This is perhaps showing those that were a bit resistant to the game and feared worst perhaps than what it was. The third component, showing that low nervousness could lead to reductions in *afraid* is probably due to the removal of uncertainty as the game is finished.

Overall, this analysis remains speculative, though it would be possible to further subdivide the CCA analysis according to image type and explore if and how participants differed in changes to their emotional state. In the interest of time and brevity however this was opted against as the effects were already very small and strongly speculative.

4.13.2 Limitations

Although the experiment showed that smiling faces increased PA state, the fact that landscapes were more effective at driving this effect remains surprising. There are two potential explanations for this finding, the first related to considerations surrounding smiling faces as stimuli to induce emotional contagion and the second involving landscapes as neutral stimuli.

Though facial mimicry may occur automatically and subconsciously (Surakka and Hietanen 1998; Dimberg et al 2000; Soussignan 2002) emotional contagion could require more prolonged exposure to these images in order to exert an effect on emotions

more strongly. Also as static images are known to induce less facial mimicry than dynamic images (Weyers et al 2006; Sato and Yoshikawa 2007; Rymarczyk et al 2011), perhaps switching the stimuli to images of neutral faces turning into smiles could also increase the power of these images.

It is possible that landscapes were underestimated as a neutral stimulus. Evolutionary analyses of environmental aesthetics suggests that landscapes may arouse people's interest and curiosity which could in turn be positively uplifting (Kaplan 1992; Orians and Heerwagen 1992). Landscapes can also lead to the retrieval of positive autobiographical memories and can even influence the extent to which one is susceptible to such imagery in adverts (Hartmann et al 2016). Therefore, it is plausible that landscapes operated as potent positive stimuli through another automatic biological affective path unrelated to emotional contagion.

Nevertheless, wherever changes with the stimuli could be possible, the effects of changes were still relatively small. It is possible that there is little scope for greater change in emotions over an average period of 160 seconds. However, the game itself may have attenuated this effect. Although the images were consciously perceived, participants were unaware of the manipulation of *image type*. As such the video-game was a way of exploring the effect of exposures to different image types and it is possible that it was confounded by the distractor task of playing the game which could have been too mentally demanding. The game required mental efforts in memory and many people have biases as to their expected performance in such games. It is therefore possible this could also have affected their emotional state negatively.

4.13.3 Future Directions

Overall the experiment was effective at driving positive changes in affect, however the translational hypothesis that emotional contagion could be induced significantly in this short time period did not hold. Therefore, the potential for this technique to be employed in the short term should not be exaggerated.

There are still some questions that remain unanswered including whether neutral faces could have been less effective at driving the change and whether people might find this useful as a conscious boost to their mood if the intervention was not subconscious.

4.14 Chapter discussion

This chapter presented three experiments undertaken with similar goals: to explore whether the established emotional contagion effect could be translated into an applied intervention that presented smiling faces to enhance emotional states.

The first Pocket Smile experiment established that mood can increase over long periods of time (up to 6 months) when seeing smiling faces. The second Pocket Smile experiment demonstrated that such changes to mood are correlated with the total number of smiling faces seen. This experiment also established that smiling faces are more effective at driving positive change compared to another positive image type that lacked the facial contagion component despite being considered positive stimuli in the literature; landscapes. The third experiment found that emotional contagion could not be induced effectively at a shorter time scale within a video game and without participants' awareness. This last experiment was informative in determining the limitations of the translated effect. It is therefore advisable to attempt to change mood rather than short term affect with this method.

These findings form a series of novel findings that contribute to the field while also defining the target of most effectiveness of this method. When summarised, the recommendation that could form the basis for a prediction on devising a most effective method of enhancing positive state would advise to aim to modify long term mood in a consciuos manner rather than short term emotions unconsciously. Although these findings limit the scope of this method as a potential intervention, it also specifies it and allows the opportunity to test further where it could have higher impact. From a theoretical standpoint, this series of experiments could not be considered to support the "strong" version of facial feedback[10] as emotional contagion was found to increase mood but not transform a negative state into a positive one. However, it should be noted that it is problematic to conclude in a study that on mood or emotional state that "no positive emotional state was present in participants prior to participation in the current study". Indeed, this would be a necessary requirement to make the claim in favour of the "strong" version of the FFH albeit although it would be impossible to empirically demonstrate a total absence of an emotional state (i.e. "I *never* felt that emotion in my life before") that is only felt after participation in a scientific study.

[10] The reader may recall from Chapter 1, that the "strong" FFH considers it could be possible to induce an emotion where none existed while the "weak" version contents itself with positing an emotion can be attenuated only when one was already present (Schneider 2008).

Experiment 3: A gamified version of Pocket Smile at the Science Museum

A main component of this chapter involved leveraging mobile digital technology to conduct research. The ubiquity of mobile technology allowed the collection of data from a much broader and larger sample than could conceivably be collected in the lab. In total, 7,780 participants were engaged in these experiments. Even when only considering the Pocket Smile 2 experiment, participants completed over 8,360 days of experiment, equivalent to over 22 years of participation. These technologies offer vast potential for scientific research and several recommendations have been made throughout this chapter that could aid future researchers in utilising it successfully. Important recommendations not previously stated but very important include budgeting for marketing, embedding desired social media SDKs within the code of the app for intelligent tracking of advert effectiveness, careful design of the database for ease of analysing the data, creating accurate requirements for the formating of the data collected, and having an open channel of communication with users so that an agile development process may benefit from feedback to drive participation.

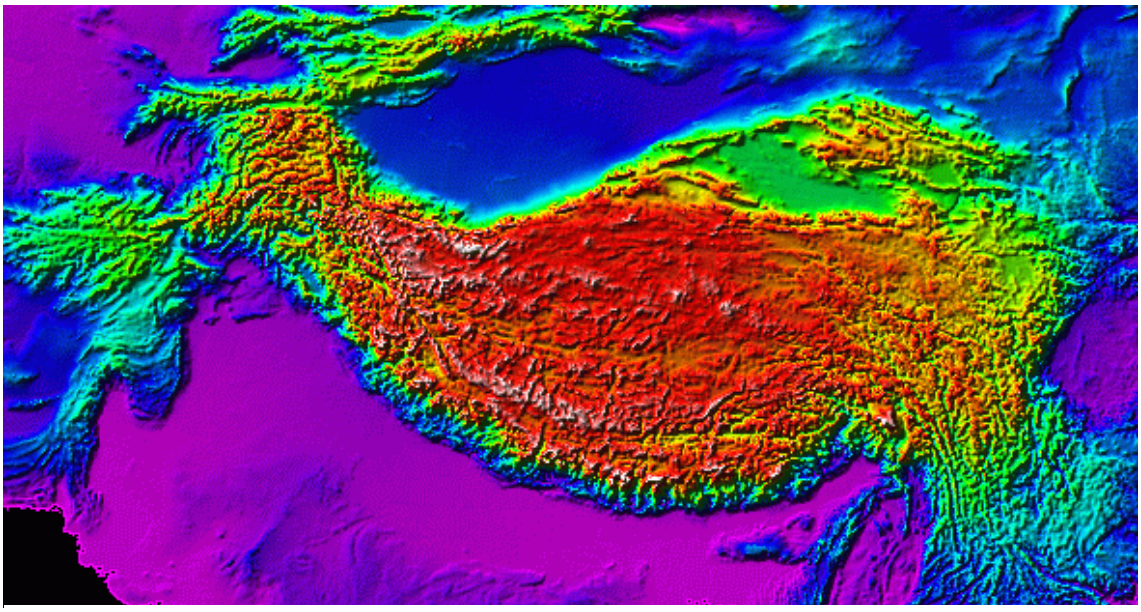
Overall, the Pocket Smile experiments demonstrated the effectiveness of using smiling faces as main components of a nudge-based intervention for low mood. The findings of the final experiment suggest it was not possible to generalise this effect in the short term. Therefore it is most appropriate to pursue the usefulness of facial contagion to improve low mood rather than momentary changes in affect. This consists of an original and valuable contribution to those who aim to modify mood positively and takes the field one step closer to a better devised intervention.

Together these experiments addressed the questions posed in the introduction as it appears to be possible to translate the neuroscientific principles of facial contagion into nudge-based interventions and are associated with lasting positive changes in mood. As a nudge-based intervention, Pocket Smile could be conceived as a complementary treatment during officially recommended therapy. If this form of intervention is found to be effective once validated with a clinically depressed population, it could hold potential to reduce the cost of health care, close the treatment gap in accessing support, and improve other health outcomes in numerous ways (Kumar et al 2013). With further research into implementation and validation with a depressed population, Pocket Smile could be administered during referral waiting times for individuals for which there is no currently viable alternative to address a national healthcare system weakness and priority (NHS England 2015).

Alternatively, the next steps in the translational process could see these findings integrated within a commercialised version of the app. For example, a social network

based on sharing smiles from loved ones could be conceived based on these findings, inspired by several comments and feedback such as this user's: "*I guess it's not like in real life when there is a person you know and care about and it's always good to see them smile.*". The Pocket Smile app could be modified into a social network consisting of people selecting or creating pictures of themselves smiling and uploading these to customisable groups of friends and family members. The app would retain its current functionalities of delivering the images throughout the day in a controlled manner and could be expanded to include groups based on cities, countries, and general interests. This would also have the added benefit of personalising the *Social* element of the EAST framework to a greater extent.

Chapter 5: Wearing your heart on your face: Predicting physiology from physiognomy with machine learning



Elevation map of Tibetan plateau where the average height is 5000m. Image from Arizona University

5.1 Introduction

5.1.1 Shifting from function to structure

In the preceding chapters, the correlations between functional features of the face and physiological variables of interest were explored. In one case, the physiology of emotional lability in MND patients was studied to assess its potential as a diagnostic marker (Chapter 3). In the interventional setting, attempts were made at influencing affect and mood through the induction of a functional effect - emotional contagion - accompanying a physical change in the face (Chapter 4).

Here the consequences of structural change to the face secondary to changes in function will be explored. The physical parameters of faces will be considered as potentially informative variables that might predict underlying physiological markers of interest, with translational utility in a clinical setting and beyond. This project is in line with the PhD's overall mission because it consists of the same translational problem: the intrinsic high dimensionality of the biological entity that is the face. It is clear that people can extract useful information from a face including identity (McKone et al 2009), ethnicity (Tanaka et al 2004), gender (O'Toole et al 1998), age (Rhodes and Anastasi 2012), emotional state (Ekman et al 1972), health (Jones et al 2012), and sexual orientation (Rule and Ambady 2008). If there is any connection between the physiology of the face and other physical markers, it is suspected that it will not be reducible to a specific parameter of any single kind. Rather, in line with the high-dimensional approach, capturing the rich patterns of the face by integrating across many variables through machine learning methods is expected to enrich outcomes in terms of power and generality.

5.1.2 The structure of the face and clinical outcomes

The connection between changes in structure of the face and related clinical measurements has been an established topic of study in several interesting contexts. The way in which the face develops is dependent on the way other parts of the body develop (Björk 1955). The structure of the face has a complex architecture and can be informative in many contexts.

For example, analysing the state of the skin can provide information associated with ageing (Rhodes 2009). The surface of the skin has a slightly acidic film, measured in terms of pH, that protects from potential contaminants that might penetrate it such

as bacteria and viruses. Marrakchi and Maibach (2007) found older people exhibited higher pH values on their facial skin surface than younger people. This difference however was not present when comparing pH levels on the forearm skin surface, suggesting the face skin surface might demonstrate changes associated with ageing.

Wrinkles are also known to be informative when estimating age, with a greater number of wrinkles associated with older age. Since wrinkling the skin is necessary to create facial expressions at all ages, over time, these become more pronounced. This is attributed to a decrease in elasticity of the skin which gathers sheaves of fibres in the points of most contraction, making the skin less homogenous (Wu et al 1995). Other skin cues that change with age include a reduction in and greyer hair, thicker eyebrows, and changing in the relative size and shape of the lips (George and Hole 2000). Additionally, expression of certain genes have been associated with being perceived as older (Gao et al 2016; Liu et al 2016). Considering that facial age estimates are associated with likelihood of mortality (Gunn et al 2015) it is valid to look at the face as a potentially informative clinical measure.

Insights about the structure of the face paired with machine learning methods can predict measures such as age with over 90% accuracy (Jagtap and Kokare 2016). For example, in the context of congenital diseases, 30-40% of genetic disorders will be associated with a structural facial abnormality (Hart and Hart 2009). By treating the face as a high-dimensional set of variables, Ferry et al (2014) trained an algorithm to identify patterns of facial dysmorphia with 91 genetic disorders. The model automatically extracted facial features from photographs and compared the points for similarity to average faces of syndromes. This system provided diagnostic information to identify rare genetic disorders in children and could be used to further tailor treatments and screening for further research on genome sequencing.

Furthermore Obafemi-Ajayi et al (2015) compared face images of healthy and autistic children and found that their machine learning technique could correctly diagnose those with the disorder. Even though the investigation included a relatively small clinical sample (N= 62), three sub-clusters were reliably identifiable that were associated with diagnostically meaningful behavioural differences. Similarly, Williamson et al (2014) parametrised facial and vocal responses of individuals that self-identified as depressed and found that the timing of motor movements involved in creating facial expressions could be used as information in predicting severity of depression.

Thus, applying machine learning methods to facial features can be informative when properly parameterised and predict high-level clinical measures.

5.1.3 The human face at high altitude

This chapter will establish whether changes in facial structure of humans living at altitude can be parameterised from photographs and used to predict biological markers of clinical relevance.

Living in high-altitude regions involves adapting to severe environmental challenges for human populations including the low availability of oxygen. It is common for residents of towns at elevations exceeding 4000 meters such as in Peru and Tibet to experience oxygen concentrations in the air around 40% lower than at sea level. When living in such conditions, multigenerational residents develop evolutionary high altitude adaptations that allow them to thrive in a hypoxic environment (Yi et al 2010). Comparisons of native Tibet populations who have lived at altitude for approximately 25,000 years compared to Han populations that have entered the Tibetan plateau relatively recently (<70 years ago) demonstrate these biological adaptations. Ethnic Tibetans exhibit higher arterial oxygen saturation at birth and during the first four months of life compared to Han infants (Niermeyer et al 1995) while a comparison of adults demonstrates superior oxygen levels while exercising (Zhuang et al 1996).

For these adaptations to occur it is necessary for the body to carry more oxygen which requires producing more haemoglobin and may in turn lead to the blood becoming thicker. This can be measured by haematocrit (HCT) levels which reflect red blood cells percentage in the blood and are commonly measured in sporting competitions to test for doping (Böning et al 2011). To produce higher HCT levels more haemoglobin needs to be produced. Since haemoglobin is produced in the bone marrow it is expected that bones storing bone marrow will expand accordingly to accommodate for the larger capacity. This is important because the face is composed of a bone marrow producing bone: the maxilla. The maxilla is an upper jaw bone that holds the upper teeth to below the eyes and attaches to the zygomatic bone (Figure 5.1). In some areas of the maxilla, bone marrow composes up to 40% of the tissue (Lindhe et al 2013). It is hypothesised that in populations that adapted to low-oxygen high-altitude environments there will be expansion in the volume of this bone marrow.

This expansion was documented in patients with major Thalassemia, a genetic disorder that severely affects production of haemoglobin. As their condition leads to haemolytic anaemia, where red blood cells die prematurely, the cheek bones expand under pressure from the bone marrow cavity becoming enlarged in an attempt to continuously produce more haemoglobin (Rund et al 1997; Phadke and Agarwal 2003; George 2013). It is not expected that living at high altitude can have such an extreme effect on physiognomy as Thalassemia; however, the haemoglobin dysregulation being similar, the effects may still be present and measurable over the face. The current project will explore whether similar facial bone abnormalities due to living in lower oxygen environments is measurable through parametrisation of faces and whether it can be clinically useful to predict the underlying physiology with machine learning methods.

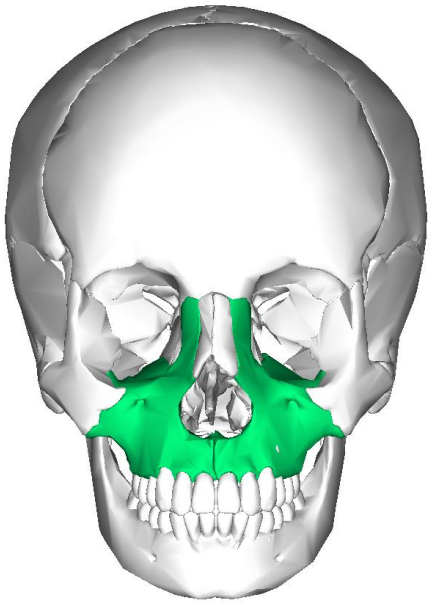


Figure 5.1. Depiction of human skull with the maxilla in green.

5.1.4 Translational utility

Although augmenting haemoglobin concentration may improve aerobic performance it may also predispose individuals to higher risk of cardiovascular diseases (CVD), particularly pulmonary hypertension (Barer et al 1983; Schreier et al 2014) and lead to mortality (Sorlie et al 1981; Gagnon et al 1994). It may therefore be useful to have an effective way of assessing risk of CVD caused by hypoxic environments based on cheap superficial measures such as photographs. If this experiment confirms the hypothesis that the effects of maxillary expansion on the face are measurably related to HCT or other cardiovascular measures it may hold a range of translational applications.

If effective, the final product could integrate the functioning algorithm into a digital pipeline that would consist of 3 steps to perform a clinical recommendation. First, photographs of front-facing faces would be taken and imported into the pipeline. Second, the images would be processed according to defined criteria and analysed by the SVM algorithm (see Chapter 1 for introduction to SVMs). Finally, this would generate an automatic result detailing the likelihoods of different CVD risk. This could guide diagnostics and focus attention on individuals at highest risk. Measuring the effects of maxillary expansion on the face currently necessitates a computed tomography scan which is cumbersome, expensive, and difficult if not impossible to transport, install, and maintain in small towns at high altitude. This solution would necessitate only an internet connection while the processing would be done offsite. The possibility of associating different known biological data with the images would also be useful in expanding the training set and increasing the system's performance.

If successful, this could be used as a screening tool for researchers and medical teams wishing to treat or investigate people at risk of pulmonary hypertension. This would be of obvious advantage for native populations, as living at high altitude is a strong risk factor for pulmonary hypertension, measurable by echocardiogram. Currently, there is no way of screening for individuals who would benefit from receiving an echocardiogram and very few medical teams have taken these machines to high altitude due to their size and weight. For example, Allemann et al (2000) and Sartori et al (2002) separately acquired echocardiographic measures from different individuals in the Italian Alps. This required the hiring of mountain guides to lead the scientific teams to the location safely as well as assistance from the Swiss Army to transport the machine to height. With current technology, it would be very simple to take a quick photograph of someone's face to act as a screening diagnostic for which individuals could benefit from

descending to their nearest hospital or for pre-screening purposes for inclusion in clinical trials.

Additionally, it could be used as a developmental tracking measure that would be of academic and clinical interest in terms of studying the developmental changes of facial morphology over time. It will also be interesting to explore how the expansion of the maxilla will force reconfiguration of the rest of the face. This is something that remains unknown, however, it is possible that during this investigation other features of the face might prove to contribute to biological markers of interest and drive further research.

5.1.5 Scope

This chapter will explore whether the effects of maxillary expansion on the face can be measured by parametrising photographs of front facing individuals from Tibetan and Han Chinese populations living at altitude. It will also explore if predictable physiological changes with clinical utility can be predicted from face parameters and vice versa. Further exploratory questions will include assessing whether machine learning methods can be used to predict acquired biological measures that could be predictive of CVD (e.g. echocardiographic or haematological measures such as Oxygen Saturation (SaO2) levels), whether more expansion is associated with more time spent at altitude, and whether relative redness of the face (i.e. blushing) can be an indicative measure of abnormal cardiovascular markers.

5.2 Methods

5.2.1 Tibetan Dataset

This dataset was obtained from an ongoing study led by Lan Zhao and Martin Wilkins at Imperial College London who travelled to Tibet and collected physiological data on 9000+ local individuals. They performed this research to explore the effects of living at high altitude on CVD and obtained ethics approval from the Imperial College Research Ethics Committee. The research team also took front facing photographs from participants and 980 of these were obtained for the present study.

For each participant, the following demographic information was recorded: age, gender, height, weight, body mass index (BMI), ethnic origin (Tibetan or Han Chinese),

and how many years they had lived at altitude in Tibet. A measure of proportion of *Time* spent at height *in Tibet* (TiT) was generated by dividing the age by the length of time spent at altitude. Several haematological and echocardiographic measures were obtained including levels of HCT, oxygen saturation of haemoglobin (SaO2), systolic (SBP) and diastolic blood pressure (DBP) and heart rate.

The echocardiographic measures (in *italics* below) provided a detailed overview of heart function and likelihood to experience CVD, especially pulmonary hypertension where the right ventricle (RV), responsible for pumping de-oxygenated blood from the heart to the lungs, may fail (Danchin et al 1987; Voelkel et al 2006). Specifically, *Tricuspid Annular Plane Systolic Excursion (TAPSE)* is a measure of overall RV function. As the pressure in the pulmonary vascular bed rises, it is propagated back to the right ventricle, raising the RV pressure (RVP). The RVP *gradient* should measure the resistance found within the heart chamber. The *RV systolic pressure* indicates the pressure within the pulmonary artery and as the chamber contracts, the *RV area* measures dilatation of the chamber resulting from decompensation in its ability to pump against the raised pressure. The *RV anterior wall thickness* measures the thickness of the muscle which hypertrophies as the heart tries to work harder initially to overcome the pressure. Another measure of RV function is the *RV fractional area change* which indicates how well the heart is contracting as a function of its overall volume. A dilated heart might not contract much as a ratio of its overall size. When the heart becomes enlarged, the tricuspid valve starts to leak, yielding a measurable *tricuspid regurgitation velocity*.

5.2.2 Whole-face feature extraction and normalisation

To predict the physiological variables of interest from the faces, the 980 photographs underwent different parametrisation procedures that captured the high dimensionality of the faces major landmarks coordinates.

First, all images were loaded into MATLAB and adjusted for quality which involved ensuring the images were faces, rotating the images, and equalising the lighting and contrast where needed. The images were then cropped to the smallest square pixel area possible around the face to minimise the information load required to process each image. An automatic feature detection algorithm parametrised 83 points on each face. These points were extracted and used to create a rigid transformation of each image by using the intraocular distance as anchor to normalise the faces. This led to comparability between the faces as they were now equally aligned over the eyes and created a mean image to use as template for further processing. To re-align the images with the rest of

the face features (size, rotation, tilt), a nonlinear transformation was performed on all the rigidly aligned faces again by repeating the face detection and extraction steps to the template image. Finally, the face detection step was performed again to extract the 83 points of the normalised images.

This procedure yielded 166 X and Y coordinates which were then used as direct predictor variables in the SVM (see figure 5.2). Overall this process led to the elimination of several images whose pictorial defects or features (e.g. glasses, lighting, blur, background) resulted in errors, further reducing the studied population to a total of 663.

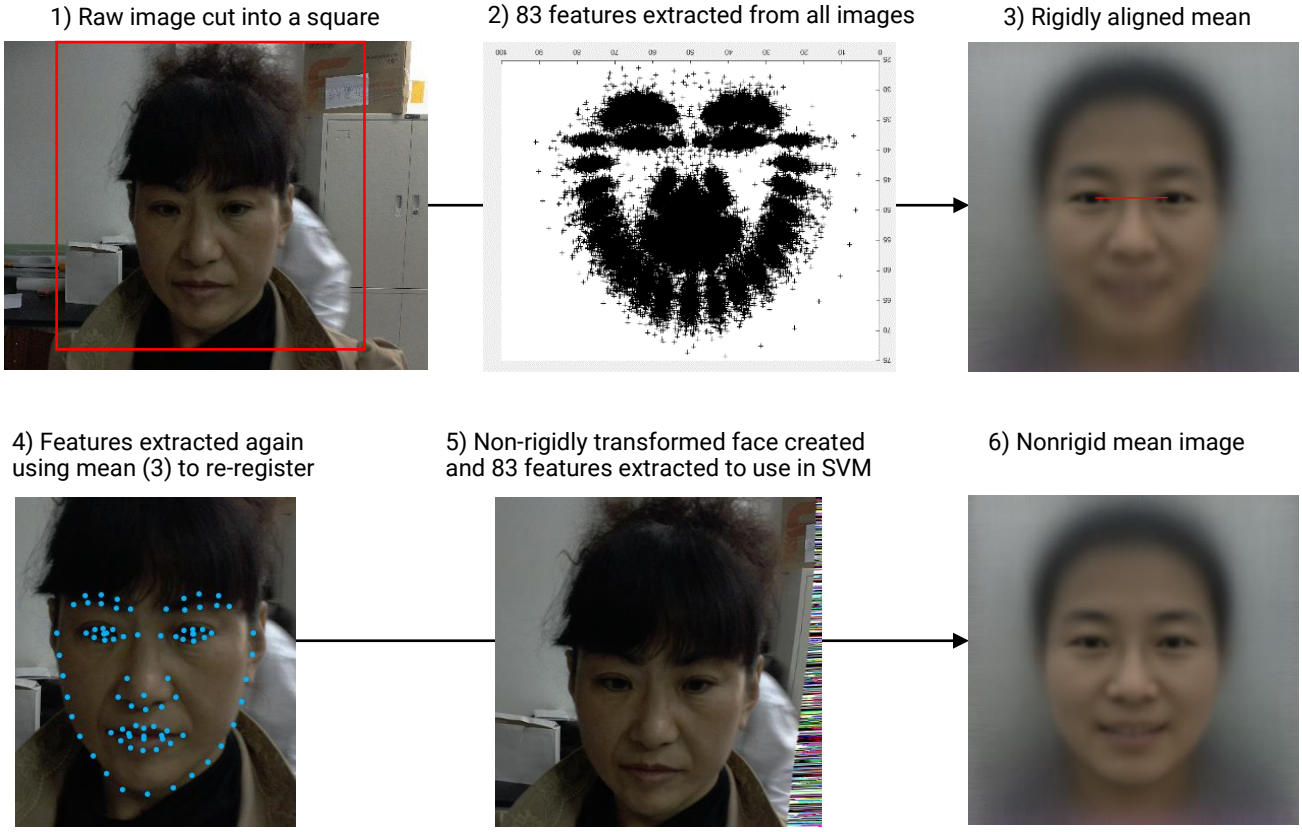


Figure 5.2 Example face parameterisation and feature extraction for SVM. The images were loaded into MATLAB and 1) cut into a square and quality adjusted, 2) 83 features were detected and extracted in point form and intraocular distance was used to 3) create a rigidly aligned mean image which was then 4) used as a template to non-rigidly align the faces. 5) The features were extracted once again and used as predictor variables in the SVM. 6) The mean of these non-rigidly transformed images exemplified the higher comparability of the outputs of this procedure.

5.2.3 Region-of-interest creation

To test whether relative redness of the face could be predictive of certain physiological markers like SaO2, the luminance channel was extracted and quantified from regions-of-interest (ROIs) of the face. To achieve this, the luminance channel of the nonrigidly aligned images were extracted and segmented into 15 different ROIs of 50 pixels. These

ROIs were the major landmarks of the face where changes are expected due to regional expansion.

These facial ROIs were further reduced in complexity by downsampling them into coarser images (Figure 5.3). To reduce the dimensionality of this massive result, the top 100 most informative data within the ROIs was selected with Principal Component Analysis (PCA) which was used in SVMs as predictor variables. Rather than concentrating solely on the red channel, including all light treated each pixel as an independent variable which made the system robust to affine changes such as those likely induced by variable illumination.

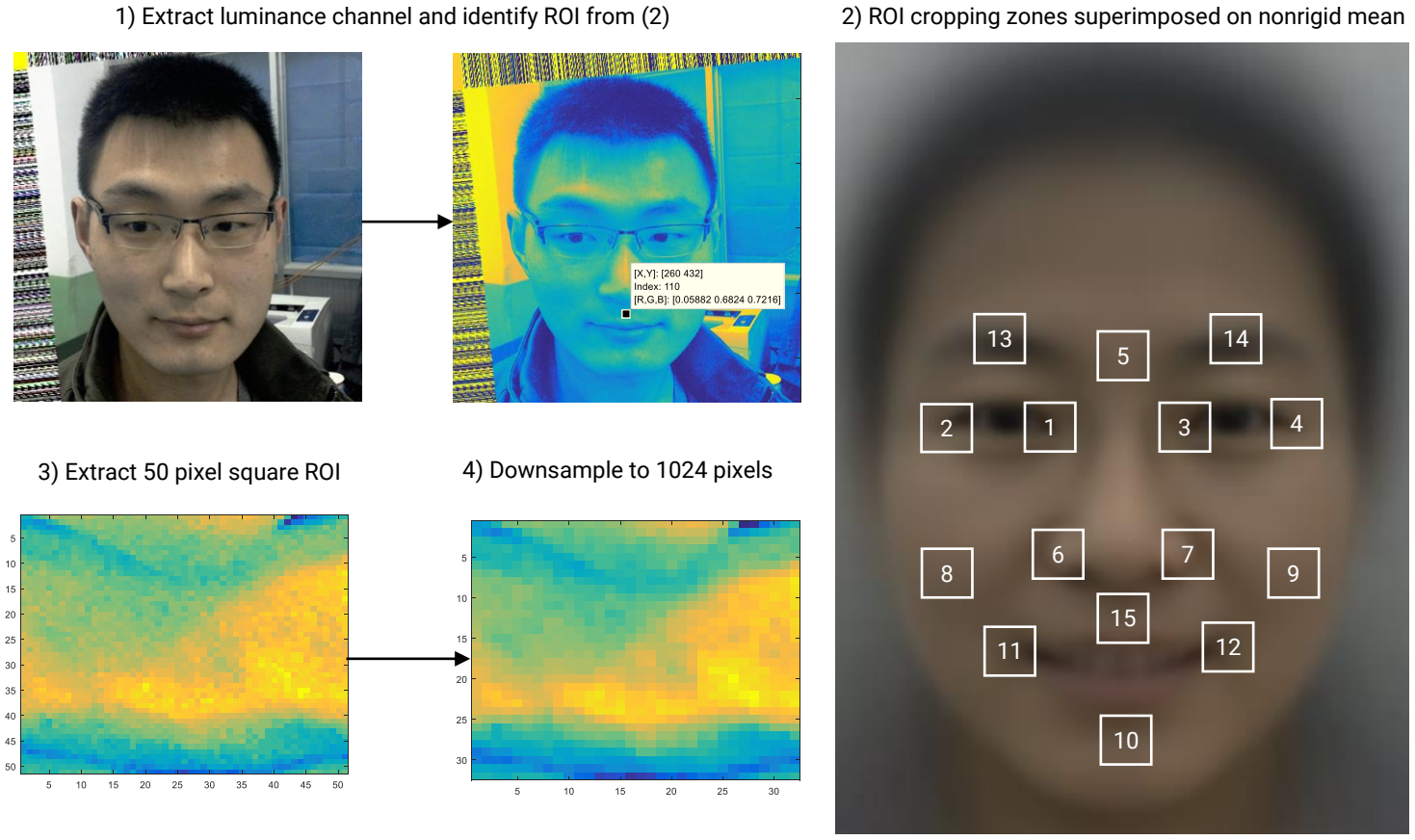


Figure 5.3 Example region-of-Interest (ROI) creations and zones. Each processed image was opened and 1) the luminance channel was extracted. 2) Each of the 15 ROIs was cropped and 3) extracted into 50 pixels around that particular point of interest and 4) downsampled to a manageable size whose 100 top PCA components were then used in the SVM as predictors. The 15 ROIs were the following 1) Inner Right Eye, 2) Outer Right Eye, 3) Inner Left Eye, 4) Outer Left Eye, 5) Nasal Root, 6) Right Nostril, 7) Left Nostril, 8) Right Maxilla, 9) Left Maxilla, 10) Chin, 11) Right mouth corner, 12) Left mouth corner, 13) Right eyebrow, 14) Left eyebrow, 15) Philtrum.

5.2.4 SVM-based prediction of variables of interest

HCT, SaO2, and echocardiographic measures were selected as physiological variables plausibly affected by high altitude to be predicted from the photographs of faces. Certain demographic variables were also included as covariates and served to conduct sanity checks on the validity of the SVM procedure.

For example, the *gender* variable should be easily predicted by a working SVM as the differences are obvious from photographs of faces. For the ROI analyses, the models were tested for predictive validity of gender and where a patch was significantly above chance, these were then tested for predictive power of physiological variables.

The variables to be predicted were normalized between -1 and 1 and matched to the photographs. If a photograph or physiological data was missing, the individual was dropped. For physiological variables, the 35% highest and lowest scores were selected, creating two separate groups. The 166 points coordinates of each nonrigidly aligned face were inputted into the SVM as predictor variables for the physiological variable of interest. The SVM model was validated with stratified 10-fold linear cross-validation by randomly dividing the data into 10 random parts, in which each class was represented in the same proportions as in the full dataset. Each part was used as training for learning the labels and informed the subsequent training of the next part which yielded the parameters (*C:* varying from -5 to 15) that yielded highest accuracy to run the classifier on the whole dataset. Once the best model was found, the SVMs were run with LibSVM (Chang and Lin 2011) on 50 different partitions of randomly selected 75% parts of the data, constantly shifting the selection of data to predict the labels with a radial basis function (RBF) kernel which estimated similarity of each new data point to the model.

The results from the binary classification task were presented in confusion matrices which show the success of the system at correctly classifying a test point into one of the following 4 categories: True Positive (TP) and True Negative (TN) if it predicts the label correctly or False Positive (FP) if it predicts a point as a positive when it is a negative and False Negative (FN) if predicted as a negative when it is a positive. These values then allow for the calculation of sensitivity (true positives / positives) and specificity (true negatives / negatives) and are used to generate more conservative balanced accuracies (sensitivity + specificity)/2.

If the aggregate mean of balanced accuracies were above 50% then this indicated the variable performed above chance. Specifically, the lowest confidence

interval (CI) needed to remain above 50%. The CIs were set at 95% and Bonferroni corrected by the number of variables predicting the same measure. For example, where HCT was predicted by 1) face parameters and 2) gender the alpha was corrected to 0.025 or 97.5% which was rounded to 98% yielding a z-score of 2.326. For parameters predicted by four variables, the alphas were corrected to 0.0125, or 99% (z-score = 2.575).

However, the *clinical threshold* for potential translational utility required the mean balanced accuracy for any predictor variable to reach at least 70%. If the aggregate mean of balanced accuracies reached this threshold, a two-samples t-test was used to compare it to other well-performing predictors. In these cases, the photographs of the highest and lowest scoring individuals for the measure would be averaged and compared visually for differences to test whether any morphological differences could be intuited visually.

5.3 Results

5.3.1 The population

All population measures are summarised in Table 5.1. Overall there were evenly balanced groups of Tibetan and Han Chinese individuals in ethnic origin and of males and females. The average individual was middle aged; however, there was great variability in years (+-15 years from mean). Physiologically, the population demonstrated normal HCT levels but about 5% lower than normal oxygen saturation as measured by SaO2 levels, indicative of mild hypoxemia. The average BMI score was within the 'healthy' range (e.g. 20-25; World Health Organization (2000)) as was the average heart rate (e.g. 60-100; Shaffer et al (2014)).

Table 5.1 Characteristics of the population (N = 663).

Other Characteristics		Echocardiographic measures	
Description	*n* (SD or percentage)	Description	measure (SD)
Ethnic origin		Systolic BP	115 (16.9)
Tibetan	358 (54%)		
Han Chinese	305 (46%)	Diastolic BP	79.4 (11.8)
Age			
Mean years	32.7 (14.7)	Diastolic RV area	14.1 cm² (3.2)
Gender			
Male	292 (44%)	RV fractional area change	49.7% (6)
Female	371 (56%)		
Proportional TiT		RV anterior wall thickness	3.5 mm (0.5)
Mean years	24 (16.2)		
Height		RV area	8.9 cm² (2.4)
Mean centimetres	161.8 (7.7)		
Weight		RV systolic pressure	23 (7.2)
Mean kilograms	60.2 (13.2)		
BMI		RV pressure gradient	25.4 (6.2)
Mean score	22.6 (3.6)		
		Tricuspid regurgitation velocity	2.6 (0.2)
Haematocrit			
Mean HCT levels	50% (0.06)	TAPSE	23.1 (2.3)
Oxygen saturation			
Mean SaO2 levels	88.9% (4.3)	Tricuspid valve stenosis	15.9 (2.2)
Hear Rate			
Mean heart rate	80.4 (12.9)	Right atrium pressure	3.9 (1)

Time in Tibet (TiT) = years of age / years spent at altitude; BP = blood pressure; RV = right ventricular; TAPSE = Tricuspid Annular Plane Systolic Excursion.

5.3.2 Sanity checks

The data used in the SVMs were only of the 663 with complete image data. However, it is possible that different number of cases appear in the results below for different measures as specific physiological measures could be missing from different participants. This explains the differences in number of *predicted* and *actual* results.

5.3.2.1 Whole-face parametrisation

To test system performance, predicting the *gender* variable from different data was a good problem as it is easily identifiable from photographs and can serve as a sanity check to explore if the method works. With only face parameters, gender was predicted at 85% balanced accuracy while adding age and ethnicity produced a comparable 84% (Table 5.2). This validated the system as whole-face parameters accurately predicted gender labels above chance and held greater predictive power than age or ethnicity as neither inclusion affected results significantly (Figure 5.4).

Table 5.2 Confusion Matrices for SVM predictions of Gender from whole-face parametrisation

Predicting Gender					
From face parameters			From face parameters, age, and ethnic origin		
BAcc = 85% (98% CI 84.41 to 85.59)			BAcc = 84% (98% CI 83.51 to 84.49)		
n=140	Predicted: M	Predicted: F	n = 197	Predicted: M	Predicted: F
Actual: M	68	10	Actual: M	68	10
Actual: F	11	51	Actual: F	69	50

Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; M = Male; F = Female.

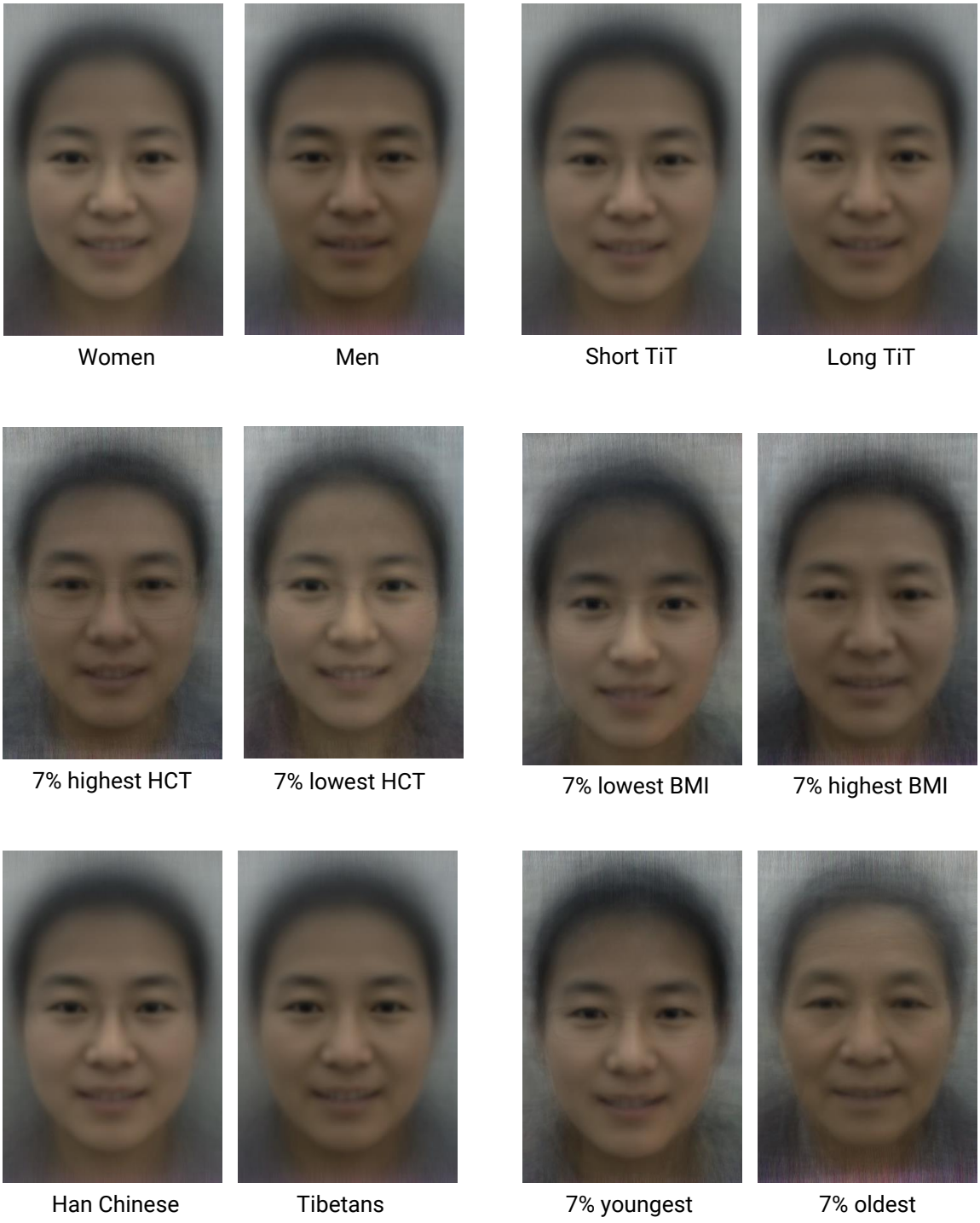


Figure 5.4 Example mean images from the population were selected to demonstrate differences found as assessed by face parameters. These draw differences between (A) women and men (B) long and short proportional time spent in Tibet (TiT), (C) highest and lowest haematocrit levels, (D) highest and lowest body mass indexes (BMI), (E) ethnic origin, and (F) according to age.

5.3.2.2 Region-of-Interest parametrisation

Similarly, to the whole-face parametrisation, the ROI zones were assessed for discriminatory capacity of the *gender* variable. Combining all face ROI parameters together did not result in above-chance prediction of gender. Furthermore, including age and ethnic origin as test co-variates increased performance by 7% above chance (Table 5.2). When considering each ROI separately as a predictor of gender, none performed significantly above chance (Figure 5.5).

The SVM results suggest that this ROI parametrisation of the faces would likely not hold value as a dimensionality reduction system for face parametrisation. Thus, ROI analyses were not performed further, and prediction models were considered only with whole face parametrisation.

Table 5.2 Confusion Matrices for SVM predictions of Gender from face ROIs

Predicting Gender					
From face ROIs			From face ROIs, age, and ethnic origin		
BAcc = 50% (98% CI 50 to 50)			BAcc = 57% (98% CI 56.41 to 57.59)		
n=140	Predicted: M	Predicted: F	n = 140	Predicted: M	Predicted: F
Actual: M	78	0	Actual: M	55	22
Actual: F	62	0	Actual: F	36	27

Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; M = Male; F = Female.

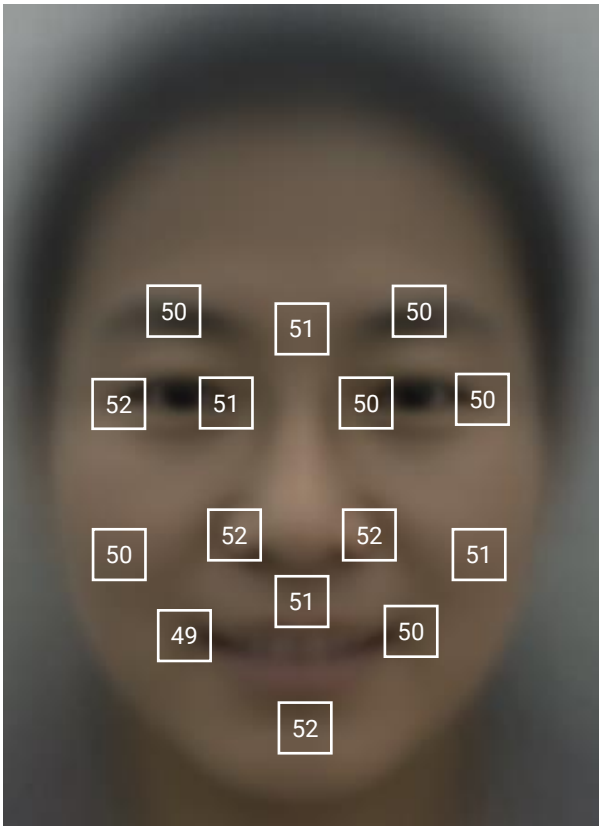


Figure 5.5 ROI based predictions of gender. Each ROI contains the balanced accuracies from predicting gender only with the parameters contained in the square. None significantly predicted gender above chance.

5.3.3 Predicting Haematocrit levels

A SVM predicting HCT levels from face parameters returned a balanced accuracy of 77%. Although significant, this was 10% worse performance than when predicting HCT from gender alone, which yielded a balanced accuracy of 87%. A two-sample t-test confirmed this was a significant difference (t (2,98) = 13.3, p<0.001).

Table 5.4 Confusion Matrices for SVM predictions of HCT from whole face parametrisation

From face parameters			From gender		
BAcc = 77% (98% CI 75.98 to 78.02)			BAcc = 87% (98% CI 86.17 to 87.83)		
n= 82	Predicted: H	Predicted: L	n = 82	Predicted: H	Predicted: L
Actual: H	32	9	Actual: H	37	4
Actual: L	10	31	Actual: L	6	35

Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.

Furthermore, a two-sample t-test found HCT levels were significantly different between men and women (t(1,917) = 18.3, p<0.001) which is in line with clinical methodology, suggesting higher levels in men than in women (Clark and Kruse 1990). Further testing was therefore run separately on men and women by binarizing intra-gender high and low variability in HCT levels and assessing predictability from face parameters or gender labels alone. The face parameters performed significantly above chance in contrast to *gender only* labels which performed at chance since any variability within this label was removed during segmentation of the groups by gender (Table 5.5).

Table 5.5 Confusion Matrices for SVM predictions of HCT from whole face

Predicting HCT for women					
From face parameters			From sex		
BAcc = 53% (98% CI 50.7 to 55)			BAcc = 50% (98% CI 50 to 50)		
n= 48	Predicted: H	Predicted: L	n = 46	Predicted: H	Predicted: L
Actual: H	14	10	Actual: H	23	0
Actual: L	13	11	Actual: L	23	0
Predicting HCT for men					
From face parameters			From sex		
BAcc = 60% (98% CI 56.98 to 63)			BAcc = 50% (98% CI 50 to 50)		
n= 38	Predicted: H	Predicted: L	n = 36	Predicted: H	Predicted: L
Actual: H	10	9	Actual: H	18	0
Actual: L	7	12	Actual: L	18	0

Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.

Predictions of HCT from face parameters for men performed an average 7% better than for women but 17% worse than if combined with women's face parameters. A two-sample t-test comparing balanced accuracy results of HCT predictions from both genders and men only found this 17% performance variation to be significantly different (t (2,98) = 12.7, p<0.001).

A final prediction of HCT was performed with proportional TiT and ethnicity separately finding these to both and independently perform equally above chance ($t(2,98) = 1.87$, $p>0.05$).

Table 5.6 Confusion Matrices for SVM predictions of HCT

From ethnicity			From proportional TiT		
BAcc = 63% (98% CI 61.6 to 64.4)			BAcc = 61% (98% CI 59.8 to 62)		
n= 82	Predicted: H	Predicted: L	n = 84	Predicted: H	Predicted: L
Actual: H	27	14	Actual: H	26	16
Actual: L	16	25	Actual: L	17	25

Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.

5.3.4 Predicting oxygen saturation

Predictions of SaO2 from face parameters performed 3% above chance though compared with a two-sample t-test to covariate predictors, performed the same as gender ($t(2,98) = 1.3$, $p>0.05$), and significantly worse than ethnicity ($t(2,98) = 2.4$, $p<0.05$) and proportional TiT ($t(2,98) = 7.0$, $p<0.001$).

Table 5.7 Confusion Matrices for SVM predictions of SaO2

From face parameters			From gender		
BAcc = 53% (99% CI 51.9 to 54.1)			BAcc = 52% (99% CI 50.8 to 53.2)		
n=100	Predicted: H	Predicted: L	n = 98	Predicted: H	Predicted: L
Actual: H	27	23	Actual: H	22	27
Actual: L	23	27	Actual: L	20	29
From ethnicity			From proportional TiT		
BAcc = 56% (99% CI 54.8 to 57.2)			BAcc = 60% (99% CI 58.8 to 61.2)		
n= 98	Predicted: H	Predicted: L	n = 99	Predicted: H	Predicted: L
Actual: H	29	20	Actual: H	30	20
Actual: L	24	25	Actual: L	20	29

Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.

5.3.5 Predicting BMI

Individual SVMs with either face parameters, gender, ethnicity, or proportional TiT did not predict BMI above chance (where BMI was binarized as high (>25) or low (<24.9); Table 5.7).

Table 5.8 Confusion Matrices for SVM predictions of BMI

From face parameters			From gender		
BAcc = 50% (99% CI 50 to 50)			BAcc = 50% (99% CI 50 to 50)		
n=173	Predicted: H	Predicted: L	n = 174	Predicted: H	Predicted: L
Actual: H	139	0	Actual: H	139	0
Actual: L	34	0	Actual: L	35	0
From ethnicity			From proportional TiT		
BAcc = 50% (99% CI 50 to 50)			BAcc = 50% (99% CI 50 to 50)		
n= 173	Predicted: H	Predicted: L	n = 173	Predicted: H	Predicted: L
Actual: H	138	0	Actual: H	137	0
Actual: L	35	0	Actual: L	36	0
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

5.3.6 Predicting proportional time in Tibet

When predicting how long people had spent in Tibet (separating those above and under the median of 64% of their lives spent in Tibet), HCT, gender, and face parameters did not perform above chance while ethnicity performed significantly above chance at 81% (Table 5.9). This was unsurprising as it suggested people were more likely to spend time in Tibet if they were originally from the geographical area.

Table 5.9 Confusion Matrices for SVM predictions of proportional TiT

From face parameters			From gender		
BAcc = 50% (99% CI 50 to 50)			BAcc = 49% (99% CI 48.14 to 49.86)		
n=172	Predicted: H	Predicted: L	n = 172	Predicted: H	Predicted: L
Actual: H	86	0	Actual: H	40	46
Actual: L	86	0	Actual: L	41	45
From ethnicity			From HCT		
BAcc = 81% (99% CI 80.47 to 81.53)			BAcc = 51% (99% CI 50.63 to 51.37)		
n= 172	Predicted: H	Predicted: L	n = 176	Predicted: H	Predicted: L
Actual: H	66	20	Actual: H	83	5
Actual: L	13	73	Actual: L	82	6
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

5.3.7 Predicting ethnic origin

When attempting to predict whether individuals were more likely to be Tibetans or Han Chinese face parameters, gender labels, and HCT levels only performed at chance, while proportional TiT was a significant predictor at 82% (Table 5.10).

Table 5.10 Confusion Matrices for SVM predictions of ethnic origins

From face parameters			From gender		
BAcc = 50% (99% CI 50 to 50)			BAcc = 50% (99% CI 50 to 50)		
n=172	Predicted: H	Predicted: T	n = 173	Predicted: H	Predicted: T
Actual: H	80	0	Actual: H	0	80
Actual: T	92	0	Actual: T	0	93
From proportional TiT			From HCT		
BAcc = 82% (99% CI 81.47 to 82.53)			BAcc = 51% (99% CI 50. 23 to 51.77)		
n= 172	Predicted: H	Predicted: L	n = 176	Predicted: H	Predicted: L
Actual: H	67	12	Actual: H	74	5
Actual: T	19	74	Actual: T	83	9

Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = Han; T = Tibetan.

5.3.8 Predicting echocardiographic measures

For echocardiographic measurements three additional variables were selected to compare predictability performance with whole face information: gender, ethnicity, and BMI. These were selected as none required specialist equipment to obtain at altitude and therefore provided the most translational benefit. For all these measures, the highest and lowest 10% were binarized and used for prediction.

5.3.8.1 Systolic blood pressure

Predicting high and low systolic blood pressure, consisting of arterial pressure during contraction of the right ventricle, found only face parameters, gender, and BMI to perform significantly above chance. None reached the clinical threshold.

Table 5.11 Confusion Matrices for SVM predictions of systolic blood pressure

From face parameters			From gender		
BAcc = 66% (99% CI 64.8 to 67.2)			BAcc = 65% (99% CI 63.9 to 66.1)		
n=100	Predicted: H	Predicted: L	n = 98	Predicted: H	Predicted: L
Actual: H	36	14	Actual: H	35	14
Actual: L	20	30	Actual: L	20	29
From ethnicity			From BMI		
BAcc = 49% (99% CI 47.8 to 50.2)			BAcc = 68% (99% CI 67.05 to 68.95)		
n= 100	Predicted: H	Predicted: L	n = 100	Predicted: H	Predicted: L
Actual: H	24	26	Actual: H	45	5
Actual: L	24	26	Actual: L	27	23
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

5.3.8.2 Diastolic blood pressure

Predicting high and low diastolic blood pressure, when the heart muscle rests between beats, found all four variables performed above chance though none passed the threshold of clinical utility.

Table 5.12 Confusion Matrices for SVM predictions of diastolic blood pressure

From face parameters			From gender		
BAcc = 65% (99% CI 63.8 to 66.2)			BAcc = 64% (99% CI 62.7 to 65.3)		
n= 99	Predicted: H	Predicted: L	n = 99	Predicted: H	Predicted: L
Actual: H	34	16	Actual: H	34	15
Actual: L	19	30	Actual: L	20	30
From ethnicity			From BMI		
BAcc = 53% (99% CI 51.8 to 54.2)			BAcc = 69% (99% CI 67.9 to 70.1)		
n= 100	Predicted: H	Predicted: L	n = 100	Predicted: H	Predicted: L
Actual: H	29	21	Actual: H	45	5
Actual: L	26	24	Actual: L	27	23
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

5.3.8.3 Right ventricular systolic pressure

Predicting high and low right ventricular systolic pressure, the peak arterial pressure, showed all variables but gender performed above chance though none reached clinical threshold significance.

Table 5.13 Confusion Matrices for SVM predictions of right ventricular systolic pressure					
From face parameters			From gender		
BAcc = 53% (99% CI 52 to 54)			BAcc = 50% (99% CI 49 to 51)		
n=120	Predicted: H	Predicted: L	n = 120	Predicted: H	Predicted: L
Actual: H	38	22	Actual: H	30	30
Actual: L	34	26	Actual: L	30	30
From ethnicity			From BMI		
BAcc = 52% (99% CI 51 to 53)			BAcc = 57% (99% CI 56.15 to 57.85)		
n= 118	Predicted: H	Predicted: L	n = 118	Predicted: H	Predicted: L
Actual: H	32	27	Actual: H	50	9
Actual: L	30	29	Actual: L	42	17
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

5.3.8.4 Right ventricular pressure gradient

When predicting high and low right ventricular pressure gradient between the right ventricle and right atrium, all variables performed above chance though none reached the clinical threshold.

Table 5.14 Confusion Matrices for SVM predictions of right ventricular pressure gradient					
From face parameters			From gender		
BAcc = 51% (99% CI 50.08 to 51.92)			BAcc = 51% (99% CI 50.03 to 51.97)		
n=118	Predicted: H	Predicted: L	n = 118	Predicted: H	Predicted: L
Actual: H	33	26	Actual: H	28	31
Actual: L	32	27	Actual: L	27	32
From ethnicity			From BMI		
BAcc = 56% (99% CI 55.1 to 56.9)			BAcc = 52% (99% CI 51.17 to 52.83)		
n= 119	Predicted: H	Predicted: L	n = 118	Predicted: H	Predicted: L
Actual: H	35	25	Actual: H	48	11
Actual: L	28	31	Actual: L	45	14
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

5.3.8.5 Right atrium pressure

Predicting high and low right atrium pressure reflecting the ability to pump blood back into the arterial system from the heart, found all but the face parameter to perform above chance though none reached the clinical threshold.

Table 5.15 Confusion Matrices for SVM predictions of right atrium pressure					
From face parameters			From gender		
BAcc = 50% (99% CI 48.9 to 51.1)			BAcc = 53% (99% CI 52.11 to 53.89)		
n=119	Predicted: H	Predicted: L	n = 120	Predicted: H	Predicted: L
Actual: H	30	29	Actual: H	29	31
Actual: L	30	30	Actual: L	25	35
From ethnicity			From BMI		
BAcc = 57% (99% CI 55.8 to 58.2)			BAcc = 51% (99% CI 50.2 to 51.8)		
n= 116	Predicted: H	Predicted: L	n = 119	Predicted: H	Predicted: L
Actual: H	35	23	Actual: H	13	47
Actual: L	27	31	Actual: L	11	48
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

5.3.8.6 Diastolic right ventricular area

When predicting high and low diastolic right ventricular area, a known predictor of pulmonary hypertension, all variables performed above chance, but none passed the threshold of clinical utility.

Table 5.16 Confusion Matrices for SVM predictions of diastolic right ventricular area					
From face parameters			From gender		
BAcc = 61% (99% CI 60.04 to 61.96)			BAcc = 67% (99% CI 66 to 68)		
n=120	Predicted: H	Predicted: L	n = 120	Predicted: H	Predicted: L
Actual: H	38	22	Actual: H	43	17
Actual: L	24	36	Actual: L	22	38
From ethnicity			From BMI		
BAcc = 55% (99% CI 54.01 to 55.99)			BAcc = 62% (99% CI 61.1 to 62.9)		
n= 119	Predicted: H	Predicted: L	n = 117	Predicted: H	Predicted: L
Actual: H	31	29	Actual: H	54	5
Actual: L	24	35	Actual: L	39	19
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

5.3.8.7 Right ventricular area

When predicting the high and low area of the right ventricle, a measure of expansion of the heart chamber, all variables performed above chance though none reached the clinical utility threshold.

Table 5.17 Confusion Matrices for SVM predictions of right ventricular area

From face parameters			From gender		
BAcc = 68% (99% CI 67 to 69)			BAcc = 62% (99% CI 61 to 63)		
n=118	Predicted: H	Predicted: L	n = 118	Predicted: H	Predicted: L
Actual: H	42	17	Actual: H	37	22
Actual: L	21	38	Actual: L	22	37
From ethnicity			From BMI		
BAcc = 58% (99% CI 57.05 to 58.95)			BAcc = 66% (99% CI 65.15 to 66.85)		
n= 118	Predicted: H	Predicted: L	n = 118	Predicted: H	Predicted: L
Actual: H	31	28	Actual: H	54	5
Actual: L	22	37	Actual: L	36	23

Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.

5.3.8.8 Right ventricular fractional area change

When predicting high and low right ventricular fractional area change, a measure dependent on both diastole and systole function, only face parameters and gender labels predicted it above chance. None reached the threshold of clinical utility.

Table 5.18 Confusion Matrices for SVM predictions of right ventricular fractional area change

From face parameters			From gender		
BAcc = 52% (99% CI 51.01 to 52.99)			BAcc = 58% (99% CI 57.06 to 58.94)		
n=120	Predicted: H	Predicted: L	n = 120	Predicted: H	Predicted: L
Actual: H	30	30	Actual: H	31	29
Actual: L	27	33	Actual: L	21	39
From ethnicity			From BMI		
BAcc = 48% (99% CI 47.15 to 48.85)			BAcc = 50% (99% CI 49.15 to 50.85)		
n= 120	Predicted: H	Predicted: L	n = 120	Predicted: H	Predicted: L
Actual: H	31	29	Actual: H	14	46
Actual: L	33	27	Actual: L	14	46

Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.

5.3.8.9 Right ventricular anterior wall thickness

Predicting high and low right ventricular wall thickness, a measure of the muscle mass required to pump the blood to the lungs, all variables performed above chance. In addition, the face parameters and BMI passed the threshold of clinical utility, performing equally well (t(2,98) = -1.6, p>0.05).

Table 5.19 Confusion Matrices for SVM predictions of right ventricular anterior wall thickness

From face parameters			From gender		
BAcc = 72% (99% CI 70.9 to 73.1)			BAcc = 56% (99% CI 55.1 to 56.9)		
n=120	Predicted: H	Predicted: L	n = 118	Predicted: H	Predicted: L
Actual: H	44	16	Actual: H	35	24
Actual: L	18	42	Actual: L	27	32
From ethnicity			From BMI		
BAcc = 52% (99% CI 51 to 53)			BAcc = 73% (99% CI 72.26 to 73.74)		
n= 118	Predicted: H	Predicted: L	n = 118	Predicted: H	Predicted: L
Actual: H	28	31	Actual: H	58	1
Actual: L	25	34	Actual: L	31	28
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

In accordance with the procedure delineated in section 5.3.8, average faces were created post-hoc for the highest and lowest individuals dependent on RVAWT (figure 5.6). From visually inspecting the images, differences included darker skin tone and wider faces in those with higher RVAWT levels. Manual estimation of the intra-cheek distance on the mean images confirmed the high RVAWT average face to be 13mm longer. Converting the images to grayscale and averaging the pixel intensity provided a measure of brightness and confirmed that high RVAWT images were darker by 0.9 pixels (px).

Figure 5.6. Average faces from individuals exhibiting highest (left) and lowest (right) RVAWT with red lines depicting the intra-cheek distance and squares showing differing brightness pixel levels.

5.3.8.10 Tricuspid regurgitation velocity

When measuring high and low tricuspid regurgitation velocity, the ability of the tricuspid valve to close during systole, preventing leakages of blood from the right ventricle into the right atrium, all variables performed above chance though none reached the threshold of clinical utility.

Table 5.20 Confusion Matrices for SVM predictions of tricuspid regurgitation velocity

From face parameters			From gender		
BAcc = 52% (99% CI 51.08 to 52.92)			BAcc = 51% (99% CI 50.03 to 51.97)		
n=119	Predicted: H	Predicted: L	n = 119	Predicted: H	Predicted: L
Actual: H	32	27	Actual: H	28	32
Actual: L	30	30	Actual: L	26	33
From ethnicity			From BMI		
BAcc = 55% (99% CI 54 to 56)			BAcc = 52% (99% CI 51.24 to 52.76)		
n= 118	Predicted: H	Predicted: L	n = 119	Predicted: H	Predicted: L
Actual: H	34	25	Actual: H	48	12
Actual: L	28	31	Actual: L	45	14
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

5.3.8.11 Tricuspid annular plane systolic excursion (TAPSE)

When measuring high and low TAPSE, the degree of displacement of the tricuspid annular plane, all variables performed above chance though none reached the threshold of clinical utility.

Table 5.21 Confusion Matrices for SVM predictions of TAPSE

From face parameters			From gender		
BAcc = 54% (99% CI 53.17 to 54.83)			BAcc = 55% (99% CI 54.03 to 55.97)		
n=119	Predicted: H	Predicted: L	n = 119	Predicted: H	Predicted: L
Actual: H	34	25	Actual: H	35	25
Actual: L	29	31	Actual: L	28	31
From ethnicity			From BMI		
BAcc = 60% (99% CI 59.1 to 60.9)			BAcc = 58% (99% CI 57.03 to 58.97)		
n= 117	Predicted: H	Predicted: L	n = 118	Predicted: H	Predicted: L
Actual: H	31	26	Actual: H	50	9
Actual: L	21	37	Actual: L	40	19
Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.					

5.3.9 Tricuspid valve stenosis

When predicting high and low tricuspid valve stenosis, a measure of morphological alterations of the valve, all variables but gender and ethnicity performed above chance while none reached statistical significance.

Table 5.22 Confusion Matrices for SVM predictions of tricuspid valve stenosis

From face parameters			From gender		
BAcc = 51% (99% CI 50.08 to 51.92)			BAcc = 49% (99% CI 48.06 to 49.94)		
n=118	Predicted: H	Predicted: L	n = 119	Predicted: H	Predicted: L
Actual: H	31	28	Actual: H	32	28
Actual: L	30	29	Actual: L	33	26
From ethnicity			From BMI		
BAcc = 50% (99% CI 49.15 to 50.85)			BAcc = 53% (99% CI 52.15 to 53.85)		
n= 120	Predicted: H	Predicted: L	n = 120	Predicted: H	Predicted: L
Actual: H	32	28	Actual: H	49	11
Actual: L	32	28	Actual: L	45	15

Abbreviations: BAcc = Balanced accuracy; n is the total number of predictions made; H = High; L = Low.

5.4 Discussion

5.4.1 Whole face versus regions of interest

The initial sanity checks demonstrated the power of the SVM methods in correctly predicting gender from whole-face parametrisations at 85% accuracy. As these results were not seen for the single ROI analyses this validated the approach for whole-face analyses only and ROI luminance analyses were discarded from further steps. The whole-face consisted of 83-point X-Y coordinate values while the ROIs were richer pixel data of certain regions. This suggests that parametrising the relationship of the coordinates of the major landmarks of the face was more informative than reducing it to detailed information about single regions. This finding lends support to the high-dimensional approach adopted in this thesis whereby a low precision high dimensional parametrisation proves superior over a high precision, low dimensional one.

5.4.2 Implications for echocardiographic measures

The major aim of this chapter was to explore whether it was possible to predict echocardiographic measures of clinical utility with facial parameters. Most echocardiographic variables were predicted above chance but did not reach the clinical utility threshold above 70% for predictive performance. Some variables performed close

to this clinical threshold with face parameters predicting right ventricular area and diastolic and systolic blood pressure above 65%. A notable exception was predicting right ventricular anterior wall thickness (RVAWT) which passed the clinical utility threshold significantly for both face parameters and BMI and whose differences could not be explained by gender or ethnicity.

RVAWT is commonly used as a measure of abnormal enlargement of the right ventricle to diagnose pulmonary hypertension (Matsukubo et al 1977; Steudel et al 1998; Ozaki et al 2001; Wanstall et al 2002; Liu et al 2006). Although both the face and BMI predicted RVAWT equally well, these can be considered independent measures as face parameters do not predict BMI and are therefore unrelated. Calculating the BMI of a person requires the exact values for their age, sex, height, and weight. Obtaining these may either require specialist equipment, albeit lighter than an echocardiogram, to obtain accurate measures or rely on self-report which is associated with its own set of reliability issues. Obtaining a photo of someone's face is far simpler as most mobile phones have integrated cameras. The photograph is also an objective physical measure that can be easily transferred remotely via the internet if the researchers cannot reach patients at altitude. Therefore, the face holds more potential than BMI for translation into a screening measure for pulmonary hypertension.

It was expected that differences in predictability could be made between measures where a clear difference was visually intuitive from average images of separate tails of a variable dividing a population. The opposite, where differences in physiology based on positive performance from face parameters was also hypothesised. Comparison of average faces for the highest and lowest individuals, dependent on RVAWT, did find intuitive and measurable differences including darker skin tone and wider faces in those with higher RVWAT levels. More precise measurements of the faces are required to confirm whether this was due specifically as a result of maxillary expansion, however these results echo findings of facial expansion found in Thalassaemic patients (Rund et al 1997; Phadke and Agarwal 2003; George 2013).

Moreover, the darker skin tone found in individuals exhibiting higher RVWAT could be due to abnormal oxygen levels in the blood stream, where low oxygen saturation levels were found to correlate with darker skin tones (Takiwaki et al 2002). For example, a higher concentration of carbon monoxide in blood is associated with "cherry red" skin colouration (Hoppe 1857), another symptom of hypoxia (Gorman et al 1992; Simini 1998; Goldstein 2008). Indeed, the direction of these differences suggests confirmation for the hypothesis that a hypoxic environment may produce measurable changes in facial

expansion that are measurable from face photograph and provide clinically significant data.

5.4.3 Other physiological measures

With regards to the haematological measures, after correcting for gender biases within HCT levels, face parameters could predict HCT levels in both men and women above chance, with 60% and 53% mean balanced accuracies respectively. This difference between genders could be due to the higher HCT levels naturally occurring in male populations which could have a more marked effect on face physiology. Indeed, as HCT levels were hypothesised to be associated with expansion of the maxilla, it is possible that the SVM method was modelling these changes. Although the performance for predicting HCT levels for men was encouraging, it is premature to assign clinical utility to the results of the current system. Nevertheless, this was a promising and positive result in support of continuing research into this approach.

Face parameters performed poorly – either at chance or just above chance without clinical utility – in predicting oxygen saturation, BMI, proportional TiT, and ethnic origin. This finding was surprising, as clear differences can be seen by visually inspecting the images consisting of the extremes within each group represented in Figure 5.4 and were expected to be detected by the SVM. Nevertheless, humans have developed a special expertise at discriminating between often similar faces as functioning within a society depends on this skill. It is possible that despite the capacity for machine learning systems to detect obvious differences such as gender, different approaches may be required to detect more subtle differences from face parameters.

Predicting haematological measures from ethnicity and proportional TiT demonstrated similar performances to face parameters for both HCT and oxygen saturation. As neither of these measures were predictable by face parameters, it is an independent finding that characterises the underlying population and more consistently than face parameters as performance was equal for HCT and SaO2 levels. These two variables are correlated as individuals of Tibetan origin are more likely to have spent a long time in Tibet (at altitude) as a proportion of their lives, explaining the similar performance in predictive power. This was further confirmed by the high level of power that each variable has in predicting one another (upwards of 80%). It is possible that this finding reflects the established observations that living at height has a marked effect on these haematological measures (Niermeyer et al 1995; Yi et al 2010). Nevertheless, no further extrapolation can be made as it is not possible to distinguish between the relative

contribution of the genetic component in ethnicity compared to the environmental effect of living at altitude with the present data.

5.4.4 Limitations and future directions

Although the face parametrisation procedure generated an accurate high-dimensional model of face features, a more sensitive measure that generated more careful feature detection or more parameters could potentially yield more informative input predictors. Indeed, the current parameterisation method reduced the face to 83-point coordinates of the major landmarks. It is therefore possible that when methods are developed that increase the dimensionality level of face parametrisation, it might lead to more accurate predictions.

As the face images were taken as a favour to our research team, no strict photographic procedure was followed. Employing a rigorous acquisition procedure to take the photographs could increase the quality of the data and predictability of the output. For example, asking participants to remove eyewear, pose with a neutral expression, and ensure even lighting and background could facilitate face parametrisation and improve the quality of the data.

As the dataset originates from an ongoing study, these analyses were run on a subset of 980 individuals of the expected 9000+ sample, consisting of about 10% of the total population studied. Although positive results were found for a single echocardiographic measure, it is possible that once the full dataset is obtained from the Imperial College London research team, the system performs significantly better to merit further translation into a clinical tool.

Chapter 6: General discussion

Vladimir Bartol's (2004) maxim from *Amalut*: "Nothing is an absolute reality; all is permitted"

6.1 Overview

This thesis adopted a high-dimensional approach to a translational research question with the aims of expanding current knowledge and testing whether well-established findings can be translated into impactful health outcomes. The biological was considered as distributed across multiple variables which required high-dimensional parametrisation to capture their full variability. This was suggested to be necessary to identify solutions that may work for individuals and successfully implement the translational focus of the thesis.

Before we examine the features of the high-dimensional approach and their exemplification in the work of this thesis, it is helpful to be reminded of the characteristics and rationale behind the conventional, low-dimensional approach. A clinician—and by extension a clinical scientist—is ultimately concerned with modifying the properties of a biological system, for that is what is needed when it goes awry. This most naturally requires understanding the system, by which is meant knowing the relation between its components, or at least a critical subset of them, in a way that allows to predict what will happen when these are modified. It is this extension beyond mere prediction, that compels the biomedical scientist to seek fully explanatory models.

The problem this raises is the need for an explanatory, mechanistic model to be simple enough to permit this kind of analysis. This may—or may not—be true of any biological system. Where the inherent complexity is clearly very high, as is obviously the case with the brain, the assumption dramatically narrows the range of systems accessible to study. Note this problem is not soluble by empirical means, for an inherently complex system cannot have a good simple model, no matter how hard it is sought. The ambition to discover a facet of the system simple enough to be given an easily manipulatable causal account, may be unfulfillable.

Instead, in employing high-dimensional models, the solution I advocate here does not abandon the ambition of modifying biological systems but rather bypasses the *requirement* for understanding, at least the conventional kind of mechanistic understanding. Naturally there still remains a wish to achieve as perspicuous a description of the system under study but need not be anchored to it.

Greater fidelity to biological reality is perhaps the most important reason for seeking an enhanced dimensionality. But another proceeds from the fundamental nature of individuation on which most clinical action is fundamentally premised. A system with

many parameters will tend to have multiple good local solutions, requiring an adequate description of any given individual to refer to many of them. Therefore, not only understanding biology, but applying it to an individual, theoretically benefits from the high-dimensional approach.

Theory is one thing, its application another. For this, the specific challenges and advantages the high-dimensional approach presents in practice should be considered. The challenges are largely familiar: the more complex a model, the more independent data points it requires to be estimated with any confidence. Assuring data quality scales in difficulty and cost with data size: that much is obvious and incontestable. But in the realm of cognition and behaviour, data acquisition is facilitated by the societal frameworks through which they are naturally communicated. Mobile communications in particular have now essentially interposed an interrogable layer between the researcher and many of those with whom are engaged in behaviours and cognitions of research interest. Moreover, commodity digital devices are now the gateway to synthetic worlds with which people voluntarily engage for pleasure—computer gaming—revealing, painlessly, a great deal about human cognitive powers and dispositions. Obtaining large scale datasets is thus enabled by architecture motivated and funded from elsewhere.

But it is important also to consider the practical advantages of modelling a biological system in a high-dimensional way beyond any biological necessity for it. These are often overlooked. Where the biological signal is distributed amongst many variables—conveyed in the covariance structure, to put it technically—it becomes innately resistant to the most common forms of noise and bias. Where noise affects each individual variable independently—as is usual for instrument noise—the covariance structure remains essentially unaffected by it.

As an example, consider the recognisability of a photograph of a human face (Figure 6.1.A), with a great deal of white noise superimposed on it (Figure 6.1.B). Similarly, a simple bias, such as an affine transform—again a common feature of instrument-driven corruption—will also tend to be rejected in the same way. For example, a face that has been subjected to shear (Figure 6.1.C). However, recognising someone by only looking at a single parameter such as the length of the nose (Figure 6.1.C) will require a very high degree of precision and will therefore be more likely to be affected by noise (Figure 6.1.D). The parameters lose their informative potential for individuation if the metric is low-dimensional and reductive, but not when it is considered as part of a high-dimensional whole, as typically the case with human brains.

This approach to noise-rejection is dependent on the existence of a robust covariance structure, as is the case with the human face. Whether it is true of any biological system is to be determined case by case. Such determination is necessarily data-driven, for the more complex the covariance structure the less likely it is that we can plausibly arrive at it *a priori*. This encourages the use of data-driven modelling techniques that are shaped by the data rather than coercing it into conformity with themselves.

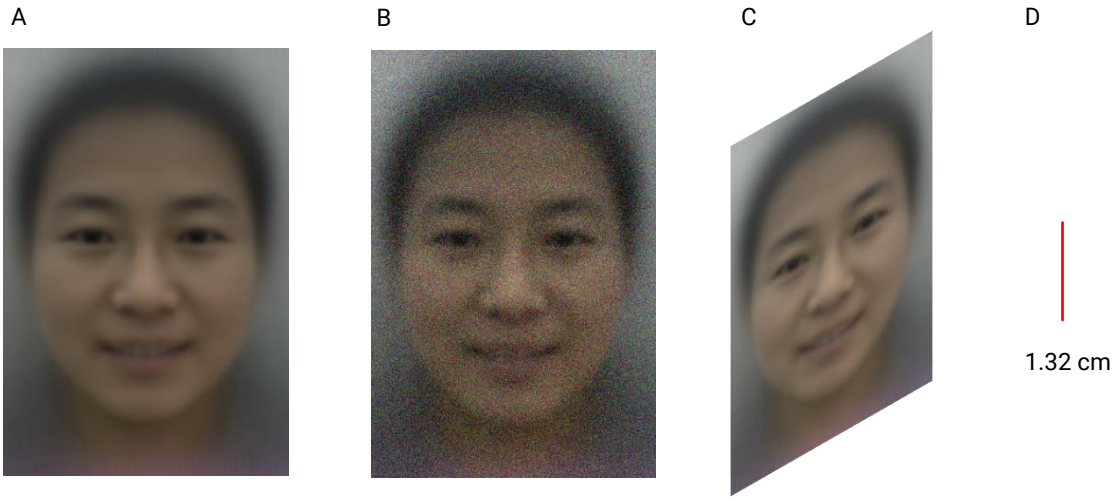


Figure 6.1. Individuating information in the face. A) The original image, B) with white noise added, C) with shear distortion and D) the single metric of the length of the nose. A complex biological system like the face is more robust to noise variations than any single precise parameter.

6.2 Assessing the high-dimensional translational approach

The human face was selected as the central high-dimensional entity to unify this endeavour as many aspects of its structure and function have been thoroughly studied. The high-dimensionality of the face is appreciable to humans as we are experts at interpreting meaning from its many configurations and contextual cues. Also, the dimensionality of the face is plausibly related to the dimensionality behind the biological processes that give it its structure and animate it. This is best demonstrated by the relationship between discrete facial configurations and felt emotions and adopting this approach revealed a set of advantages when manipulating systems.

For example, the known effects of seeing a human smiling face on affective state were implemented in chapter 4 as a series of experiments. I tested the extent to which inducing facial contagion could influence affect through emotional contagion as the basis for an intervention for low mood. The results demonstrated its effectiveness in use over long periods of time (i.e. weeks) but effects on mood were not found in the short term (i.e. minutes). The question was not lead by pure curiosity in the methods, as is often the case with neuroimaging, but forced me to implement non-traditional methods

to answer complex questions and generate real-world solutions. The experiments demonstrated how leveraging mobile technology can engage the general public in scientific experimentation and expand the reach and impact of research. As mobile phones are commonly carried by most people in developed and developing countries, their proximity to people allowed for delivery of frequent doses of interventions without disturbing participants' life significantly. These small interruptions during the day were much less intrusive than say, meditating for a half-hour. It also demonstrated the impact of multidisciplinary integration of gamification and behavioural economics "nudge-theory" to generate large datasets within a translational approach. This was the first time that neuroscientific findings were translated into a mood therapy leveraging app mobile technology, it led to a series of good practice lessons detailed in the relevant chapter as well as interesting results that advanced our knowledge of the extent to which established findings can be translated for the greater good.

These studies highlighted the challenges of the high-dimensional approach. They created huge datasets; essential in modelling biological systems from multiple variables while becoming harder to handle. Also, it became less practical to deliver the intervention high-dimensionally and respect inter-individual differences. This was attempted by allowing individuals to control the timing and level of dosages. Although this allowed a certain degree of individual control and rate of delivery of the therapy (i.e. dosage), the scale and method did not allow for greater detail than was implemented in data gathering. Indeed, the potential to individuate effects was complicated by the necessity to keep the questionnaires brief and non-intrusive which also limited the richness of the data. Longer questionnaires might have led to lower participation rates and more elaborate emotion detection techniques, such as photographs of participants complicated the design with a range of ethical considerations which I was unable to implement within the constraints of the current PhD. This is a common challenge across high-dimensional studies and in this chapter, it was inevitable to ultimately rely on self-report.

When a large and rich dataset *is* available the high-dimensional approach can generate interesting insights, as was the case in the last experimental chapter. This presented a good opportunity to use changes in the *structure* of the face to drive translational outputs. Structure and function are intertwined, for the body is adaptive, changing as input-output transformations drive it hither and thither. Here, the faces of Tibetan and Han Chinese participants were parametrised to explore which physiological measures could be predicted with machine learning methods. The prediction of certain

variables from the whole face compared to certain isolated regions confirmed that a high-dimensional parametrisation was superior to single, information-rich areas. In addition, an echocardiographic measure of heart function indicative of pulmonary hypertension, was found to be well predicted from face parameters. This was an original finding that further confirmed the validity of the approach and is a necessary first step in creating a screening tool for echocardiographic markers of cardiovascular diseases in hard to reach populations based on such measures. As is common in methodology-driven approaches, a vast number of improvements could be made to this study to ensure the robustness of the findings. For example, testing the data with a larger data set or one with populations indigenous to other geographies (e.g. Andean mountains) and other machine learning methods could further cement and independently verify the findings. However, the results remain a powerful demonstration of the potential utility of using a data-driven approach coupled with machine learning methods within a translational mindset.

Another advantage of the high-dimensional approach is that it allows for the proper consideration of the relation between the variables that define the variance between individuals. This could in turn generate data-driven models that are more individually translatable. Indeed, complex, multiparameter, adaptive systems are likely to have multiple good solutions. This may result in each individual being adequately describable only by a relatively large set of parameters. To individuate treatment, one therefore has no option but to adopt high-dimensional data-driven modelling, for the intrinsic complexity of the biological necessitates it. Moreover, the more complex a system, the harder it will be to generate a satisfactory model *a priori*. Good methods will naturally tailor themselves to the data, deriving their properties from the data itself in the most flexible way.

The second chapter perhaps exemplified this best, where I adapted this approach to better elucidate the characteristics of the very noisy facial contagion response as measured with EMG. I developed the Pompeii Adaptive Filter (PAF), an automated data-driven filtering method that would overcome many limitations found in conventional filtering. The PAF capitalised on the many studies that characterise the facial contagion response to generate four Meta-CRFs that were used as templates to extract the underlying shape of the signal of interest to each emotional reaction. The process of deconstruction through variational mode decomposition captured the full variability within the signal while remaining highly flexible and assuming very little about the fundamental properties of the data. The resulting signal combinations were tested for

highest similarity with the Meta-CRFs through cross-correlation. Therefore, rather than selecting the relevant data *a priori*, it was defined by the literature in the field and automatically selected by the method in the most flexible manner. The PAF was contrasted to conventional filtering by simulating different EMG noise levels and was found to be superior at each muscle site and for all modelled contagion responses. This was the first time that facial EMG data underwent such complex filtering as the field was quite set in common practice. It also demonstrated the potential of the data-driven translational approach as Meta-CRFs were constructed from historical data and the full high-dimensionality of the entire waveform was used in SVMs to validate the efficacy of the results. This result also exemplified the extrapolative validity of commonly applied techniques in neuroimaging statistical analysis which were the foundational inspiration for this work due to the author's background in the subject matter.

I then moved to explore whether this technique could be successful with biologically acquired data gathered in chapter 3. This chapter's primary goal was to measure facial contagion responses with EMG in MND patients and identify differences between high and low lability groups. High lability MND patients demonstrated weaker and slower starting facial contagion responses with an exaggerated reactivity to negative emotions. This demonstrated the value of translating a neuroscientific process into one with practical applications which, with further exploration, could hold real clinical potential. Indeed, this was the first differential observation made between facial reactivities of MND patients. This invites further focus from the field on the emotional dysregulation aspects of the condition rather than pure concentration on the motor deteriorative symptoms. From a methodological standpoint, this presented an opportunity to test the performance of the PAF on biological data. It did not conclusively perform better than conventional filtering since performance was not consistently superior across muscle sites. Although this could have been due to several factors including small sample size and technical issues, if differences would be further confirmed, the technique of predicting differences between groups by feeding the waveforms into a SVM holds potential as the basis of a differentiator.

Indeed, the PAF method holds potential beyond the current setting and could be used in other signal processing instances. As I took inspiration from previous research in neuroscience that models the haemodynamic response to measure brain activity, I hope this inspires researchers in other neuroscience fields such as neuroimaging to explore whether the PAF can be useful in return. For example, this holds particular potential in electroencephalographic recordings where a prototypical response is well

established for event related potentials in the field of decision making (e.g. P300 and readiness potential). If found to be effective across other domains, it could lead to shorter experiments as less trials might be required to be completed by participants to generate similar signal-to-noise ratio as with conventional filtering. Ideally, clinical practice requires highly individuating tests that can work with few trials otherwise they cannot be applied at all, whether they are good or not. This is an important consideration as a major barrier to introducing behavioural tests in clinical practice is the "sweat-factor": the number of trials needed to get good classification.

This highlights another advantage of the high-dimensional approach, in that it generates models that are more robust. If the individuating information is in the covariance between multiple variables they *may* make determining a biological state more resistant to noise and bias. As stated in the previous section, resistance to noise comes from distributing the individuating information across many variables enabling rejection of noise, which will generally not have the same covariance structure, indeed it will usually be random. Resistance to bias comes from the converse: global effects such as changes in baseline have too simple a covariance structure that tends to produce an affine transform of the data transparent to the biological signal.

Finally, there are some questions that remain to improve operating within this approach. How does one identify a finding that is ready for translation? Once individuated, which issue or disorder should it tackle and what are its limits? Should ethics be adapted to the growing mobile industry that collects more data than academic research with less repercussions and for profit? Should deceit be allowed in mobile experiments where debrief cannot be ensured unless participants request it actively and if they forfeit this right during the consent process? Solving some of these issues will enable the advancement of the field within ethical constraints while providing a needed structure to the process that might mitigate some of the risk of not obtaining positive results.

6.3 Lessons on ways-of-working

Although it is rarely a point of discussion, scientific research is after all a profession and has established ways-of-working. Several lessons from this PhD suggest certain transformations could enable the development of different ways-of-working, which could contribute to advancing the discovery and application process.

In the current thesis, the PhD student had to wear many hats including traditional responsibilities, such as writing ethics and analysing data, but also as a project manager. Indeed, without creating a team of 12 volunteers that were present 6-days a week, for over one month it would have been impossible to collect data from 4116 Science Museum visitors. Without learning agile methodologies and leading user testing, it would not have been possible to publish 4 mobile apps across tablets and smartphones with different operating systems. These types of endeavours break away from a traditional junior researcher role and provide great growth opportunities and transferable skills, useful both inside and outside academia.

Another technology that was leveraged throughout this PhD to generate results were artificial intelligence methods such as support vector machines for classification prediction. The potential for machine learning to change daily activities is increasingly being recognised and research is one of these sectors as many algorithms draw their inspiration from neuronal systems in the brain. It is convenient that in neuroscience large datasets are routinely produced as machine learning methods require these to work effectively. Learning to capture the high-dimensional variability within this data and properly parameterise it has potential to better inform research questions, provide independent lines of evidence, and discover practical applications for research. By demonstrating the versatility of these techniques, the present work hopes to have inspired other neuroscientists to explore their potential in generating new findings outside of computational neuroscience.

Since the goal of clinical practice is to prescribe as well as predict it is important to know how to change a system towards a desired goal. This is easiest when established models are sufficiently simple to be mechanistically interpretable, so that a prediction can be easily made based on the consequences of minor changes. Inevitably this has led to a preference towards fitting the simplest possible models, as trying complex ones raises concerns over their prescriptive potential. However, this hesitation may be misplaced as if prediction requires a complex model then so shall prescription: the established notion that the biological must be simple needs to be exceeded. Simplicity is here the exception rather than the rule, and this complexity needs to be characterised accurately. A solution is found in prescriptive modelling where the models generate the likelihood of desired outcomes for each course of action without requiring the researcher to understand these at all. In this way, the interventions are modelled high-dimensionally and match the complexity with which the systems are described, leading to more precise and tailored interventions.

An emphasis on the importance of the collaborative element in this type of research is required. It is impossible to both generate large datasets and demonstrate versatility in exploration of different topics without collaborations. For example, the team from Imperial College that provided the face photographs were experts at gathering haematological and echocardiographic data at altitude, but never considered photographing participants' faces. By asking for a simple extra measure to be recorded in a very complex and large study, an unprecedented level of detail was uncovered about the relationship between faces and heart function and the potential of this relationship to be translated into clinical screening measures.

If this approach were institutionalised, a new way-of-working might require a potential call for "piggybacking" requests at the pre-ethics stage. For example, prior to submitting a large-scale study for ethical approval, the researchers could open a call for data collection requests based on their measures that another researcher could benefit within their ability to deliver. This would lead to greater publications, inter-disciplinary collaboration, and better use of financial resources. Indeed, if this call could be implemented as a condition to award multidisciplinary or translational grants, it could enhance collaborations and provide the foundation for a new way to generate discoveries. This collaborative element is usually not preferred at the early career stage which is a PhD as the emphasis is on the individual learning to perform scientific research independently. However, when in cases such as the present one, where expertise is required across a variety of domains and the ambitions of the project can perhaps be too daunting to take to commercialisation or publication in the short time available, it is important to consider whether one should prioritise completion of single projects or breadth of topics. In this case, the latter was clearly preferred in order to develop a wide array of skills and provide an interesting basis for further work. If some of these experiments may appear "incomplete" as they may seem to end just when the interest of the reader is at its peak it was due to time restrictions. I believe that for scientific discovery to be successful it should advance the field by answering certain questions but also by raising many more. In this way, there will always be a need for further discovery and the search for answers never ends.

6.4 Conclusions

The largest impact of this thesis was demonstrating the potential of a translational approach in cognitive neuroscience. The parallel translational approach forced me to exit from the traditional pathway to generating science and simultaneously create and

test findings. The experimental uncertainty of this research was justified by weighing the potential impact and benefits of outcomes. Few of the chapters contain follow-up experiments due to the time involved in pursuing each one and there are of course more questions raised than answered in this research.

As a training exercise to become an independent researcher, this approach enabled me to develop innovative solutions across different areas by conducting six experiments that exploited different conceptual aspects of its structure, such as changes in face bones due to environmental pressures, and function, such as facial and emotional contagion. I also applied different data acquisition techniques such as EMG and qualitative mood questionnaires, generated big datasets by integrating experiments within mobile devices such as iPhones or gamifying them into Android tablets, and explored the potential of different analyses techniques and artificial intelligence algorithms to understand ensuing effects.

I believe that collaboration between individuals and institutions will be key to demonstrate the utility of these findings and inform further investigation across cognitive neuroscience and beyond. Thus, it may be that not all the findings presented in this thesis will be published in scientific articles without further work. However, had all experiments yielded positive results I might have been accused of being *overly cautious* at best, or *obvious*, at worst.

The impact that mobile technology demonstrated is forcing a multidisciplinary reconfiguration of how research is conducted within cognitive neuroscience. When I launched the Pocket Smile project, a single app experiment existed as a mobile app in the field: *The Great Brain Experiment* (Brown et al 2014). This app facilitated the identification and screening of participants for further neuroimaging studies through gamified psychophysics tasks. As this was new territory for the field, the researchers were content with replicating canonical research and did not generate new information despite the massive population reached (N>20,000).

More recently however, the apps in the field have become more complex. For example, in *Sea Hero Quest* (Morgan 2016), a sailor tries to re-capture the memories of his father in a sea of mazes as he loses them to dementia. This application aims to create a normative dataset about people's spatial navigational skills, instrumental in creating an early symptoms diagnostic test for dementia. Although the results are still unknown, it is clear the greater understanding of the potential for mobile technology to deliver data from millions of participants is being recognised to generate important translational advances and explore the potential impact on people's lives. Conducting

these types of projects will require a shift in funding structure and culture. Specifically, granting bodies will need to understand the importance of creating a promotional video, allocating funds for managing a social media presence, and understanding that development costs may increase at different stages to ensure the success of such projects.

In the future, obtaining data from multiple distributed devices from the same individual may provide richer and more objective reliable information than currently possible. This will become most relevant when delivering therapy at scale through distributed interventions as specific information and individualised feedback will be possible in this way. This type of progress may redefine the connection between knowledge and intervention and close the gap between data gathering and implementation.

Ultimately, the translational approach to research proved to be impactful and testing the uses of research in parallel as producing new findings holds potential to speed up the applicability of science. The approach also demonstrated that it was a high-risk endeavour where generating big effects was difficult and sensitivity to noise was high. Translating different aspects of established research about the face was proven to be useful in a variety of ways in the context of neuroscience research. The research conducted was also of interest to the wider public, with one of the projects being featured in a major UK newspaper (www.telegraph.co.uk/science/2016/12/25/smile-pocket-could-help-ease-depression-anxiety/) and internationally on German radio (www.dradiowissen.de/beitrag/pocketsmile-laecheln-als-app-therapie). The aim was to test the value of the translational and high-dimensional approach to neuroscientific research and the assessment concludes it is valid. I hope other researchers and institutions will integrate some of the findings within their work to extend their impact to those affected by neurological disorders, the general public, and advance progress in scientific research.

References

Abel, E.L., Kruger, M.L. (2010) “Smile intensity in photographs predicts longevity,” *Psychological Science*, 21(4), 542–544.

Achaibou, A., Pourtois, G., Schwartz, S., Vuilleumier, P. (2008) “Simultaneous recording of EEG and facial muscle reactions during spontaneous emotional mimicry,” *Neuropsychologia*, 46(4), 1104–1113.

Adams, V. (2016) *Metrics: What Counts in Global Health*, Duke University Press.

Agbo, A.A. (2016) “The validation of the International Positive and Negative Affect Schedule-Short Form in Nigeria,” *South African Journal of Psychology*, 0081246316630081.

Aguado, L., Román, F.J., Rodr'\iguez, S., Diéguez-Risco, T., Romero-Ferreiro, V., Fernández-Cahill, M. (2013) “Learning of facial responses to faces associated with positive or negative emotional expressions,” *The Spanish journal of psychology*, 16, E24.

Aguirre, G.K. (2012) Number of Neurons in a Voxel [online], available: https://cfn.upenn.edu/aguirre/wiki/public:neurons_in_a_voxel [accessed 8 2016].

Allemann, Y., Sartori, C., Lepori, M., Pierre, S., Mélot, C., Naeije, R., Scherrer, U., Maggiorini, M. (2000) “Echocardiographic and invasive measurements of pulmonary artery pressure correlate closely at high altitude,” *American Journal of Physiology-Heart and Circulatory Physiology*, 279(4), H2013–H2016.

Alloy, L.B., Kelly, K.A., Mineka, S., Clements, C.M. (1990) “Comorbidity of anxiety and depressive disorders: a helplessness-hopelessness perspective.”

Alpaydin, E. (2014) *Introduction to Machine Learning*, MIT press.

American Psychiatric Association (2000) *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (text Revision)*, American Psychiatric Publishing, Inc.

Anderson, A.K., Christoff, K., Panitz, D., De Rosa, E., Gabrieli, J.D. (2003) “Neural correlates of the automatic processing of threat facial signals,” *Journal of Neuroscience*, 23(13), 5627–5633.

Anderson, C. (2008) “The end of theory,” *Wired magazine*, 16(7), 7–16.

Angst, J., Gamma, A., Gastpar, M., Lépine, J.-P., Mendlewicz, J., Tylee, A. (2002) “Gender differences in depression,” *European archives of psychiatry and clinical neuroscience*, 252(5), 201–209.

Anon (2014) “New dimensions in immunology,” *Nature Biotechnology*, 32(145), 10.

Arakawa, D. R. (2012). *Emotional contagion and its relationship to mood* (Doctoral dissertation, University of Hawaii at Manoa, Honolulu, HI). Available from ProQuest Dissertations and Theses database. (UMI No. 3569048)

Ardizzi, M., Sestito, M., Martini, F., Umiltà, M.A., Ravera, R., Gallese, V. (2014) “When age matters: differences in facial mimicry and autonomic responses to peers’ emotions in teenagers and adults,” *PloS one*, 9(10), e110763.

Arichi T., Fagiolo G., Varela M., Melendez-Calderon A., Allievi A., Merchant N., Tusor N., Counsell S.J., Burdet E., Beckmann C., F., David E.A. (2012) Development of BOLD signal hemodynamic responses in the human brain. *NeuroImage*, *63*, 663–673.

Arno, A., Thomas, S. (2016) “The efficacy of nudge theory strategies in influencing adult dietary behaviour: a systematic review and meta-analysis,” *BMC Public Health*, 16(1), 676.

Ashburner, J., Klöppel, S. (2011) “Multivariate models of inter-subject anatomical variability,” *Neuroimage*, 56(2), 422–439.

Bach, D.R., Flandin, G., Friston, K.J., Dolan, R.J. (2009) “Time-series analysis for rapid event-related skin conductance responses,” *Journal of neuroscience methods*, 184(2), 224–234.

Bach, D.R., Flandin, G., Friston, K.J., Dolan, R.J. (2010) “Modelling event-related skin conductance responses,” *International Journal of Psychophysiology*, 75(3), 349–356.

Bach, D.R., Friston, K.J. (2013) “Model-based analysis of skin conductance responses: Towards causal models in psychophysiology,” *Psychophysiology*, 50(1), 15–22.

Bach, D.R., Friston, K.J., Dolan, R.J. (2013) “An improved algorithm for model-based analysis of evoked skin conductance responses,” *Biological psychology*, 94(3), 490–497.

Barer, G.R., Bee, D., Wach, R.A. (1983) “Contribution of polycythaemia to pulmonary hypertension in simulated high altitude in rats.,” *The Journal of physiology*, 336, 27.

Barker, A.T., Jalinous, R., Freeston, I.L. (1985) “Non-invasive magnetic stimulation of human motor cortex,” *The Lancet*, 325(8437), 1106–1107.

Barrett, L.F., Robin, L., Pietromonaco, P.R., Eyssell, K.M. (1998) “Are women the ‘more emotional’ sex? Evidence from emotional experiences in social context,” *Cognition \& Emotion*, 12(4), 555–578.

Bartol, V. (2004) *Alamut*, Scala House Press: Seattle, United States.

Basmajian, J.V., Stecko, G. (1962) “A new bipolar electrode for electromyography,” *Journal of Applied Physiology*, 17(5), 849–849.

Bavelas, J.B., Black, A., Lemery, C.R., Mullett, J. (1986) “‘I show how you feel’: Motor mimicry as a communicative act.,” *Journal of personality and social psychology*, 50(2), 322.

Beall, P.M., Moody, E.J., McIntosh, D.N., Hepburn, S.L., Reed, C.L. (2008) “Rapid facial reactions to emotional facial expressions in typically developing children and children with autism spectrum disorder,” *Journal of experimental child psychology*, 101(3), 206–223.

Bechtel, W., Hamilton, A. (2007) “Reduction, integration, and the unity of science: Natural, behavioral, and social sciences and the humanities,” *General philosophy of science: Focal issues*, 377–430.

Bedard, C., Kroeger, H., Destexhe, A. (2006) “Does the 1/f frequency scaling of brain signals reflect self-organized critical states?,” *Physical review letters*, 97(11), 118102.

Bedford, E. (1956) “Emotions,” in *Proceedings of the Aristotelian Society*, 281–304.

Behavioural Insights Team (2014) *EAST: Four Simple Ways to Apply Behavioural Insights*.

Behavioural Insights Team (2015) *The Behavioural Insights Team Update Report 2013-15*.

Bellman, R.E. (2015) *Adaptive Control Processes: A Guided Tour*, Princeton university press.

Bensimon, G., Lacomblez, L., Meininger, V. ft (1994) “A controlled trial of riluzole in amyotrophic lateral sclerosis,” *New England Journal of Medicine*, 330(9), 585–591.

Birnbaum, H.G., Kessler, R.C., Kelley, D., Ben-Hamadi, R., Joish, V.N., Greenberg, P.E. (2010) “Employer burden of mild, moderate, and severe major depressive disorder: mental health services utilization and costs, and work performance,” *Depression and anxiety*, 27(1), 78–89.

Björk, A. (1955) “Cranial base development: a follow-up x-ray study of the individual variation in growth occurring between the ages of 12 and 20 years and its relation to brain case and face development,” *American Journal of Orthodontics*, 41(3), 198–225.

Blomdahl, C., Gunnarsson, A.B., Guregård, S., Björklund, A. (2013) “A realist review of art therapy for clients with depression,” *The Arts in Psychotherapy*, 40(3), 322–330.

Bloom, F.E. (2001) “What does it all mean to you?,” *The journal of Neuroscience*, 21(21), 8304–8305.

Bonini, L. (2016) “The Extended Mirror Neuron Network Anatomy, Origin, and Functions,” *The Neuroscientist*.

Boucsein, W. (2012) *Electrodermal Activity*, Springer Science & Business Media.

Bower, P., Richards, D., Lovell, K. (2001) “The clinical and cost-effectiveness of self-help treatments for anxiety and depressive disorders in primary care: a systematic review.,” *Br J Gen Pract*, 51(471), 838–845.

Van Boxtel, A. (2001) “Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles,” *Psychophysiology*, 38(01), 22–34.

Van Boxtel, A. (2010) “Facial EMG as a tool for inferring affective states,” in *Proceedings of Measuring Behavior*, 104–108.

Brady, E.U., Kendall, P.C. (1992) “Comorbidity of anxiety and depression in children and adolescents.,” *Psychological bulletin*, 111(2), 244.

Breiman., L. (2001). "Random forests". *Machine learning*, *45*(1), 5–32.

Brett, M., Johnsrude, I.S., Owen, A.M. (2002) “The problem of functional localization in the human brain,” *Nature reviews neuroscience*, 3(3), 243–249.

Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M. (2010) “The balanced accuracy and its posterior distribution,” in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 3121–3124.

Brooks, B.R. (1994) “El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis,” *Journal of the neurological sciences*, 124, 96–107.

Brown, H.R., Zeidman, P., Smittenaar, P., Adams, R.A., McNab, F., Rutledge, R.B., Dolan, R.J. (2014) “Crowdsourcing for cognitive science-the utility of smartphones,” *PloS one*, 9(7), e100662.

Bühler, C. (1931) *The Social Behavior of the Child*.

Burns, M.N., Begale, M., Duffecy, J., Gergle, D., Karr, C.J., Giangrande, E., Mohr, D.C. (2011) “Harnessing context sensing to develop a mobile intervention for depression,” *Journal of medical Internet research*, 13(3), e55.

Buttner, A.P., Strasser, R. (2014) “Contagious yawning, social cognition, and arousal: an investigation of the processes underlying shelter dogs’ responses to human yawns,” *Animal cognition*, 17(1), 95–104.

Böning, D., Maassen, N., Pries, A. (2011) “The Hematocrit Paradox-How Does Blood Doping Really Work?,” *International journal of sports medicine*, 32(04), 242–246.

Cacioppo, J.T., Martzke, J.S., Petty, R.E., Tassinary, L.G. (1988) “Specific forms of facial EMG response index emotions during an interview: From Darwin to the continuous flow hypothesis

of affect-laden information processing.," *Journal of personality and social psychology*, 54(4), 592.

Calder, A. (2011) *Oxford Handbook of Face Perception*, Oxford University Press.

Cannon, W.B. (1927) "The James-Lange theory of emotions: A critical examination and an alternative theory," *The American journal of psychology*, 39(1/4), 106–124.

De Carvalho, M., Dengler, R., Eisen, A., England, J.D., Kaji, R., Kimura, J., Mills, K., Mitsumoto, H., Nodera, H., Shefner, J., others (2008) "Electrodiagnostic criteria for diagnosis of ALS," *Clinical neurophysiology*, 119(3), 497–503.

Carvalho, M.D., Swash, M. (2009) "Awaji diagnostic algorithm increases sensitivity of El Escorial criteria for ALS diagnosis," *Amyotrophic Lateral Sclerosis*, 10(1), 53–57.

Chang, C., Lin, C. (2016) "LIBSVM: a Library for Support Vector Machines.," available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Chang, C.-C., Lin, C.-J. (2011) "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 27.

Charcot, J. (1874) "De la sclérose latérale amyotrophique," *Prog Med*, 2(325), 341–453.

Chartrand, T.L., Bargh, J.A. (1999) "The chameleon effect: The perception-behavior link and social interaction.," *Journal of personality and social psychology*, 76(6), 893.

Chen, D., Xie, S. (2004) *Review of the Control Strategies Applied to Active Power Filters*, Electric Utility Deregulation, Restructuring and Power Technologies , IEEE.

Chiò, A., Logroscino, G., Hardiman, O., Swingler, R., Mitchell, D., Beghi, E., Traynor, B.G., Consortium, E., others (2009) "Prognostic factors in ALS: a critical review," *Amyotrophic Lateral Sclerosis*, 10(5-6), 310–323.

Chowdhury, R.H., Reaz, M.B., Ali, M.A.B.M., Bakar, A.A., Chellappan, K., Chang, T.G. (2013a) "Surface electromyography signal processing and classification techniques," *Sensors*, 13(9), 12431–12466.

Chowdhury, R.H., Reaz, M.B., Ali, M.A.B.M., Bakar, A.A., Chellappan, K., Chang, T.G. (2013b) "Surface electromyography signal processing and classification techniques," *Sensors*, 13(9), 12431–12466.

Christensen, P.B., Højer-Pedersen, E., Niels Bjørn Jensen, M. (1990) "Survival of patients with amyotrophic lateral sclerosis in 2 Danish counties," *Neurology*, 40(4), 600–600.

Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., Lin, C., Initiative, A.D.N., others (2012) "Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images," *Neuroimage*, 60(1), 59–70.

Ciccarelli, O., Behrens, T., Altmann, D., Orrell, R., Howard, R., Johansen-Berg, H., Miller, D., Matthews, P., Thompson, A. (2006) "Probabilistic diffusion tractography: a potential tool to assess the rate of disease progression in amyotrophic lateral sclerosis," *Brain*, 129(7), 1859–1871.

Cisek, P. (2007) "Cortical mechanisms of action selection: the affordance competition hypothesis," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1485), 1585–1599.

Clancy, E., Morin, E.L., Merletti, R. (2002) "Sampling, noise-reduction and amplitude estimation issues in surface electromyography," *Journal of Electromyography and Kinesiology*, 12(1), 1–

16.

Clark, V.L., Kruse, J.A. (1990) “Clinical methods: the history, physical, and laboratory examinations,” *JAMA*, 264(21), 2808–2809.

Contreras-Rodriguez, O., Pujol, J., Batalla, I., Harrison, B.J., Bosque, J., Ibern-Regàs, I., Hernández-Ribas, R., Soriano-Mas, C., Deus, J., López-Solà, M., others (2014) “Disrupted neural processing of emotional faces in psychopathy,” *Social cognitive and affective neuroscience*, 9(4), 505–512.

Cooper, L.A., Gonzales, J.J., Gallo, J.J., Rost, K.M., Meredith, L.S., Rubenstein, L.V., Wang, N.-Y., Ford, D.E. (2003) “The acceptability of treatment for depression among African-American, Hispanic, and white primary care patients,” *Medical care*, 41(4), 479–489.

Cortes, C., Vapnik, V. (1995) “Support-vector networks,” *Machine learning*, 20(3), 273–297.

Coward, H.R., Franklin, J.J. (1989) “Identifying the science-technology interface: matching patent data to a bibliometric model,” *Science, Technology, \& Human Values*, 14(1), 50–77.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.-O., Chupin, M., Benali, H., Colliot, O., Initiative, A.D.N., others (2011) “Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database,” *neuroimage*, 56(2), 766–781.

Danchin, N., Cornette, A., Henriquez, A., Godenir, J., Ethevenot, G., Polu, J., Sadoul, P. (1987) “Two-dimensional echocardiographic assessment of the right ventricle in patients with chronic obstructive lung disease,” *Chest*, 92(2), 229–233.

Darwin, C. (1965) *The Expression of the Emotions in Man and Animals*, University of Chicago press.

Daube, J., Rubin, D. (2009) “clinical neurophysiology. 3 rd edit.”

Davenport, R., Swingler, R., Chancellor, A., Warlow, C. (1996) “Avoiding false positive diagnoses of motor neuron disease: lessons from the Scottish Motor Neuron Disease Register.,” *Journal of Neurology, Neurosurgery \& Psychiatry*, 60(2), 147–151.

Davila-Ross, M., Allcock, B., Thomas, C., Bard, K.A. (2011) “Aping expressions? Chimpanzees produce distinct laugh types when responding to laughter of others.,” *Emotion*, 11(5), 1013.

Davis, J.I., Senghas, A., Brandt, F., Ochsner, K.N. (2010) “The effects of BOTOX injections on emotional experience.,” *Emotion*, 10(3), 433.

Deschamps, P., Schutte, I., Kenemans, J., Matthys, W., Schutter, D. (2012) “Electromyographic responses to emotional facial expressions in 6-7year olds: A feasibility study,” *International Journal of Psychophysiology*, 85(2), 195–199.

Dimberg, U. (1982) “Facial reactions to facial expressions,” *Psychophysiology*, 19(6), 643–647.

Dimberg, U. (1986) “Facial reactions to fear-relevant and fear-irrelevant stimuli,” *Biological psychology*, 23(2), 153–161.

Dimberg, U. (1987) “Facial reactions, autonomic activity and experienced emotion: A three component model of emotional conditioning,” *Biological psychology*, 24(2), 105–122.

Dimberg, U., Lundqvist, L. (1990) “Gender differences in facial reactions to facial expressions,” *Biological Psychology*, 30(2), 151–159.

Dimberg, U., Thell, S. (1988) “Facial electromyography, fear relevance and the experience of stimuli.,” *Journal of Psychophysiology*.

Dimberg, U., Thunberg, M. (1998) “Rapid facial reactions to emotional facial expressions,” *Scandinavian journal of psychology*, 39(1), 39–45.

Dimberg, U., Thunberg, M. (2007) “Speech anxiety and rapid emotional reactions to angry and happy facial expressions,” *Scandinavian journal of psychology*, 48(4), 321–328.

Dimberg, U., Thunberg, M. (2012) “Empathy, emotional contagion, and rapid facial reactions to angry and happy facial expressions,” *PsyCh Journal*, 1(2), 118–127.

Dimberg, U., Thunberg, M., Elmehed, K. (2000) “Unconscious facial reactions to emotional facial expressions,” *Psychological science*, 11(1), 86–89.

Dragomiretskiy, K., Zosso, D. (2014) “Variational mode decomposition,” *IEEE transactions on signal processing*, 62(3), 531–544.

Du, S., Martinez, A.M. (2011) “The resolution of facial expressions of emotion,” *Journal of Vision*, 11(13), 24–24.

Duchenne, D.B. (1862) “The mechanism of human facial expression or an electro-physiological analysis of the expression of the emotions (A. Cuthbertson, Trans.).”

Duque, A., Vázquez, C. (2015) “Double attention bias for positive and negative emotional faces in clinical depression: Evidence from an eye-tracking study,” *Journal of Behavior Therapy and Experimental Psychiatry*, 46, 107–114.

Durik, A.M., Hyde, J.S., Marks, A.C., Roy, A.L., Anaya, D., Schultz, G. (2006) “Ethnicity and gender stereotypes of emotion,” *Sex Roles*, 54(7-8), 429–445.

Ekman, P. (1989) “The argument and evidence about universals in facial expressions,” *Handbook of social psychophysiology*, 143–164.

Ekman, P., Friesen, W.V., Ellsworth, P. (1972) *Emotion in the Human Face: Guidelines for Research and an Integration of Findings: Guidelines for Research and an Integration of Findings*, Pergamon.

Ellis, C., Simmons, A., Jones, D.K., Bland, J., Dawson, J., Horsfield, M., Williams, S., Leigh, P. (1999) “Diffusion tensor MRI assesses corticospinal tract damage in ALS,” *Neurology*, 53(5), 1051–1051.

Else-Quest, N.M., Higgins, A., Allison, C., Morton, L.C. (2012) “Gender differences in self-conscious emotional experience: a meta-analysis.,” *Psychological bulletin*, 138(5), 947.

Engell, A., Gobbini, M., Haxby, J. (2006) “Gaze-change perception in the early visual cortex,” in *Poster Presented at the Annual Meeting of the Society for Neuroscience*.

Enticott, P.G., Kennedy, H.A., Johnston, P.J., Rinehart, N.J., Tonge, B.J., Taffe, J.R., Fitzgerald, P.B. (2014) “Emotion recognition of static and dynamic faces in autism spectrum disorder,” *Cognition and emotion*, 28(6), 1110–1118.

Van Esbroeck, A., Smith, L., Syed, Z., Singh, S., Karam, Z. (2016) “Multi-task seizure detection: addressing intra-patient variation in seizure morphologies,” *Machine Learning*, 102(3), 309–321.

Farah, M.J., Wilson, K.D., Drain, M., Tanaka, J.N. (1998) “What is‘ special’ about face perception?,” *Psychological review*, 105(3), 482.

Feinstein, A., Feinstein, K., Gray, T., O’Connor, P. (1997) “Prevalence and neurobehavioral correlates of pathological laughing and crying in multiple sclerosis,” *Archives of neurology*, 54(9), 1116–1121.

Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Abdel-Aziz, K., Hamdan, H. (2016) “A machine learning system for automated whole-brain seizure detection,” *Applied Computing and Informatics*, 12(1), 70–89.

Ferry, Q., Steinberg, J., Webber, C., FitzPatrick, D.R., Ponting, C.P., Zisserman, A., Nellåker, C. (2014) “Diagnostically relevant facial gestalt information from ordinary photos,” *Elife*, 3, e02020.

Finzi, E., Rosenthal, N.E. (2014) “Treatment of depression with onabotulinumtoxinA: a randomized, double-blind, placebo controlled trial,” *Journal of psychiatric research*, 52, 1–6.

Finzi, E., Wasserman, E. (2006) “Treatment of depression with botulinum toxin A: a case series,” *Dermatologic Surgery*, 32(5), 645–650.

Fischer, A.H. (1993) “Sex differences in emotionality: Fact or stereotype?,” *Feminism \& Psychology*, 3(3), 303–318.

Fischer, K.W., Shaver, P.R., Carnochan, P. (1990) “How emotions develop and how they organise development,” *Cognition and emotion*, 4(2), 81–127.

Fisher, K.M., Zaaimi, B., Williams, T.L., Baker, S.N., Baker, M.R. (2012) “Beta-band intermuscular coherence: a novel biomarker of upper motor neuron dysfunction in motor neuron disease,” *Brain*, aws150.

Floeter, M.K., Katipally, R., Kim, M.P., Schanz, O., Stephen, M., Danielian, L., Wu, T., Huey, E.D., Meoded, A. (2014) “Impaired corticopontocerebellar tracts underlie pseudobulbar affect in motor neuron disorders,” *Neurology*, 83(7), 620–627.

Floyd, A., Yu, Q., Piboolnurak, P., Tang, M., Fang, Y., Smith, W., Yim, J., Rowland, L., Mitsumoto, H., Pullman, S. (2009) “Transcranial magnetic stimulation in ALS Utility of central motor conduction tests,” *Neurology*, 72(6), 498–504.

Fridlund, A.J., Cacioppo, J.T. (1986) “Guidelines for human electromyographic research,” *Psychophysiology*, 23(5), 567–589.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. B., Frith, C. D., & Frackowiak, R. S. J. (1995) Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210.

Friston K.J. (2005). Models of brain function in neuroimaging. *Annual Review Psychology*, 56, 57–87.

Gagnon, D.R., Zhang, T.-J., Brand, F.N., Kannel, W.B. (1994) “Hematocrit and the risk of cardiovascular disease—the Framingham study: a 34-year follow-up,” *American heart journal*, 127(3), 674–682.

Gallagher, J. (1989) “Pathologic laughter and crying in ALS: a search for their origin,” *Acta Neurologica Scandinavica*, 80(2), 114–117.

Gao, W., Tan, J., Hüls, A., Ding, A., Liu, Y., Matsui, M.S., Vierkötter, A., Krutmann, J., Schikowski, T., Jin, L., others (2016) “Genetic variants associated with skin aging in the Chinese Han population,” *Journal of Dermatological Science*.

Gelder, B. de, Vroomen, J., Van der Heide, L. (1991) “Face recognition and lip-reading in autism,” *European Journal of Cognitive Psychology*, 3(1), 69–86.

George, E. (2013) “HbE β-Thalassaemia in Malaysia: Revisited,” *Journal of Hematology \& Thromboembolic Diseases*.

George, P.A., Hole, G.J. (2000) “The role of spatial and surface cues in the age-processing of unfamiliar faces,” *Visual Cognition*, 7(4), 485–509.

Georgoulopoulou, E., Fini, N., Vinceti, M., Monelli, M., Vacondio, P., Bianconi, G., Sola, P., Nichelli, P., Mandrioli, J. (2013) “The impact of clinical factors, riluzole and therapeutic interventions on ALS survival: a population based study in Modena, Italy,” *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 14(5-6), 338–345.

Gibson, J.J. (1977) *The Theory of Affordances* , USA.

Gilbody, S., Richards, D., Brealey, S., Hewitt, C. (2007) “Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis,” *Journal of general internal medicine*, 22(11), 1596–1602.

Gitelman D.R., Penny W.D., Ashburner J., & Friston K.J. (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. *NeuroImage*, *19*(200-207.

Goldstein, M. (2008) “Carbon monoxide poisoning,” *Journal of Emergency Nursing*, 34(6), 538–542.

Gorman, D., Clayton, D., Gilligan, J., Webb, R. (1992) “A longitudinal study of 100 consecutive admissions for carbon monoxide poisoning to the Royal Adelaide Hospital.,” *Anaesthesia and intensive care*, 20(3), 311–316.

Gueguen, N., De Gail, M.-A. (2003) “The effect of smiling on helping behavior: Smiling and good Samaritan behavior,” *Communication reports*, 16(2), 133–140.

Gunn, D.A., Larsen, L.A., Lall, J.S., Rexbye, H., Christensen, K. (2015) “Mortality is Written on the Face,” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, glv090.

Guyatt, G.H., Keller, J.L., Jaeschke, R., Rosenbloom, D., Adachi, J.D., Newhouse, M.T. (1990) “The n-of-1 randomized controlled trial: clinical usefulness: our three-year experience,” *Annals of Internal Medicine*, 112(4), 293–299.

Hristo., Z. (2016) “Pink Noise Generation with MATLAB Implementation.”

Hacker, P. (2015) *The Dialetic of the Emotions*.

Hacker, P.M. (2009) “The conceptual framework for the investigation of emotions,” in *Emotions and Understanding*, Springer, 43–59.

Halpern, D. (2015) *Inside the Nudge Unit: How Small Changes Can Make a Big Difference*, Random House.

Hansen, P.G., Skov, L.R., Skov, K.L. (2016) “Making healthy choices easier: regulation versus nudging,” *Annual review of public health*, 37, 237–251.

Harker, L., Keltner, D. (2001) “Expressions of positive emotion in women’s college yearbook pictures and their relationship to personality and life outcomes across adulthood.,” *Journal of personality and social psychology*, 80(1), 112.

Harrison, N.A., Morgan, R., Critchley, H.D. (2010) “From facial mimicry to emotional empathy: A role for norepinephrine?,” *Social neuroscience*, 5(4), 393–400.

Hart, T., Hart, P. (2009) “Genetic studies of craniofacial anomalies: clinical implications and applications,” *Orthodontics \& craniofacial research*, 12(3), 212–220.

Hartmann, P., Apaolaza, V., Eisend, M. (2016) “Nature Imagery in Non-Green Advertising: The Effects of Emotion, Autobiographical Memory, and Consumer’s Green Traits,” *Journal of*

Advertising, 1–14.

Hatfield, E., Cacioppo, J.T., Rapson, R.L. (1994) *Emotional Contagion*, Cambridge university press.

Hatfield, E., Bensman, L., Thornton, P. D., Rapson, L. (2014a) " New Perspectives on Emotional Contagion: A Review of Classic and Recent Research on Facial Mimicry and Contagion", *Interpersona*, 8(2), 159-179.

Hatfield, E., Carpenter, M., & Rapson, R. L. (2014b). Emotional contagion as a precursor to collective emotions. In C. von Scheve & M. Salmela (Eds.), *Collective emotions: Perspectives from psychology, philosophy, and sociology* (pp. 108-122). Oxford, United Kingdom: Oxford University Press.

Haverkamp, L.J., Appel, V., Appel, S.H. (1995) "Natural history of amyotrophic lateral sclerosis in a database population Validation of a scoring system and a model for survival prediction," *Brain*, 118(3), 707–719.

Haxby, J.V., Gobbini, M.I. (2011) "The Oxford Handbook of Face Perception," in Gillian Rhodes, Andy Calder, Mark Johnson, and James V. Haxby, ed.

Heerey, E.A., Crossley, H.M. (2013) "Predictive and reactive mechanisms in smile reciprocity," *Psychological science*, 0956797612472203.

Hermans, E.J., Putman, P., Van Honk, J. (2006) "Testosterone administration reduces empathetic behavior: a facial mimicry study," *Psychoneuroendocrinology*, 31(7), 859–866.

Hermens, H.J., Freriks, B., Disselhorst-Klug, C., Rau, G. (2000) "Development of recommendations for SEMG sensors and sensor placement procedures," *Journal of electromyography and Kinesiology*, 10(5), 361–374.

Hermens, H.J., Freriks, B., Merletti, R., Stegeman, D., Blok, J., Rau, G., Disselhorst-Klug, C., Hägg, G. (1999) "European recommendations for surface electromyography," *Roessingh Research and Development*, 8(2), 13–54.

Hertenstein, M.J., Hansel, C.A., Butts, A.M., Hile, S.N. (2009) "Smile intensity in photographs predicts divorce later in life," *Motivation and Emotion*, 33(2), 99–105.

Hess, U., Blairy, S. (2001) "Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy," *International journal of psychophysiology*, 40(2), 129–141.

Hess, U., Fischer, A. (2014) "Emotional mimicry: Why and when we mimic emotions," *Social and Personality Psychology Compass*, 8(2), 45–57.

Hess, U., Senécal, S., Kirouac, G., Herrera, P., Philippot, P., Kleck, R.E. (2000) "Emotional expressivity in men and women: Stereotypes and self-perceptions," *Cognition \& Emotion*, 14(5), 609–642.

Hind, D., Cotter, J., Thake, A., Bradburn, M., Cooper, C., Isaac, C., House, A. (2014) "Cognitive behavioural therapy for the treatment of depression in people with multiple sclerosis: a systematic review and meta-analysis," *BMC psychiatry*, 14(1), 1.

Hinsz, V.B., Tomhave, J.A. (1991) "Smile and (half) the world smiles with you, frown and you frown alone," *Personality and Social Psychology Bulletin*, 17(5), 586–592.

Hjortsjö, C.-H. (1969) *Man's Face and Mimic Language*, Studen litteratur.

Hofree, G., Ruvolo, P., Bartlett, M.S., Winkielman, P. (2014) "Bridging the mechanical and the human mind: spontaneous mimicry of a physically present android," *PloS one*, 9(7), e99934.

Hollands, G.J., Shemilt, I., Marteau, T.M., Jebb, S.A., Kelly, M.P., Nakamura, R., Suhrcke, M., Ogilvie, D. (2013) *Altering Choice Architecture to Change Population Health Behaviour: A Large-Scale Conceptual and Empirical Scoping Review of Interventions within Micro-Environments*.

Hong, Y.-H., Lee, K.-W., Sung, J.-J., Chang, K.-H., Song, I.C. (2004) "Diffusion tensor MRI as a diagnostic tool of upper motor neuron involvement in amyotrophic lateral sclerosis," *Journal of the neurological sciences*, 227(1), 73–78.

Hoppe, F. (1857) "Über die Einwirkung des Kohlenoxydgases auf das Hämatoglobulin," *Virchows Archiv*, 11(3), 288–289.

Hsu, C.-C., Hsu, Y.-C., Chang, K.-H., Lee, C.-Y., Chong, L.-W., Lin, C.-L., Shang, C.-S., Sung, F.-C., Kao, C.-H. (2015) "Depression and the Risk of Peptic Ulcer Disease: A Nationwide Population-Based Study," *Medicine*, 94(51).

Ironside, R. (1956) "Disorders of laughter due to brain lesions," *Brain*, 79(4), 589–609.

Iwata, N.K., Kwan, J.Y., Danielian, L.E., Butman, J.A., Tovar-Moll, F., Bayat, E., Floeter, M.K. (2011) "White matter alterations differ in primary lateral sclerosis and amyotrophic lateral sclerosis," *Brain*, 134(9), 2642–2655.

Jagtap, J., Kokare, M. (2016) "Human age classification using facial skin aging features and artificial neural network," *Cognitive Systems Research*, 40, 116–128.

Jain, G., Roy, A., Harikrishnan, V., Yu, S., Dabbous, O., Lawrence, C. (2013) "Patient-reported depression severity measured by the PHQ-9 and impact on work productivity: results from a survey of full-time employees in the United States," *Journal of occupational and environmental medicine*, 55(3), 252–258.

James, W. (1884) "II.—What is an emotion?," *Mind*, (34), 188–205.

James, W. (1894) "Discussion: The physical basis of emotion.," *Psychological Review*, 1(5), 516.

Jeong, Y.-J., Hong, S.-C., Lee, M.S., Park, M.-C., Kim, Y.-K., Suh, C.-M. (2005) "Dance movement therapy improves emotional responses and modulates neurohormones in adolescents with mild depression," *International journal of neuroscience*, 115(12), 1711–1720.

Jonas, E., Kording, K. (2016) "Could a neuroscientist understand a microprocessor?," *bioRxiv*.

Jones, A.L., Kramer, R.S., Ward, R. (2012) "Signals of personality and health: The contributions of facial shape, skin texture, and viewing angle.," *Journal of Experimental Psychology: Human Perception and Performance*, 38(6), 1353.

Jones, M.C. (1926) "The development of early behavior patterns in young children," *The Pedagogical Seminary and Journal of Genetic Psychology*, 33(4), 537–585.

Jones, S.S. (2009) "The development of imitation in infancy," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1528), 2325–2335.

De Jong, P.J., Peters, M., Vanderhallen, I. (2002) "Disgust and disgust sensitivity in spider phobia:: Facial emg in response to spider and oral disgust imagery," *Journal of anxiety disorders*, 16(5), 477–493.

Jongkreangkrai, C., Vichianin, Y., Tocharoenchai, C., Arimura, H., Initiative, A.D.N. (2016) "Computer-aided classification of Alzheimer's disease based on support vector machine with combination of cerebral image features in MRI," in *Journal of Physics: Conference Series*, 012036.

Kanwisher, N., McDermott, J., Chun, M.M. (1997) "The fusiform face area: a module in human extrastriate cortex specialized for face perception," *The Journal of neuroscience*, 17(11),

4302–4311.

Kaplan, S. (1992) “Environmental preference in a knowledge-seeking, knowledge-using organism.”

Karim, J., Weisz, R., Rehman, S.U. (2011) “International positive and negative affect schedule short-form (I-PANAS-SF): Testing for factorial invariance across cultures,” *Procedia-Social and Behavioral Sciences*, 15, 2016–2022.

Katz, J.E., Rice, R.E. (2009) “Public views of mobile medical devices and services: A US national survey of consumer sentiments towards RFID healthcare technology,” *International journal of medical informatics*, 78(2), 104–114.

Kauer, S.D., Reid, S.C., Crooke, A.H.D., Khor, A., Hearps, S.J.C., Jorm, A.F., Sanci, L., Patton, G. (2012) “Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial,” *Journal of medical Internet research*, 14(3), e67.

Kesić, S. (2015) “Systems biology, emergence and antireductionism,” *Saudi Journal of Biological Sciences*.

Kessler, R.C., Avenevoli, S., Merikangas, K.R. (2001) “Mood disorders in children and adolescents: an epidemiologic perspective,” *Biological psychiatry*, 49(12), 1002–1014.

Kim, J.S. (1997) “Pathological laughter and crying in unilateral stroke.,” *Stroke; a journal of cerebral circulation*, 28(11), 2321.

Kirsch, I., Deacon, B.J., Huedo-Medina, T.B., Scoboria, A., Moore, T.J., Johnson, B.T. (2008) “Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration,” *PLoS Med*, 5(2), e45.

Konrad, P. (2005) “The abc of emg,” *A practical introduction to kinesiological electromyography*, 1, 30–35.

Kraft, T.L., Pressman, S.D. (2012) “Grin and bear it the influence of manipulated facial expression on the stress response,” *Psychological science*, 23(11), 1372–1378.

Kreifeldt, J.G. (1971) “Signal versus noise characteristics of filtered EMG used as a control source,” *IEEE Transactions on Biomedical Engineering*, (1), 16–22.

Kroenke, K., Spitzer, R.L., Williams, J.B. (2001) “The Phq-9,” *Journal of general internal medicine*, 16(9), 606–613.

Krumhuber, E.G., Kappas, A., Manstead, A.S. (2013) “Effects of dynamic aspects of facial expressions: a review,” *Emotion Review*, 5(1), 41–46.

Krumhuber, E.G., Likowski, K.U., Weyers, P. (2014) “Facial mimicry of spontaneous and deliberate Duchenne and non-Duchenne smiles,” *Journal of Nonverbal Behavior*, 38(1), 1–11.

Kuesten, C., Chopra, P., Bi, J., Meiselman, H.L. (2014) “A global study using PANAS (PA and NA) scales to measure consumer emotions associated with aromas of phytonutrient supplements,” *Food Quality and Preference*, 33, 86–97.

Kumar, S., Nilsen, W.J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., Riley, W.T., Shar, A., Spring, B., Spruijt-Metz, D., others (2013) “Mobile health technology evaluation: the mHealth evidence workshop,” *American journal of preventive medicine*, 45(2), 228–236.

Künecke, J., Hildebrandt, A., Recio, G., Sommer, W., Wilhelm, O. (2014) “Facial EMG responses to emotional expressions are related to emotion perception ability,” *PloS one*, 9(1), e84053.

Kung, S., Alarcon, R.D., Williams, M.D., Poppe, K.A., Moore, M.J., Frye, M.A. (2013) "Comparing the Beck Depression Inventory-II (BDI-II) and Patient Health Questionnaire (PHQ-9) depression measures in an integrated mood disorders practice," *Journal of affective disorders*, 145(3), 341–343.

LaFrance, M., Hecht, M.A. (1995) "Why smiles generate leniency," *Personality and Social Psychology Bulletin*, 21(3), 207–214.

Laird, J.D. (1974) "Self-attribution of emotion: the effects of expressive behavior on the quality of emotional experience.," *Journal of personality and social psychology*, 29(4), 475.

Lange, C. (1885) "The Mechanism of the Emotions ."

Larsen, J.T., Norris, C.J., Cacioppo, J.T. (2003) "Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii," *Psychophysiology*, 40(5), 776–785.

Lazebnik, Y. (2002) "Can a biologist fix a radio?—Or, what I learned while studying apoptosis," *Cancer cell*, 2(3), 179–182.

Lee, T.-W., Josephs, O., Dolan, R.J., Critchley, H.D. (2006) "Imitating expressions: emotion-specific neural substrates in facial mimicry," *Social cognitive and affective neuroscience*, 1(2), 122–135.

Lerner, D., Adler, D.A., Rogers, W.H., Lapitsky, L., McLaughlin, T., Reed, J., others (2010) "Work performance of employees with depression: the impact of work stressors," *American Journal of Health Promotion*, 24(3), 205–213.

Lewis, M.B. (2012) "Exploring the positive and negative implications of facial feedback.," *Emotion*, 12(4), 852.

LiKamWa, R., Liu, Y., Lane, N.D., Zhong, L. (2011) "Can your smartphone infer your mood," in *PhoneSense Workshop*, 1–5.

Lillie, E.O., Patay, B., Diamant, J., Issell, B., Topol, E.J., Schork, N.J. (2011) "The n-of-1 clinical trial: the ultimate strategy for individualizing medicine?," *Personalized medicine*, 8(2), 161–173.

Lindhe, J., Bressan, E., Cecchinato, D., Corrá, E., Toia, M., Liljenberg, B. (2013) "Bone tissue in different parts of the edentulous maxilla and mandible," *Clinical oral implants research*, 24(4), 372–377.

Linos, E., Reinhard, J., Ruda, S. (2015) Promoting Diversity in the Police [online], available: http://www.behaviouralinsights.co.uk/trial-results/behavioural-insights-and-home-affairs/ [accessed 3 2017].

Liu, F., Hamer, M.A., Deelen, J., Lall, J.S., Jacobs, L., van Heemst, D., Murray, P.G., Wollstein, A., de Craen, A.J., Uh, H.-W., others (2016) "The MC1R Gene and Youthful Looks," *Current Biology*, 26(9), 1213–1220.

Liu, J.Q., Zelko, I.N., Erbynn, E.M., Sham, J.S., Folz, R.J. (2006) "Hypoxic pulmonary hypertension: role of superoxide and NADPH oxidase (gp91phox)," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 290(1), L2–L10.

De Luca, C.J. (1997) "The use of surface electromyography in biomechanics," *Journal of applied biomechanics*, 13, 135–163.

Lundqvist, D., Flykt, A., Öhman, A. (1998) *The Karolinska Directed Emotional Faces-KDEF. CD-ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet, Stockholm, Sweden*.

Lundqvist, D.L.-O. (1995) “Facial EMG reactions to facial expressions: a case of facial emotional contagion?,” *Scandinavian journal of psychology*, 36(2), 130–141.

Ly, K.H., Trüschel, A., Jarl, L., Magnusson, S., Windahl, T., Johansson, R., Carlbring, P., Andersson, G. (2014) “Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial,” *BMJ open*, 4(1), e003440.

Lydon, D.M., Nixon, E. (2014) “Facial mimicry interference affects reaction time but not accuracy on an emotion recognition task in children, adolescents and adults,” *The Irish Journal of Psychology*, 35(2-3), 106–120.

Lyubomirsky, S., Dickerhoof, R., Boehm, J.K., Sheldon, K.M. (2011) “Becoming happier takes both a will and a proper way: an experimental longitudinal intervention to boost well-being.,” *Emotion*, 11(2), 391.

Lyubomirsky, S., Lepper, H.S. (1999) “A measure of subjective happiness: Preliminary reliability and construct validation,” *Social indicators research*, 46(2), 137–155.

Löwe, B., Unützer, J., Callahan, C.M., Perkins, A.J., Kroenke, K. (2004) “Monitoring depression treatment outcomes with the patient health questionnaire-9,” *Medical care*, 42(12), 1194–1201.

Madani, Y., Azad, S., Nachev, P., Collas, D. (2013) “Laughing at funerals,” *QJM*, 106(10), 951–952.

Mah, Y.-H., Husain, M., Rees, G., Nachev, P. (2014) “Human brain lesion-deficit inference remapped,” *Brain*, 137(9), 2522–2531.

Mancini, G., Ferrari, P.F., Palagi, E. (2013a) “In play we trust. Rapid facial mimicry predicts the duration of playful interactions in geladas,” *PloS one*, 8(6), e66481.

Mancini, G., Ferrari, P.F., Palagi, E. (2013b) “Rapid facial mimicry in geladas,” *Scientific reports*, 3, 1527.

Marder, E., Bucher, D. (2007) “Understanding circuit dynamics using the stomatogastric nervous system of lobsters and crabs,” *Annu. Rev. Physiol.*, 69, 291–316.

Marks, J. (1982) “A theory of emotion,” *Philosophical Studies*, 42(2), 227–242.

Marrakchi, S., Maibach, H.I. (2007) “Biophysical parameters of skin: map of human face, regional, and age-related differences,” *Contact dermatitis*, 57(1), 28–34.

Matsukubo, H., Matsuura, T., Endo, N., Asayama, J., Watanabe, T. (1977) “Echocardiographic measurement of right ventricular wall thickness. A new application of subxiphoid echocardiography.,” *Circulation*, 56(2), 278–284.

Matsumoto, D. (1987). The role of facial response in the experience of emotion: More methodological problems and a meta-analysis. *Journal of Personality and Social Psychology*, 52, 769-774.

Matzke, B., Herpertz, S.C., Berger, C., Fleischer, M., Domes, G. (2013) “Facial reactions during emotion recognition in borderline personality disorder: a facial electromyography study,” *Psychopathology*, 47(2), 101–110.

McFie, J., Piercy, Mf., Zangwill, O. (1950) “Visualspatial agnosia associated with lesions of the right cerebral hemisphere.,” *Brain: a journal of neurology*.

McIntosh, D.N. (1996) “Facial feedback hypotheses: Evidence, implications, and directions,” *Motivation and emotion*, 20(2), 121–147.

McKone, E., Crookes, K., Kanwisher, N., others (2009) “The cognitive and neural development of face recognition in humans,” *The cognitive neurosciences*, 4, 467–482.

Medicines, Healthcare products Regulatory Agency (2016) *Medical Device Stand-Alone Software Including Apps (including IVDMDs)*.

Mewett, D.T., Nazeran, H., Reynolds, K.J. (2001) “Removing power line noise from recorded EMG,” in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, 2190–2193.

Meyer D, Leisch F, Hornik K. (2003). "The support vector machine under test." *Neurocomputing*, *55*, 59–71.

Miller, R.G., Mitchell, J., Lyon, M., Moore, D.H. (2007) “Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND),” *The Cochrane Library*.

Mitchell, G.K., Hardy, J.R., Nikles, C.J., Carmont, S.-A.S., Senior, H.E., Schluter, P.J., Good, P., Currow, D.C. (2015) “The effect of methylphenidate on fatigue in advanced cancer: an aggregated n-of-1 trial,” *Journal of pain and symptom management*, 50(3), 289–296.

Mojtabai, R., Olfson, M., Sampson, N.A., Jin, R., Druss, B., Wang, P.S., Wells, K.B., Pincus, H.A., Kessler, R.C. (2011) “Barriers to mental health treatment: results from the National Comorbidity Survey Replication,” *Psychological medicine*, 41(08), 1751–1761.

Monti M.M. (2011). Statistical analysis of fMRI time-series: a critical review of the GLM approach. *Frontiers in Human Neuroscience*, 5 , 28.

Moody, E.J., McIntosh, D.N., Mann, L.J., Weisser, K.R. (2007) “More than mere mimicry? The influence of emotion on rapid facial reactions to faces.,” *Emotion*, 7(2), 447.

Morgan, J. (2016) “Gaming for dementia research: a quest to save the brain,” *The Lancet Neurology*, 15(13), 1313.

Morris, J.S., Öhman, A., Dolan, R.J. (1998) “Conscious and unconscious emotional learning in the human amygdala,” *Nature*, 393(6684), 467–470.

Moura, M.C., Novaes, M.R.C.G., Eduardo, E.J., Zago, Y.S., Freitas, R.D.N.B., Casulari, L.A. (2015) “Prognostic factors in amyotrophic lateral sclerosis: a population-based study,” *PloS one*, 10(10), e0141500.

Mukerji, C., Perszyk, D., McPartland, J.C. (2013) “Face Recognition,” in *Encyclopedia of Autism Spectrum Disorders*, Springer, 1222–1226.

Munsat, T., Andres, P., Skerry, L. (1990) “Amyotrophic Lateral Sclerosis, Progress in Clinical Neurologic Trials,” in F. Clifford Rose , ed., Demos Publications: New York, 65–76.

Murry, T., Cannito, M.P., Woodson, G.E. (1994) “Spasmodic dysphonia: emotional status and botulinum toxin treatment,” *Archives of Otolaryngology-Head \& Neck Surgery*, 120(3), 310–316.

Musiat, P., Goldstone, P., Tarrier, N. (2014) “Understanding the acceptability of e-mental health-attitudes and expectations towards computerised self-help treatments for mental health problems,” *BMC psychiatry*, 14(1), 1.

Nachev, P. (2015) “The Laws of the Brain.” *Unpublished Manuscript*.

Nawa, N.E., Ando, H. (2014) “Classification of self-driven mental tasks from whole-brain activity patterns,” *PloS one*, 9(5), e97296.

Neal, D.T., Chartrand, T.L. (2011) “Embodied emotion perception amplifying and dampening facial feedback modulates emotion perception accuracy,” *Social Psychological and Personality Science*, 2(6), 673–678.

Neighbors, H.W., Woodward, A.T., Bullard, K.M., Ford, B.C., Taylor, R.J., Jackson, J.S. (2008) “Mental health service use among older African Americans: the National Survey of American Life,” *The American Journal of Geriatric Psychiatry*, 16(12), 948–956.

Neufeld, L., Aaron, G., Garrett, G., Baker, S., Dary, O., Van, M. (2016) “Food fortification for impact: a data-driven approach,” *Bulletin of the World Health Organisation*, 94, 631–632.

Neumann, R., Schulz, S.M., Lozo, L., Alpers, G.W. (2014) “Automatic facial responses to near-threshold presented facial displays of emotion: Imitation or evaluation?,” *Biological psychology*, 96, 144–149.

Newsom-Davis, I., Abrahams, S., Goldstein, L., Leigh, P. (1999) “The emotional lability questionnaire: a new measure of emotional lability in amyotrophic lateral sclerosis,” *Journal of the neurological sciences*, 169(1), 22–25.

NHS England (2015) *Guidance to Support the Introduction of Access and Waiting Time Standards for Mental Health Services in 2015/16*.

Niermeyer, S., Yang, P., Zhuang, J., Moore, L.G. (1995) “Arterial oxygen saturation in Tibetan and Han infants born in Lhasa, Tibet,” *New England Journal of Medicine*, 333(19), 1248–1252.

Nietzsche, F. (1887) *On the Genealogy of Morality: A Polemic Tract*.

Nischal, A., Tripathi, A., Nischal, A., Trivedi, J., others (2012) “Suicide and antidepressants: what current evidence indicates,” *Mens sana monographs*, 10(1), 33.

Nitze, I., Schulthess, U. and Asche, H. (2012) “*Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification*,” presented at Proc. of the 4th GEOBIA, 7-9 May 2012, Rio de Janeiro, Brazil (2012).

Nomura, M., Ohira, H., Haneda, K., Iidaka, T., Sadato, N., Okada, T., Yonekura, Y. (2004) “Functional association of the amygdala and ventral prefrontal cortex during cognitive evaluation of facial expressions primed by masked angry faces: an event-related fMRI study,” *Neuroimage*, 21(1), 352–363.

Nulman, I., Koren, G., Rovet, J., Barrera, M., Pulver, A., Streiner, D., Feldman, B. (2012) “Neurodevelopment of children following prenatal exposure to venlafaxine, selective serotonin reuptake inhibitors, or untreated maternal depression,” *American Journal of Psychiatry*.

Nutting, P.A., Rost, K., Dickinson, M., Werner, J.J., Dickinson, P., Smith, J.L., Gallovic, B. (2002) “Barriers to initiating depression treatment in primary care practice,” *Journal of General Internal Medicine*, 17(2), 103–111.

O’Toole, A.J., Deffenbacher, K.A., Valentin, D., McKee, K., Huff, D., Abdi, H. (1998) “The perception of face gender: The role of stimulus structure in recognition and classification,” *Memory \& Cognition*, 26(1), 146–160.

Oakley, J. (1992) “Morality and the Emotions.”

Obafemi-Ajayi, T., Miles, J.H., Takahashi, T.N., Qi, W., Aldridge, K., Zhang, M., Xin, S.-Q., He, Y., Duan, Y. (2015) “Facial structure analysis separates autism spectrum disorders into meaningful clinical subgroups,” *Journal of autism and developmental disorders*, 45(5), 1302–

1317.

Oberman, L.M., Winkielman, P., Ramachandran, V.S. (2007) “Face to face: Blocking facial mimicry can selectively impair recognition of emotional expressions,” *Social neuroscience*, 2(3-4), 167–178.

Oberman, L.M., Winkielman, P., Ramachandran, V.S. (2009) “Slow echo: facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders,” *Developmental science*, 12(4), 510–520.

Ogawa S., & Sung Y. (2007). *Scholarpedia*, *2*(10), 3105.

Olivieri, N.F. (2003) “Patients’ health or company profits? The commercialisation of academic research,” *Science and Engineering Ethics*, 9(1), 29–41.

Orians, G.H., Heerwagen, J.H. (1992) “Evolved responses to landscapes.”

Ozaki, M., Kawashima, S., Yamashita, T., Ohashi, Y., Rikitake, Y., Inoue, N., Hirata, K.-I., Hayashi, Y., Itoh, H., Yokoyama, M. (2001) “Reduced hypoxic pulmonary vascular remodeling by nitric oxide from the endothelium,” *Hypertension*, 37(2), 322–327.

Palagi, E., Leone, A., Mancini, G., Ferrari, P. (2009) “Contagious yawning in gelada baboons as a possible expression of empathy,” *Proceedings of the National Academy of Sciences*, 106(46), 19262–19267.

Palmieri, A., Abrahams, S., Sorarù, G., Mattiuzzi, L., D’Ascenzo, C., Pegoraro, E., Angelini, C. (2009) “Emotional Lability in MND: relationship to cognition and psychopathology and impact on caregivers,” *Journal of the neurological sciences*, 278(1), 16–20.

Pampallona, S., Bollini, P., Tibaldi, G., Kupelnick, B., Munizza, C. (2004) “Combined pharmacotherapy and psychological treatment for depression: a systematic review,” *Archives of general psychiatry*, 61(7), 714–719.

De Paris, S.M. des H. (1862) *Bulletins et Mémoires de La Société Médicale Des Hôpitaux de Paris*, Masson.

Paukner, A., Anderson, J.R. (2006) “Video-induced yawning in stumptail macaques (Macaca arctoides),” *Biology Letters*, 2(1), 36–38.

Perry, C., Chhatralia, K., Damesick, D., Hobden, S., Volpe, L. (2015) “Behavioural insights in health care,” *London: The Health Foundation*, 18–29.

Phadke, S.R., Agarwal, S. (2003) “Phenotype score to grade the severity of thalassemia intermedia,” *Indian journal of pediatrics*, 70(6), 477–481.

Pichon, S., Miendlarzewska, E.A., Eryilmaz, H., Vuilleumier, P. (2014) “Cumulative activation during positive and negative events and state anxiety predicts subsequent inertia of amygdala reactivity,” *Social cognitive and affective neuroscience*, nsu044.

Plant, E.A., Hyde, J.S., Keltner, D., Devine, P.G. (2000) “The gender stereotyping of emotions,” *Psychology of Women Quarterly*, 24(1), 81–92.

Platek, S.M., Mohamed, F.B., Gallup, G.G. (2005) “Contagious yawning and the brain,” *Cognitive brain research*, 23(2), 448–452.

Pringle, N.P., Mudhar, H.S., Collarini, E.J., Richardson, W.D. (1992) “PDGF receptors in the rat CNS: during late neurogenesis, PDGF alpha-receptor expression appears to be restricted to glial cells of the oligodendrocyte lineage,” *Development*, 115(2), 535–551.

Prkachin, K.M., Berzins, S., Mercer, S.R. (1994) “Encoding and decoding of pain expressions: a judgement study,” *Pain*, 58(2), 253–259.

Punja, S., Schmid, C.H., Hartling, L., Urichuk, L., Nikles, C.J., Vohra, S. (2016) “To meta-analyze or not to meta-analyze? A combined meta-analysis of N-of-1 trial data with RCT data on amphetamines and methylphenidate for pediatric ADHD,” *Journal of clinical epidemiology*.

Quackenbush, J. (2007) “Extracting biology from high-dimensional biological data,” *Journal of Experimental Biology*, 210(9), 1507–1517.

Reaz, M., Hussain, M., Mohd-Yasin, F. (2006) “Techniques of EMG signal analysis: detection, processing, classification and applications,” *Biological procedures online*, 8(1), 11–35.

Rhodes, M.G. (2009) “Age estimation of faces: A review,” *Applied Cognitive Psychology*, 23(1), 1–12.

Rhodes, M.G., Anastasi, J.S. (2012) “The own-age bias in face recognition: a meta-analytic and theoretical review.,” *Psychological bulletin*, 138(1), 146.

Richards, D., Richardson, T. (2012) “Computer-based psychological treatments for depression: a systematic review and meta-analysis,” *Clinical psychology review*, 32(4), 329–342.

Rinn, W.E. (1984) “The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions.,” *Psychological bulletin*, 95(1), 52.

Rizzolatti, G., Craighero, L., Fadiga, L. (2002) “The mirror system in humans,” *Advances in Consciousness Research*, 42, 37–62.

Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L. (1996) “Premotor cortex and the recognition of motor actions,” *Cognitive brain research*, 3(2), 131–141.

Robinson, L.A., Berman, J.S., Neimeyer, R.A. (1990) “Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research.,” *Psychological bulletin*, 108(1), 30.

Robinson, R.G., Parikh, R.M., Lipsey, J.R., Starkstein, S.E., Price, T.R. (1993) “Pathological laughing and crying following stroke: validation of a measurement scale and a double-blind treatment study,” *American Journal of Psychiatry*, 150, 286–286.

Romero, R., Espinoza, J., Gotsch, F., Kusanovic, J., Friel, L., Erez, O., Mazaki-Tovi, S., Than, N., Hassan, S., Tromp, G. (2006) “The use of high-dimensional biology (genomics, transcriptomics, proteomics, and metabolomics) to understand the preterm parturition syndrome,” *BJOG: An International Journal of Obstetrics & Gynaecology*, 113(s3), 118–135.

Rosenfeld, J., Strong, M.J. (2015) “Challenges in the understanding and treatment of amyotrophic lateral sclerosis/motor neuron disease,” *Neurotherapeutics*, 12(2), 317–325.

Ross, M.D., Menzler, S., Zimmermann, E. (2008) “Rapid facial mimicry in orangutan play,” *Biology letters*, 4(1), 27–30.

Rost, K., Zhang, M., Fortney, J., Smith, J., Smith Jr, G.R. (1998) “Rural-urban differences in depression treatment and suicidality,” *Medical Care*, 36(7), 1098–1107.

Rowland, L.P. (2001) “How amyotrophic lateral sclerosis got its name: the clinical-pathologic genius of Jean-Martin Charcot,” *Archives of neurology*, 58(3), 512–515.

Rule, N.O., Ambady, N. (2008) “Brief exposures: Male sexual orientation is accurately perceived at 50ms,” *Journal of Experimental Social Psychology*, 44(4), 1100–1105.

Rund, D., Oron-Karni, V., Filon, D., Goldfarb, A., Rachmilewitz, E., Oppenheim, A. (1997) "Genetic Analysis of-Thalassemia Intermedia in Israel," *American journal of hematology*, 54, 16–22.

Rutledge, L. L., & Hupka, R. B. (1985). The facial feedback hypothesis: Methodological concerns and new supporting evidence. *Motivation and Emotion*, 9(3), 219-240.

Rymarczyk, K., Biele, C., Grabowska, A., Majczynski, H. (2011) "EMG activity in response to static and dynamic facial expressions," *International Journal of Psychophysiology*, 79(2), 330–333.

Sartori, C., Allemann, Y., Duplain, H., Lepori, M., Egli, M., Lipp, E., Hutter, D., Turini, P., Hugli, O., Cook, S., others (2002) "Salmeterol for the prevention of high-altitude pulmonary edema," *New England Journal of Medicine*, 346(21), 1631–1636.

Sato, W., Yoshikawa, S. (2007) "Spontaneous facial mimicry in response to dynamic facial expressions," *Cognition*, 104(1), 1–18.

Schachter, S., Singer, J. (1962) "Cognitive, social, and physiological determinants of emotional state.," *Psychological review*, 69(5), 379.

Scharlemann, J.P., Eckel, C.C., Kacelnik, A., Wilson, R.K. (2001) "The value of a smile: Game theory with a human face," *Journal of Economic Psychology*, 22(5), 617–640.

Scherer, K.R. (2005) "What are emotions? And how can they be measured?," *Social science information*, 44(4), 695–729.

Schhnall, S., Laird, J. (2003) "Keep Smiling: Enduring Effects of Facial Expressions and Postures on Emotional Experience and Memory," *Cognition and Emotion*, 17(5), 787–797.

Schmidhuber. J. (2015) "Deep learning in neural networks: An overview." *Neural Networks*, *61*, 85–117.

Schneider, K. G. (2008). *Expressive control and emotion perception: The impact of expressive suppression and mimicry on sensitivity to facial expressions of emotion* (Unpublished doctoral dissertation). Department of Psychology and Neuroscience, Duke University, Durham, NC.

Schork, N.J. (2015) "Personalized medicine: time for one-person trials," *Nature*, 520(7549), 609–11.

Schreier, D.A., Hacker, T.A., Hunter, K., Eickoff, J., Liu, A., Song, G., Chesler, N. (2014) "Impact of increased hematocrit on right ventricular afterload in response to chronic hypoxia," *Journal of applied physiology*, 117(8), 833–839.

Schwartz, G.E., Fair, P.L., Salt, P., Mandel, M.R., Klerman, G.L. (1976) "Facial muscle patterning to affective imagery in depressed and nondepressed subjects," *Science*, 192(4238), 489–491.

Scotton, W.J., Scott, K.M., Moore, D.H., Almedom, L., Wijesekera, L.C., Janssen, A., Nigro, C., Sakel, M., Leigh, P.N., Shaw, C., others (2012) "Prognostic categories for amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis*, 13(6), 502–508.

Scuffham, P.A., Nikles, J., Mitchell, G.K., Yelland, M.J., Vine, N., Poulos, C.J., Pillans, P.I., Bashford, G., Del Mar, C., Schluter, P.J., others (2010) "Using N-of-1 trials to improve patient management and save costs," *Journal of general internal medicine*, 25(9), 906–913.

Seder, J.P., Oishi, S. (2012) "Intensity of smiling in Facebook photos predicts future life satisfaction," *Social Psychological and Personality Science*, 3(4), 407–413.

Segerstråle, U.C.E., Molnár, P.E. (1997) *Nonverbal Communication: Where Nature Meets Culture.*, Lawrence Erlbaum Associates, Inc.

Seligman, M.E., Steen, T.A., Park, N., Peterson, C. (2005) "Positive psychology progress: empirical validation of interventions.," *American psychologist*, 60(5), 410.

Shaffer, F., McCraty, R., Zerr, C.L. (2014) "A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability," *Frontiers in psychology*, 5, 1040.

Shimai, S., Otake, K., Utsuki, N., Ikemi, A., Lyubomirsky, S. (2004) "[Development of a Japanese version of the Subjective Happiness Scale (SHS), and examination of its validity and reliability].," *[Nihon koshu eisei zasshi] Japanese journal of public health*, 51(10), 845–853.

Shoeb, A.H., Guttag, J.V. (2010) "Application of machine learning to epileptic seizure detection," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 975–982.

Simini, B. (1998) "Cherry-red discolouration in carbon monoxide poisoning," *The Lancet*, 352(9134), 1154.

Simon, G.E., Fleck, M., Lucas, R., Bushnell, D.M. (2004) "Prevalence and predictors of depression treatment in an international primary care study," *American Journal of Psychiatry*, 161(9), 1626–1634.

Simonyan, K., Zisserman, A. (2014) "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*.

Slayton, S.C., D'Archer, J., Kaplan, F. (2010) "Outcome studies on the efficacy of art therapy: A review of findings," *Art Therapy*, 27(3), 108–118.

Sloan, D.M., Bradley, M.M., Dimoulas, E., Lang, P.J. (2002) "Looking at facial expressions: Dysphoria and facial EMG," *Biological psychology*, 60(2), 79–90.

Smith, F.W., Schyns, P.G. (2009) "Smile through your fear and sadness transmitting and identifying facial expression signals over a range of viewing distances," *Psychological Science*, 20(10), 1202–1208.

Smith, R.A., Berg, J.E., Pope, L.E., Callahan, J.D., Wynn, D., Thisted, R.A. (2004) "Validation of the CNS emotional lability scale for pseudobulbar affect (pathological laughing and crying) in multiple sclerosis patients," *Multiple Sclerosis Journal*, 10(6), 679–685.

Sonnby--Borgström, M. (2002) "Automatic mimicry reactions as related to differences in emotional empathy," *Scandinavian journal of psychology*, 43(5), 433–443.

Sonnby-Borgström, M., Jönsson, P. (2003) "Models-of-self and models-of-others as related to facial muscle reactions at different levels of cognitive control," *Scandinavian journal of psychology*, 44(2), 141–151.

Sorlie, P.D., Garcia-Palmieri, M.R., Costas, R., & Havlik, R.J. (1981) "Hematocrit and risk of coronary heart disease: the Puerto Rico Heart Health Program," *American heart journal*, 101(4), 456–461.

Soussignan, R. (2002) "Duchenne smile, emotional experience, and autonomic reactivity: a test of the facial feedback hypothesis.," *Emotion*, 2(1), 52.

Spreng*, R.N., McKinnon*, M.C., Mar, R.A., Levine, B. (2009) "The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures," *Journal of personality assessment*, 91(1), 62–71.

Statnikov, A. Wang, L. & Aliferis, C. A. (2008). "A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification," *BMC Bioinformatics*, *9*(319).

Steffener, J.D., Tabert, M. Reuben, A., & Stern Y. (2010) "Investigating hemodynamic response variability at the group level using basis functions". *Neuroimage*, *49*, 2113-2122.

Steudel, W., Scherrer-Crosbie, M., Bloch, K.D., Weimann, J., Huang, P.L., Jones, R.C., Picard, M.H., Zapol, W.M. (1998) "Sustained pulmonary hypertension and right ventricular hypertrophy after chronic hypoxia in mice with congenital deficiency of nitric oxide synthase 3.," *Journal of Clinical Investigation*, 101(11), 2468.

Strack, F. (2016) "Reflection on the Smiling Registered Replication Report," *Perspectives on Psychological Science*, 1745691616674460.

Strack, F., Martin, L.L., Stepper, S. (1988) "Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis.," *Journal of personality and social psychology*, 54(5), 768.

Strange, K. (2005) "The end of 'naive reductionism': rise of systems biology or renaissance of physiology?," *American Journal of Physiology-Cell Physiology*, 288(5), C968–C974.

Van Straten, A., Hill, J., Richards, D., Cuijpers, P. (2015) "Stepped care treatment delivery for depression: a systematic review and meta-analysis," *Psychological medicine*, 45(02), 231–246.

Supuk, T.G., Skelin, A.K., Cic, M. (2014) "Design, development and testing of a low-cost sEMG system and its use in recording muscle activity in human gait," *Sensors*, 14(5), 8235–8258.

Surakka, V., Hietanen, J.K. (1998) "Facial and emotional reactions to Duchenne and non-Duchenne smiles," *International Journal of Psychophysiology*, 29(1), 23–33.

Swami, V., Stieger, S., Voracek, M., Dressler, S.G., Eisma, L., Furnham, A. (2009) "Psychometric evaluation of the Tagalog and German Subjective Happiness Scales and a cross-cultural comparison," *Social Indicators Research*, 93(2), 393–406.

Swan, M. (2009) "Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking," *International journal of environmental research and public health*, 6(2), 492–525.

Swan, M. (2013) "The quantified self: Fundamental disruption in big data science and biological discovery," *Big Data*, 1(2), 85–99.

Szendro, P., Vincze, G., Szasz, A. (2001) "Pink-noise behaviour of biosystems," *European Biophysics Journal*, 30(3), 227–231.

Taigman, Y., Yang, M., Ranzato, M., Wolf, L. (2014) "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1701–1708.

Takiwaki, H., Miyaoka, Y., Kohno, H., Arase, S. (2002) "Graphic analysis of the relationship between skin colour change and variations in the amounts of melanin and haemoglobin," *Skin research and technology*, 8(2), 78–83.

Tanaka, J.W., Kiefer, M., Bukach, C.M. (2004) "A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study," *Cognition*, 93(1), B1–B9.

Tateno, A., Jorge, R.E., Robinson, R.G. (2004) "Pathological laughing and crying following traumatic brain injury," *The Journal of neuropsychiatry and clinical neurosciences*, 16(4), 426–434.

Thaler, Sunstein (2008) "Nudge: Improving decisions about health, wealth and happiness."

Thomas, K.C., Ellis, A.R., Konrad, T.R., Holzer, C.E., Morrissey, J.P. (2009) “County-level estimates of mental health professional shortage in the United States,” *Psychiatric Services*.

Thompson, B. (2005) “Encyclopedia of statistics in behavioral science,” Wiley Online Library.

Thompson, E.R. (2007) “Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS),” *Journal of cross-cultural psychology*, 38(2), 227–242.

Tidd, K.L., Lockard, J.S. (1978) “Monetary significance of the affiliative smile: A case for reciprocal altruism,” *Bulletin of the Psychonomic Society*, 11(6), 344–346.

Tomkins, S.S. (1962) “Primary site of the affects: The face.”

Tomonaga, Y., Haettenschwiler, J., Hatzinger, M., Holsboer-Trachsler, E., Rufer, M., Hepp, U., Szucs, T.D. (2013) “The economic burden of depression in Switzerland,” *Pharmacoeconomics*, 31(3), 237–250.

Torous, J., Staples, P., Shanahan, M., Lin, C., Peck, P., Keshavan, M., Onnela, J.-P. (2015) “Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder,” *JMIR mental health*, 2(1).

Turner, M., Bakker, M., Sham, P., Shaw, C., Leigh, P., Al-Chalabi, A. (2002) “Prognostic modelling of therapeutic interventions in amyotrophic lateral sclerosis,” *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, 3(1), 15–21.

Vanina, Y., Podolskaya, A., Sedky, K., Shahab, H., Siddiqui, A., Munshi, F., Lippmann, S. (2002) “Body weight changes associated with psychopharmacology,” *Psychiatric services*, 53(7), 842–847.

Voelkel, N.F., Quaife, R.A., Leinwand, L.A., Barst, R.J., McGoon, M.D., Meldrum, D.R., Dupuis, J., Long, C.S., Rubin, L.J., Smart, F.W., others (2006) “Right ventricular function and failure,” *Circulation*, 114(17), 1883–1891.

Vrana, S.R. (1993) “The psychophysiology of disgust: Differentiating negative emotional contexts with facial EMG,” *Psychophysiology*, 30(3), 279–286.

Vrana, S.R., Rollock, D. (1998) “Physiological response to a minimal social encounter: Effects of gender, ethnicity, and social context,” *Psychophysiology*, 35(04), 462–469.

Vuilleumier, P., Armony, J.L., Driver, J., Dolan, R.J. (2001) “Effects of attention and emotion on face processing in the human brain: an event-related fMRI study,” *Neuron*, 30(3), 829–841.

Wagenmakers, E.-J., Beek, T., Dijkhoff, L., Gronau, Q.F., Acosta, A., Adams, R., Albohn, D., Allard, E., Benning, S., Blouin-Hudon, E.-M., others (2016) “Registered Replication Report ,” *Perspectives on Psychological Science*, 1745691616674458.

Wahl, O.F. (1999) “Mental health consumers’ experience of stigma.,” *Schizophrenia bulletin*, 25(3), 467.

Wanstall, J.C., Gambino, A., Jeffery, T.K., Cahill, M.M., Bellomo, D., Hayward, N.K., Kay, G.F. (2002) “Vascular endothelial growth factor-B-deficient mice show impaired development of hypoxic pulmonary hypertension,” *Cardiovascular research*, 55(2), 361–368.

Ward, L., Greenwood, P. (2007) “1/f noise,” *Scholarpedia*, 2, 1537.

Waters, K. (1987). "A muscle model for animating three-dimensional facial expressions", *Computer Graphics*, 21(4), 17-24.

Watts, S., Mackenzie, A., Thomas, C., Griskaitis, A., Mewton, L., Williams, A., Andrews, G. (2013) "CBT for depression: a pilot RCT comparing mobile phone vs. computer," *BMC psychiatry*, 13(1), 1.

Weber, A., Heger, S., Sinkgraven, R., Heckmann, M., Elsner, P., Rzany, B. (2005) "Psychosocial aspects of patients with focal hyperhidrosis. Marked reduction of social phobia, anxiety and depression and increased quality of life after treatment with botulinum toxin A," *British Journal of Dermatology*, 152(2), 342–345.

Weyers, P., Mühlberger, A., Hefele, C., Pauli, P. (2006) "Electromyographic responses to static and dynamic avatar emotional facial expressions," *Psychophysiology*, 43(5), 450–453.

Weyers, P., Mühlberger, A., Kund, A., Hess, U., Pauli, P. (2009) "Modulation of facial reactions to avatar emotional faces by nonconscious competition priming," *Psychophysiology*, 46(2), 328–335.

WHO (2012) *Global Burden of Mental Disorders and the Need for*.

Wild, B., Erb, M., Bartels, M. (2001) "Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences," *Psychiatry research*, 102(2), 109–124.

Williamson, J.R., Quatieri, T.F., Helfer, B.S., Ciccarelli, G., Mehta, D.D. (2014) "Vocal and facial biomarkers of depression based on motor incoordination and timing," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, 65–72.

Wilson, S.A.K. (1924) "Some problems in neurology," *Journal of Neurological Psychopathology*, 4, 299–333.

Wittgenstein, L. (1958) "The blue and brown books."

Wittgenstein, L., Anscombe, G.E.M., Wright, G.H. (1967) *Zettel*, Univ of California Press.

Wolf, K., Mass, R., Ingenbleek, T., Kiefer, F., Naber, D., Wiedemann, K. (2005) "The facial pattern of disgust, appetence, excited joy and relaxed joy: An improved facial EMG study," *Scandinavian Journal of Psychology*, 46(5), 403–409.

Woo, J.H., Wang, S., Melhem, E.R., Gee, J.C., Cucchiara, A., McCluskey, L., Elman, L. (2014) "Linear associations between clinically assessed upper motor neuron disease and diffusion tensor imaging metrics in amyotrophic lateral sclerosis," *PloS one*, 9(8), e105753.

Woodworth, R.J., O'Brien-Malone, A., Diamond, M.R., Schüz, B. (2016) "Happy Days: Positive Psychology interventions effects on affect in an N-of-1 trial," *International Journal of Clinical and Health Psychology*, 16(1), 21–29.

World Health Organisation (2016) WHO Depression Factsheet [online], available: http://www.who.int/mediacentre/factsheets/fs369/en/ [accessed 7–12 2016].

World Health Organization (2000) *Obesity: Preventing and Managing the Global Epidemic*, World Health Organization.

Wortzel, H.S., Oster, T.J., Anderson, C.A., Arciniegas, D.B. (2008) "Pathological laughing and crying," *CNS drugs*, 22(7), 531–545.

Wu, Y., Thalmann, N.M., Thalmann, D. (1995) "A dynamic wrinkle model in facial animation and skin ageing," *The journal of visualization and computer animation*, 6(4), 195–205.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., others (2010) "Sequencing of 50 human exomes reveals adaptation to

high altitude," *Science*, 329(5987), 75–78.

Yu, W., Liu, T., Valdez, R., Gwinn, M., Khoury, M.J. (2010) "Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes," *BMC Medical Informatics and Decision Making*, 10(1), 1.

Zerhouni, E. (2003) "The NIH roadmap," *Science*, 302(5642), 63–72.

Zerhouni, E.A. (2005) "Translational and clinical science—time for a new vision," *New England Journal of Medicine*, 353(15), 1621–1623.

Zhuang, J., Droma, T., Sutton, J.R., Groves, B.M., McCullough, R.E., McCullough, R.G., Sun, S., Moore, L.G. (1996) "Smaller alveolar-arterial O 2 gradients in Tibetan than Han residents of Lhasa (3658 m)," *Respiration physiology*, 103(1), 75–82.

Ziemann, U., Eisen, A. (2004) "TMS for ALS: why and why not," *Clinical Neurophysiology*, 115(6), 1237–1238.

Appendices

Appendix A - Supplementary information for Chapter 3

A.1 Patient Information Sheets (PIS)

UCL University College London Hospitals NHS
NHS Foundation Trust

Participant Information Sheet: Patients with suspected motor neuron disease without upper motor neuron involvement

Study title: A novel measure of upper motor neuron involvement in motor neuron disease
Protocol reference number: version number and date
Invitation to participate
We would like to invite you to take part in our research study. Before you decide we would like you to understand why the research is being done and what it would involve for you. **One of our team will go through the information sheet with you and answer any questions you have.** This should take about 10 minutes. Talk to others about the study if you wish. Part 1 tells you the purpose of this study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study. Ask us if there is anything that is not clear.

Part 1

1.1 **What is the purpose of the study?**
We wish to develop a simple, non-invasive test for patients with suspected motor neuron disease (MND). We hope this test will allow us to distinguish early on in the course of any symptoms between patients with MND who have upper motor neuron involvement and those who do not. Upper motor neuron involvement refers to the presence of changes within the surface of the brain, rather than only within the deeper parts of the brain and spinal cord. The distinction is important because the presence or absence of upper motor neuron involvement may be helpful in predicting the patient's future needs, and having this information early may help patients and their doctors make choices about the management of the illness early on. In this study we are seeking to characterise the performance of the test in patients with established MND, with and without upper motor neuron involvement, and to compare the data with a set derived from normal participants.

1.2 **Why have I been invited?**
As someone who has no upper motor neuron involvement your participation will help identify a pattern of physiological response on which a future diagnostic test for upper motor involvement in MND may be based.

1.3 **Do I have to take part?**
It is up to you to decide to join the study. We will describe the study and go through this information sheet. If you agree to take part, we will then ask you to sign a consent form. You are free to withdraw at any time, without giving a reason. This will not affect the standard of care you receive, now or in the future.

1.4 What will happen to me if I take part?

If you choose to take part, you will be asked to perform a simple computer-based task involving responding with a button press to images of faces projected on a screen in front of you. During the task, an eye tracker and recording electrodes placed on your face and arm will measure your body's reaction to the images. The eye-tracker is a small, light camera that rests on your head and monitors the movement of the eyes remotely without taking any video pictures. The recording electrodes are placed in contact with the skin in a few places over the face and forearm and held lightly with skin-sensitive stickers. At the end, you will be asked to complete a short questionnaire. In total, the experiment should take no longer than an hour and a half.

1.5 What are the potential risks and disadvantages in taking part?

All of the procedures involved are non-invasive and pose no risk to health. The only potential problem is minor skin allergy to the stickers used in positioning the electrodes: these are chosen to be hypoallergenic, so the risk of this is exceedingly low.

1.6 What are the possible benefits of taking part?

There is no benefit for you individually, but if the study is successful, it may form the basis of a future test that may be of benefit to others.

1.7 What happens when the research study stops?

We will follow ethical and legal practice and all information about you will be handled in confidence. Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The details are included in Part 2.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

Part 2

2.0 What will happen if I don't want to carry on with the study?

You are free to withdraw at any time, without giving a reason. This will not affect the standard of care you receive, now or in the future.

2.1 What if there is a problem?

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions [Dr Parashkev Nachev on mobile 07825093778]. If you remain unhappy and wish to complain formally, you

can do this by contacting the NHS Complaints department. Details can be obtained from NHS direct on 0845 4647].

2.2 **Will my taking part in this study be kept confidential?**
Yes. All information which is collected about you during the course of the research will be kept strictly confidential. All data will be collected and stored in accordance with the Data Protection Act 1998. All personal details and gathered data will be stored securely and accessed only by members of the research team. It will not be possible to identify your responses from any reports or publications.

2.3 **Involvement of the General Practitioner.**
Since participation in this study will not affect your health in any way, it will not be necessary to contact your General Practitioner.

2.4 **What will happen to the results of the research study?**
We intend to use the data collected to develop a test for patients with suspected motor neuron disease. We also intend to publish the results so as to assist other researchers in their work. No patient identifiable information will be published. Should you wish to receive an electronic copy of any publications arising from the study please notify a member of the research team who will email you a copy if and when available.

2.5 **Who is organising and funding the research?**
This research is funded by the Engineering and Physical Sciences Research Council and the University College London Institute of Biomedical Engineering.

2.6 **Who has reviewed the study?**
All research in the NHS is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by the NHS Research Ethics Committee under reference 13/LO/1693.

Contact for further information

Should you need any further information, please ask a member of the research team or contact Dr. Parashkev Nachev on 07825093778 or p.nachev@ucl.ac.uk

Thank you for taking the time to read this information sheet.

We appreciate your acceptance to participate and for taking the time to read the information sheet. You will be given a copy of an information sheet and a signed consent form for your personal records.

Date .....

University College London Hospitals NHS
NHS Foundation Trust

Participant Information Sheet: Patients with motor neuron disease with upper motor neuron involvement

Study title: A novel measure of upper motor neuron involvement in motor neuron disease
Protocol reference number: version number and date
Invitation to participate
We would like to invite you to take part in our research study. Before you decide we would like you to understand why the research is being done and what it would involve for you. **One of our team will go through the information sheet with you and answer any questions you have.** This should take about 10 minutes. Talk to others about the study if you wish. Part 1 tells you the purpose of this study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study. Ask us if there is anything that is not clear.

Part 1

1.1 **What is the purpose of the study?**
We wish to develop a simple, non-invasive test for patients with suspected motor neuron disease (MND). We hope this test will allow us to distinguish early on in the course of any symptoms between patients with MND who have upper motor neuron involvement and those who do not. Upper motor neuron involvement refers to the presence of changes within the surface of the brain, rather than only within the deeper parts of the brain and spinal cord. The distinction is important because the presence or absence of upper motor neuron involvement may be helpful in predicting the patient's future needs, and having this information early may help patients and their doctors make choices about the management of the illness early on. In this study we are seeking to characterise the performance of the test in patients with established MND, with and without upper motor neuron involvement, and to compare the data with a set derived from normal participants.

1.2 **Why have I been invited?**
As someone who has upper motor neuron involvement in the context of MND your participation will help identify a pattern of physiological response on which a future diagnostic test for upper involvement in MND may be based.

1.3 **Do I have to take part?**
It is up to you to decide to join the study. We will describe the study and go through this information sheet. If you agree to take part, we will then ask you to sign a consent form. You are free to withdraw at any time, without giving a reason. This will not affect the standard of care you receive, now or in the future.

1.4 What will happen to me if I take part?

If you choose to take part, you will be asked to perform a simple computer-based task involving responding with a button press to images of faces projected on a screen in front of you. During the task, an eye tracker and recording electrodes placed on your face and arm will measure your body's reaction to the images. The eye-tracker is a small, light camera that rests on your head and monitors the movement of the eyes remotely without taking any video pictures. The recording electrodes are placed in contact with the skin in a few places over the face and forearm and held lightly with skin-sensitive stickers. At the end, you will be asked to complete a short questionnaire. In total, the experiment should take no longer than an hour and a half.

1.5 What are the potential risks and disadvantages in taking part?

All of the procedures involved are non-invasive and pose no risk to health. The only potential problem is minor skin allergy to the stickers used in positioning the electrodes: these are chosen to be hypoallergenic, so the risk of this is exceedingly low.

1.6 What are the possible benefits of taking part?

There is no benefit for you individually, but if the study is successful, it may form the basis of a future test that may be of benefit to others.

1.7 What happens when the research study stops?

We will follow ethical and legal practice and all information about you will be handled in confidence. Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The details are included in Part 2.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

Part 2

2.0 What will happen if I don't want to carry on with the study?

You are free to withdraw at any time, without giving a reason. This will not affect the standard of care you receive, now or in the future.

2.1 What if there is a problem?

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions [Dr Parashkev Nachev 07825093778]. If you remain unhappy and wish to complain formally, you can do this by contacting the NHS Complaints department. Details can be obtained from [NHS direct: 0845 4647].

2.2 Will my taking part in this study be kept confidential?

Yes. All information which is collected about you during the course of the research will be kept strictly confidential. All data will be collected and stored in accordance with the Data Protection Act 1998. All personal details and gathered data will be stored securely and accessed only by members of the research team. It will not be possible to identify your responses from any reports or publications.

2.3 Involvement of the General Practitioner.

Since participation in this study will not affect your health in any way, it will not be necessary to contact your General Practitioner.

2.4 What will happen to the results of the research study?

We intend to use the data collected to develop a test for patients with suspected motor neuron disease. We also intend to publish the results so as to assist other researchers in their work. No patient identifiable information will be published. Should you wish to receive an electronic copy of any publications arising from the study please notify a member of the research team who will email you a copy if and when available.

2.5 Who is organising and funding the research?

This research is funded by the Engineering and Physical Sciences Research Council and the University College London Institute of Biomedical Engineering.

2.6 Who has reviewed the study?

All research in the NHS is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by the NHS Research Ethics Committee under reference 13/LO/1693.

Contact for further information

Shall you need any further information, please ask a member of the research team or contact Dr Parashkev Nachev on 07825093778 or p.nachev@ucl.ac.uk

Thank you

We appreciate your acceptance to participate and for taking the time to read the information sheet. You will be given a copy of an information sheet and a signed consent form for your personal records.

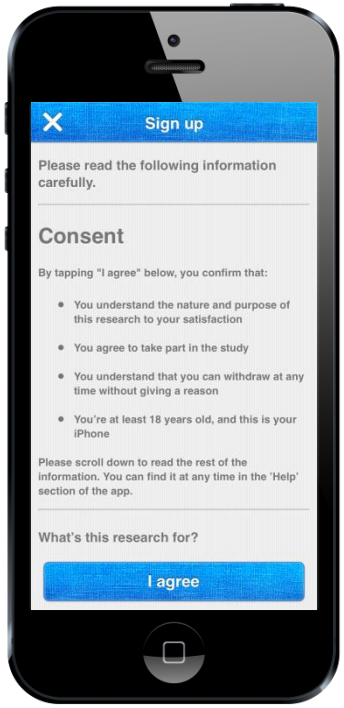
Date .....

Appendix B - Supplementary information for Chapter 4

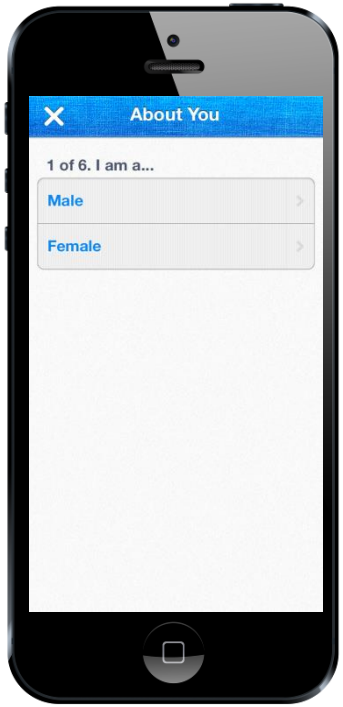
B.1 User interface screenshots

Loading

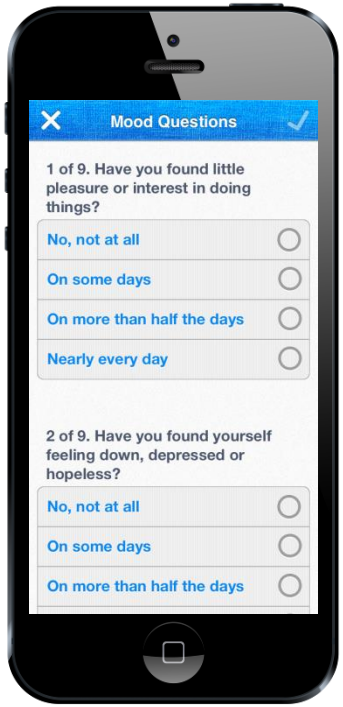
Welcome



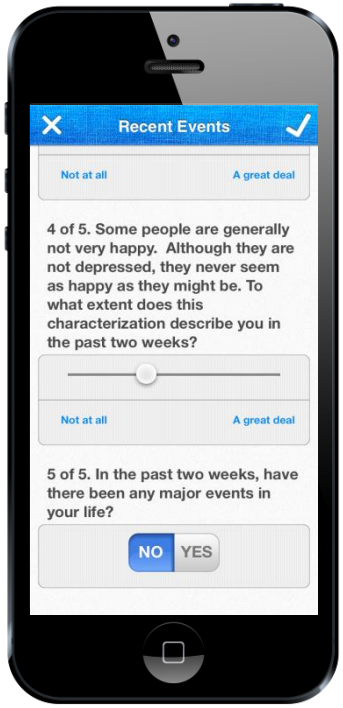
Consent (B.2)



Demographics



Depression survey (B.4)



Happiness survey (B.5)

B.2 Consent form in app

The text below is presented after participants selected the *sign up* welcome page item. It is also available at all times from the *FAQ* section within the app as a participant information sheet.

Please read the following information carefully.

Consent

By tapping "I agree" below, you confirm that:

- **You understand the nature and purpose of this research to your satisfaction**
- **You agree to take part in the study**
- **You understand that you can withdraw at any time without giving a reason**
- **You're at least 18 years old, and this is your iPhone**

Please scroll down to read the rest of the information. You can find it at any time in the 'Help' section of the app.

7.1.1.1 What's this research for?

Just like seeing people yawn makes you feel like yawning, so seeing people smile make you feel like smiling. This is a pilot study in which we seek to determine if people's mood can be improved by seeing faces of smiling people. You will be helping to design and test an app that may help improve people's mood without the use of drugs or other interventions associated with the risks of side effects.

7.1.1.2 What will I do?

After providing some basic demographic and health-related information you will be asked to complete two short mood questionnaires. You will then start receiving notifications on your phone showing smiling faces. This will happen by default 10 times a day. You can change the number of notifications and the times of the day you want to receive them in the 'Settings' section of the app.

Each notification will prompt you to open Pocket Smile and look at a smiling face. If you don't want to look at the face when prompted, that is fine, you will be prompted again at the next notification when a different face will appear. For the purposes of the experiment, it would be best if you leave the default settings so that you can see at least 10 faces a day, but feel free to adjust to whatever frequency suits you best.

After 10 days, you will be asked to complete the short mood questionnaire again. Once completed, Pocket Smile will present the questionnaire once a month while continuing to notify you to look at the smiling faces. You can keep taking part as long as you want.

It is important that you try to always keep the app minimized when not using it. The app is not power hungry at all, so it will not drain your battery.

7.1.1.3 How long will it take?

The sign-up process should not take more than 10 minutes. The daily viewing of smiling faces is like reading a notification; it may last from a couple of seconds to as long as you want. The monthly mood questionnaire will take about two minutes at most.

For your participation to be most useful for the study, it would be good if you could participate at least 3 months. However, you can keep taking part in the study for as long (or short) a period as you want and can leave the study without giving a reason.

7.1.1.4 What data will I be sharing?

None of the data collected will be personally identifiable. Pocket Smile will keep track of usage statistics, including how many images you see, the time spent looking at each image, and responses to the monthly mood questionnaires. All data will be safely and anonymously transferred to our secure data store.

7.1.1.5 What will you do with this data?

All gathered data will be stored securely and accessed only by members of the research team. We will use statistical techniques for our academic research. We will be looking at the effect of app usage on people's mood, taking into account other potential influences.

We'll be posting results and updates on www.pocketsmile.icn.ucl.ac.uk, so please visit if you're interested in seeing what we find. We also hope to present our findings in academic journals and at conferences.

7.1.1.6 Is it confidential?

Yes. We will keep all information of individuals strictly confidential and never report any responses of individual people. Information will only be used and reported at the group level. It will not be possible to identify your responses from any reports or publications. We do not collect any personally identifiable information such as your name or contact details.

Your data will not be disclosed to third parties unless (1) we're required by law to do so, or (2) exclusively for the purposes of academic research at a recognised institution, under a strict contractual agreement with other academic researchers who also agree on the confidentiality principles outlined here.

7.1.1.7 Is it secure?

Yes. All communication between Pocket Smile and our data store is over an SSL encrypted connection. This is the same kind that is used for online banking and secure shopping transactions. The data is stored in a firewalled and fully updated Linux server which can only be accessed over a secure connection.

7.1.1.8 Is it easy to quit?

Yes. Participation is completely voluntary. You are free to withdraw at any time, without giving any reason. Simply delete this app from your iPhone or take a break from the study by changing notifications per day to zero on the 'Settings' screen within the app.

We can also delete all your data from our data store if you ask us. We won't be able to remove your data from research that we've already published, however.

7.1.1.9 How much data does it use?

Sending your questionnaire responses uses as much data as sending a brief email (around 50KB).

An unlimited data plan is not necessary to take part. The app itself is approximately 17MB to download.

7.1.1.10 Can I take part if I'm not in the UK?

Yes. The app is available to download all over the world in English!

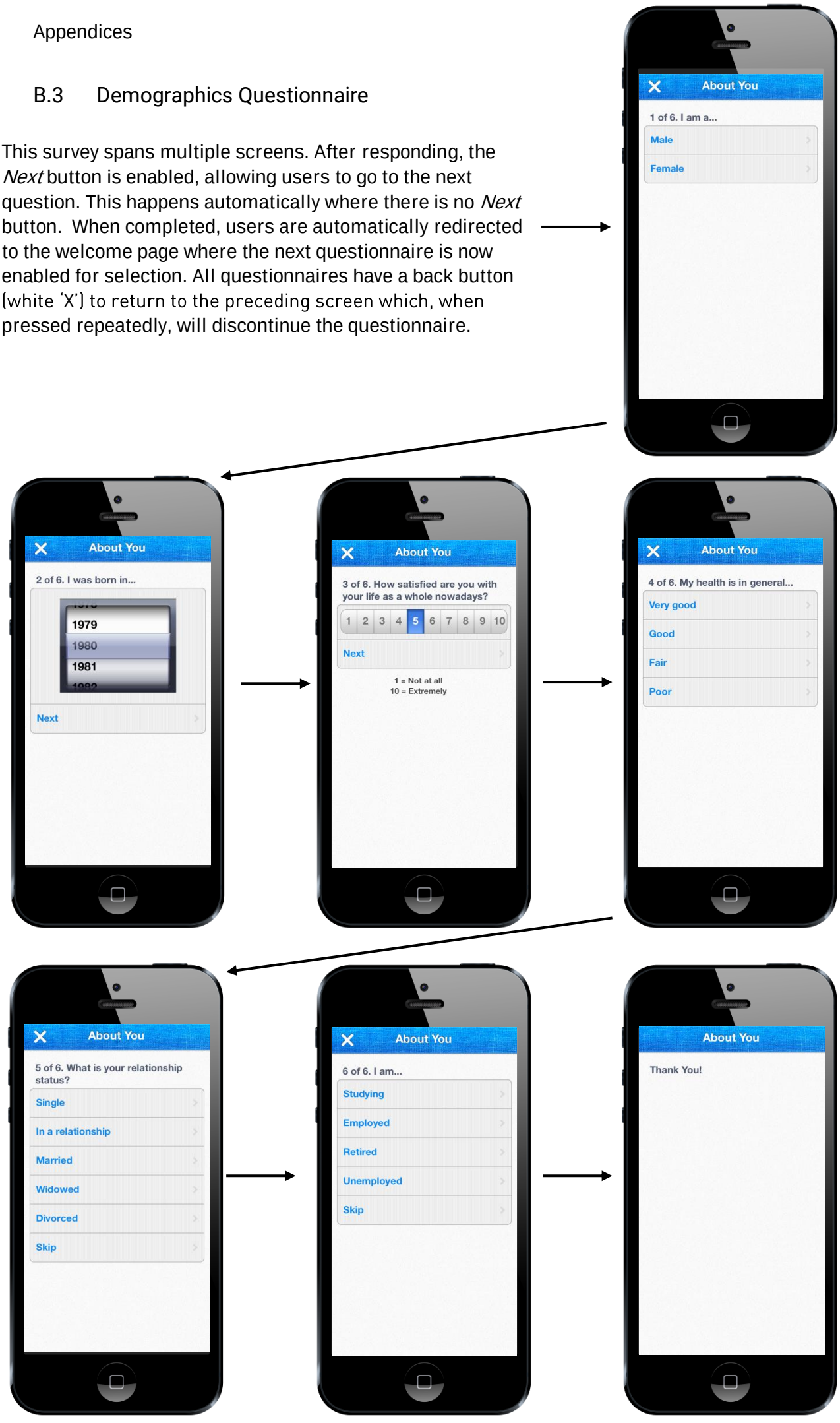
7.1.1.11 I have another question…

If there's anything else you'd like to know, please visit www.pocketsmile.icn.ucl.ac.uk or contact Javier Elkin or Dr Parashkev Nachev:
To do this right now, tap and hold this address: contact-pocketsmile@ucl.ac.uk
Or write to us at the Institute of Cognitive Neuroscience, Alexandra House, 17 Queen Square, London WC1N 3AR

Thank you!

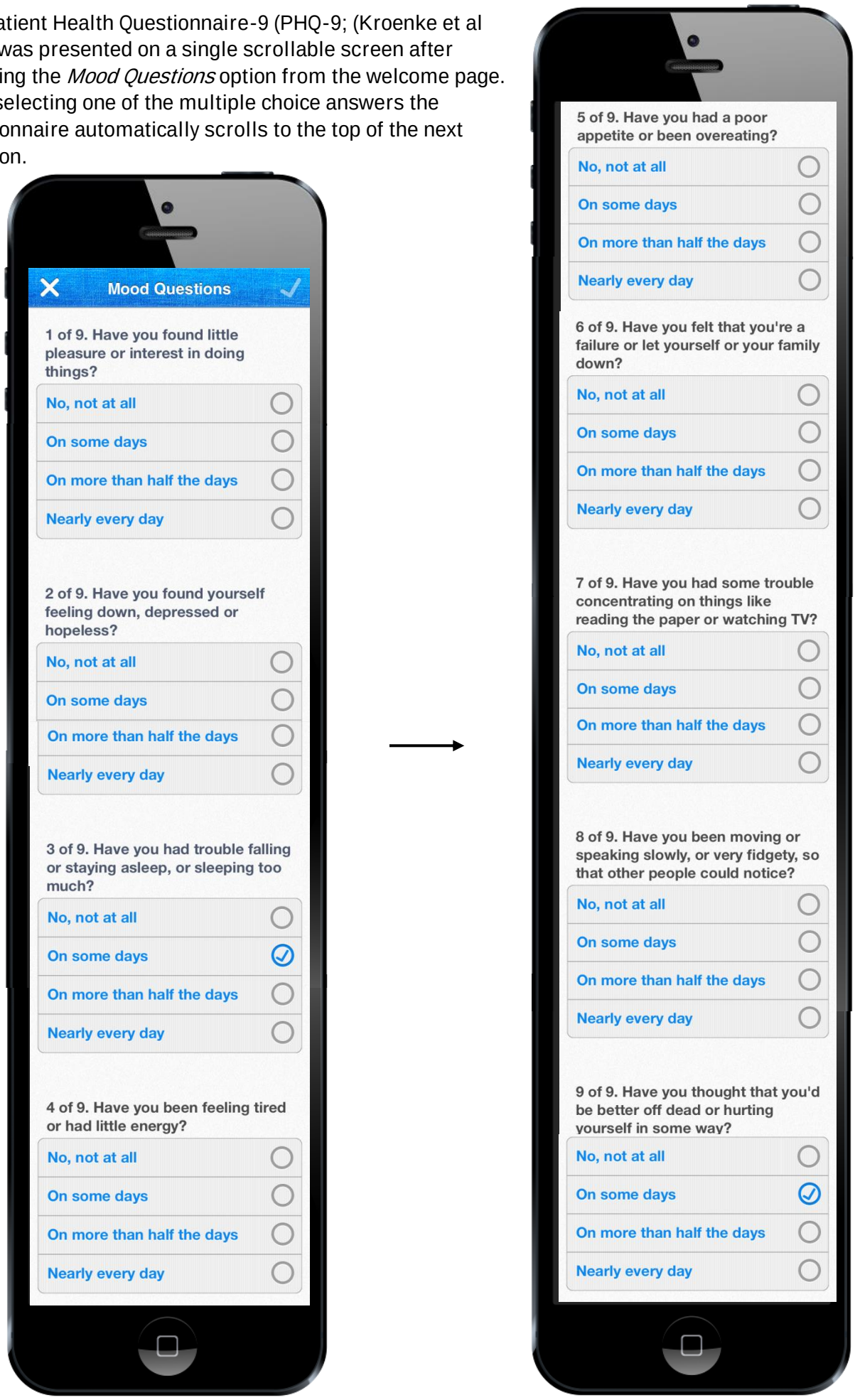
B.3 Demographics Questionnaire

This survey spans multiple screens. After responding, the *Next* button is enabled, allowing users to go to the next question. This happens automatically where there is no *Next* button. When completed, users are automatically redirected to the welcome page where the next questionnaire is now enabled for selection. All questionnaires have a back button (white 'X') to return to the preceding screen which, when pressed repeatedly, will discontinue the questionnaire.



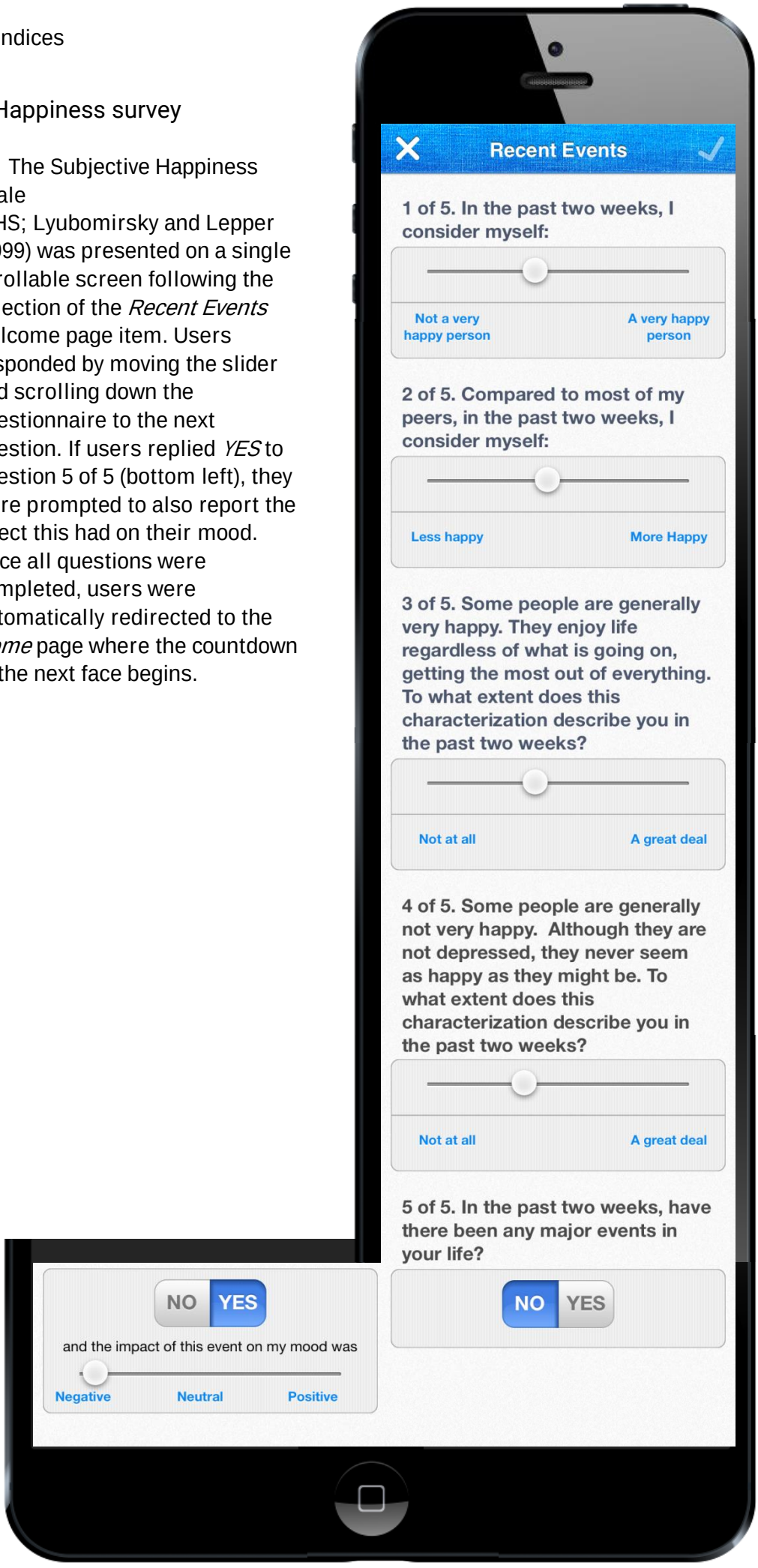
B.4 Depression survey

The Patient Health Questionnaire-9 (PHQ-9; (Kroenke et al 2001) was presented on a single scrollable screen after selecting the *Mood Questions* option from the welcome page. After selecting one of the multiple choice answers the questionnaire automatically scrolls to the top of the next question.

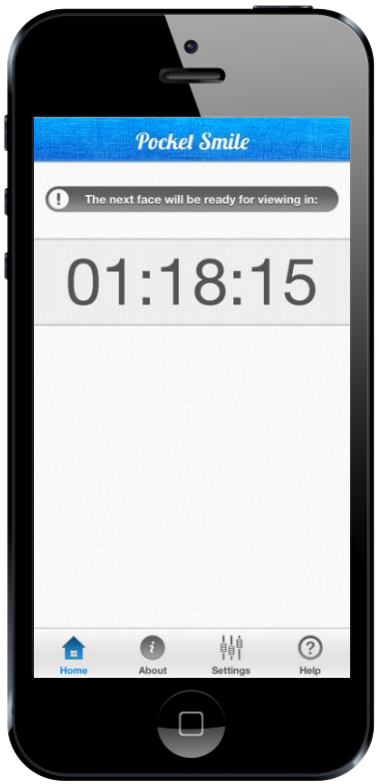


B.5 Happiness survey

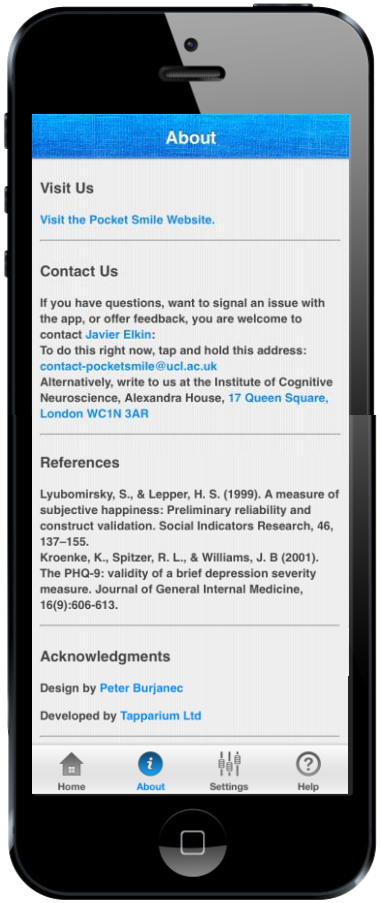
The Subjective Happiness Scale (SHS; Lyubomirsky and Lepper (1999) was presented on a single scrollable screen following the selection of the *Recent Events* welcome page item. Users responded by moving the slider and scrolling down the questionnaire to the next question. If users replied *YES* to question 5 of 5 (bottom left), they were prompted to also report the effect this had on their mood. Once all questions were completed, users were automatically redirected to the *Home* page where the countdown to the next face begins.



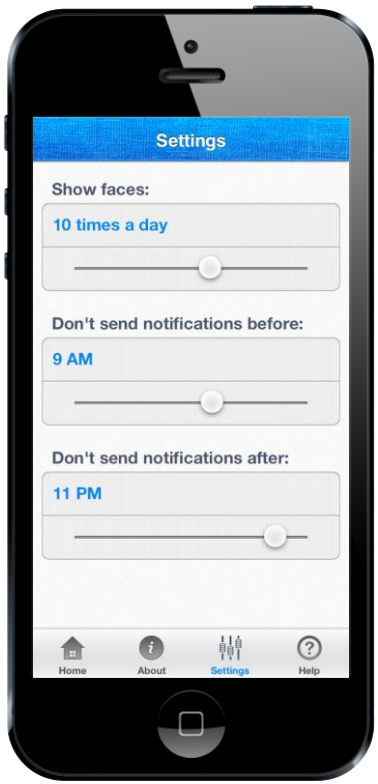
B.6 In-App Menus



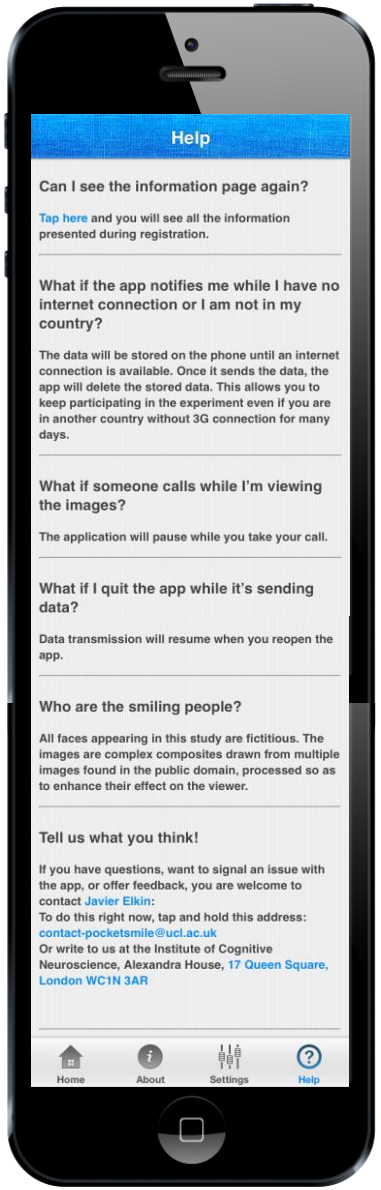
Home



About

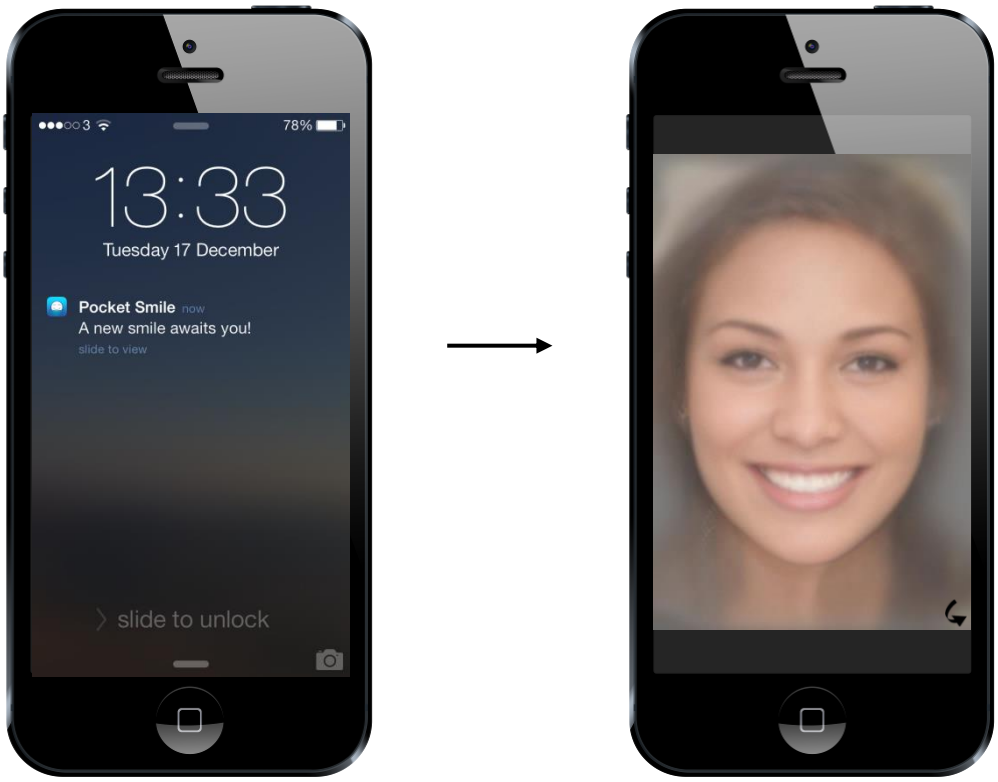


Settings



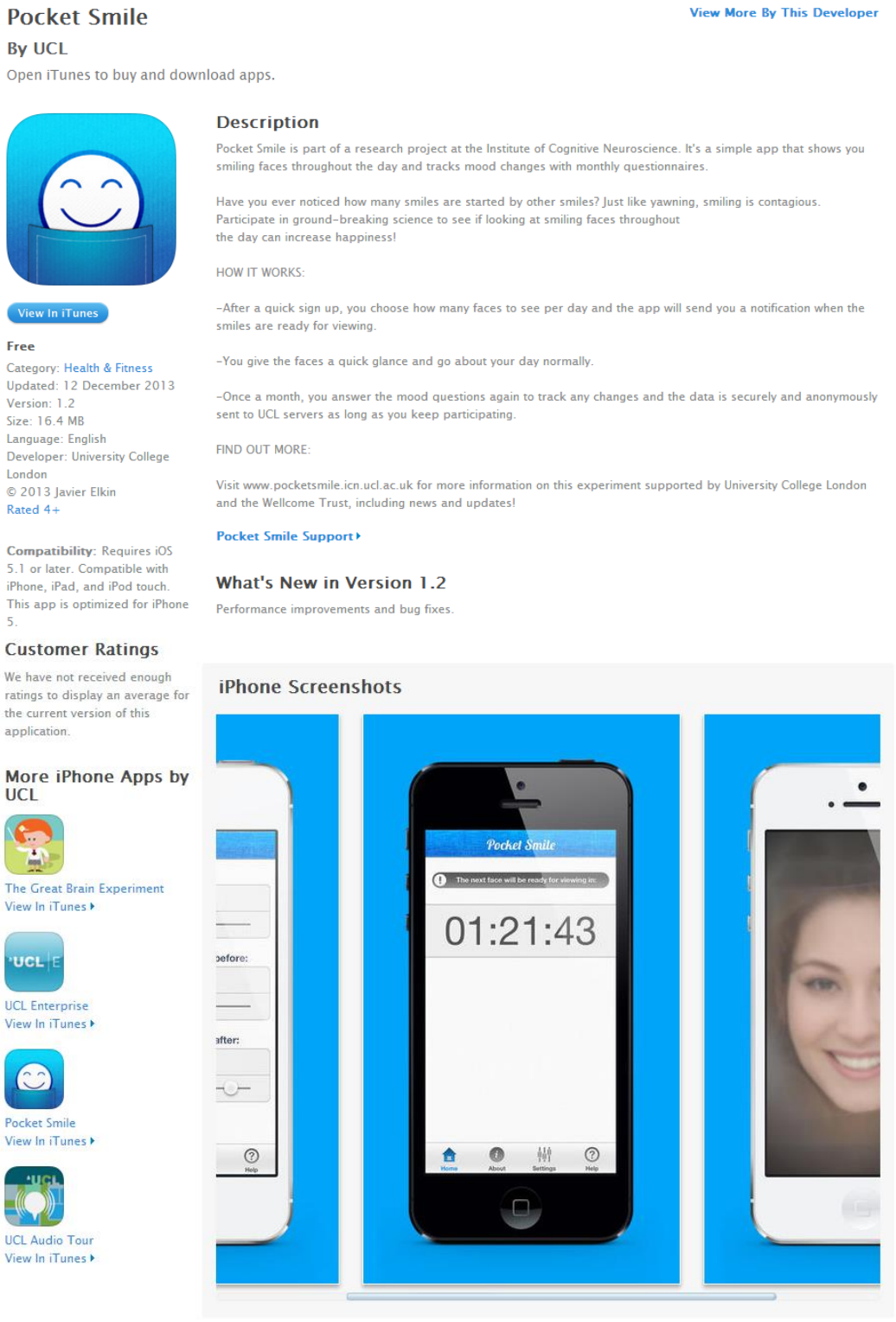
Help

B.7 Notification and Stimuli



B.8 App Store description

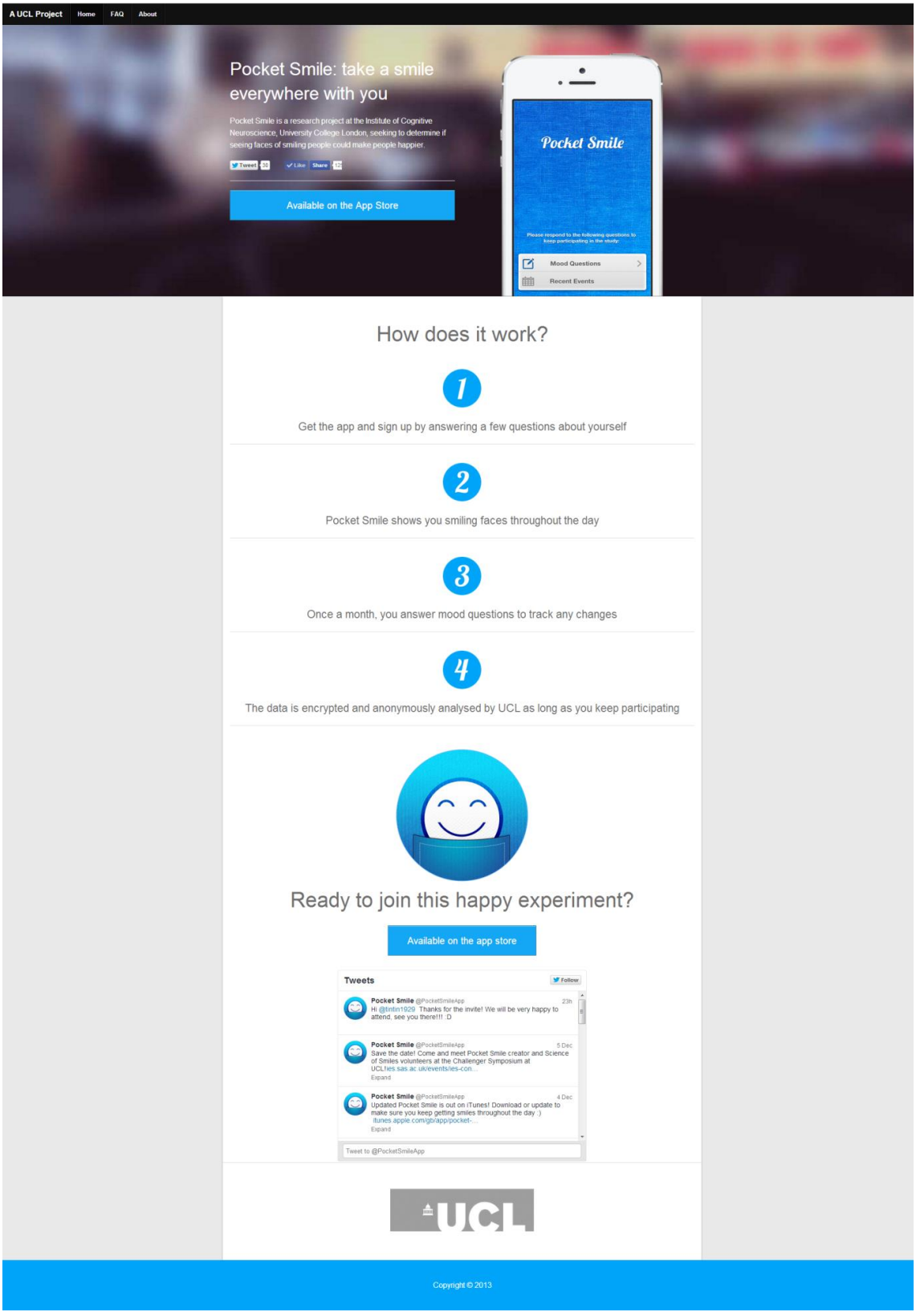
This is a screenshot of the iTunes page where users can download Pocket Smile. Retrieved 16.12.2013 link: https://itunes.apple.com/gb/app/pocket-smile/id709226955?mt=8



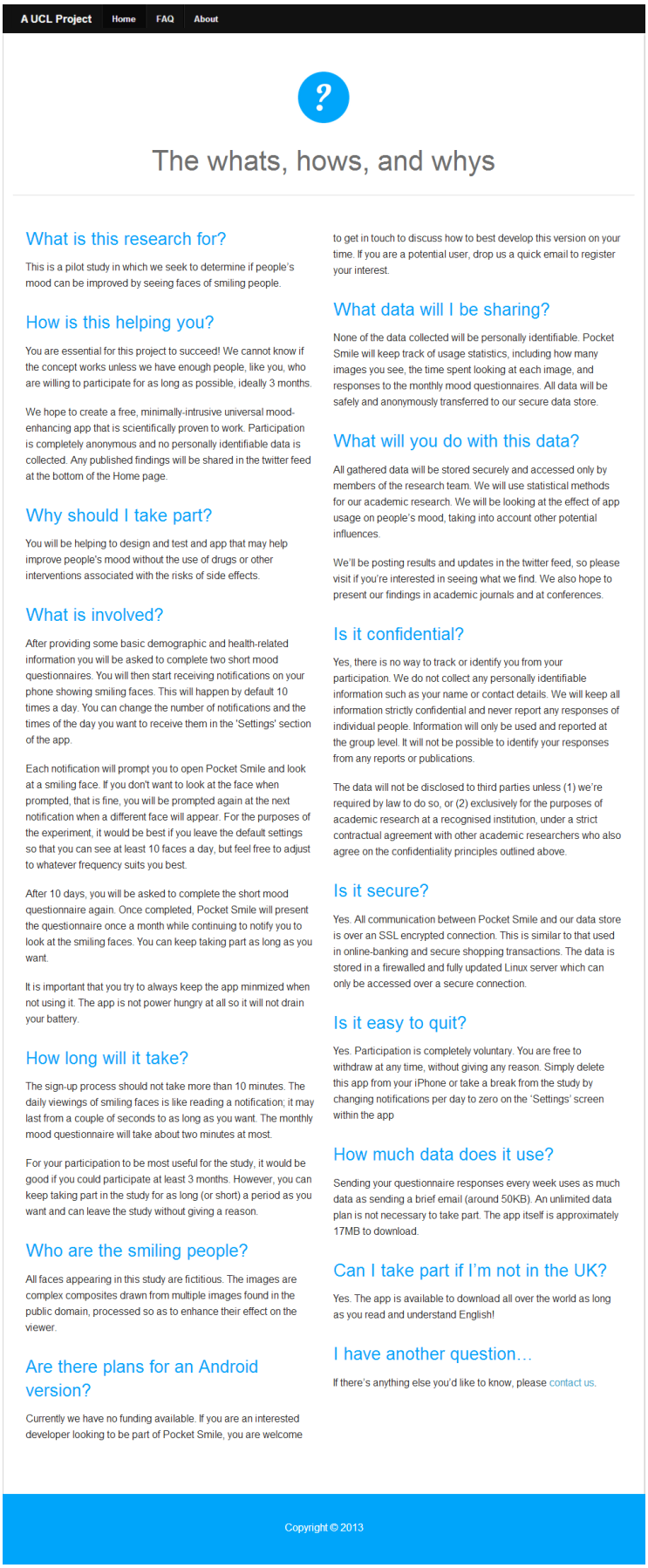
B.9 Website

The following pages show the site as it appeared on 16.12.13: www.pocketsmile.icn.ucl.ac.uk

Home Page



Frequently Asked Questions Page



A UCL Project Home FAQ About

The whats, hows, and whys

What is this research for?

This is a pilot study in which we seek to determine if people's mood can be improved by seeing faces of smiling people.

How is this helping you?

You are essential for this project to succeed! We cannot know if the concept works unless we have enough people, like you, who are willing to participate for as long as possible, ideally 3 months.

We hope to create a free, minimally-intrusive universal mood-enhancing app that is scientifically proven to work. Participation is completely anonymous and no personally identifiable data is collected. Any published findings will be shared in the twitter feed at the bottom of the Home page.

Why should I take part?

You will be helping to design and test and app that may help improve people's mood without the use of drugs or other interventions associated with the risks of side effects.

What is involved?

After providing some basic demographic and health-related information you will be asked to complete two short mood questionnaires. You will then start receiving notifications on your phone showing smiling faces. This will happen by default 10 times a day. You can change the number of notifications and the times of the day you want to receive them in the 'Settings' section of the app.

Each notification will prompt you to open Pocket Smile and look at a smiling face. If you don't want to look at the face when prompted, that is fine, you will be prompted again at the next notification when a different face will appear. For the purposes of the experiment, it would be best if you leave the default settings so that you can see at least 10 faces a day, but feel free to adjust to whatever frequency suits you best.

After 10 days, you will be asked to complete the short mood questionnaire again. Once completed, Pocket Smile will present the questionnaire once a month while continuing to notify you to look at the smiling faces. You can keep taking part as long as you want.

It is important that you try to always keep the app minmized when not using it. The app is not power hungry at all so it will not drain your battery.

How long will it take?

The sign-up process should not take more than 10 minutes. The daily viewings of smiling faces is like reading a notification; it may last from a couple of seconds to as long as you want. The monthly mood questionnaire will take about two minutes at most.

For your participation to be most useful for the study, it would be good if you could participate at least 3 months. However, you can keep taking part in the study for as long (or short) a period as you want and can leave the study without giving a reason.

Who are the smiling people?

All faces appearing in this study are fictitious. The images are complex composites drawn from multiple images found in the public domain, processed so as to enhance their effect on the viewer.

Are there plans for an Android version?

Currently we have no funding available. If you are an interested developer looking to be part of Pocket Smile, you are welcome to get in touch to discuss how to best develop this version on your time. If you are a potential user, drop us a quick email to register your interest.

What data will I be sharing?

None of the data collected will be personally identifiable. Pocket Smile will keep track of usage statistics, including how many images you see, the time spent looking at each image, and responses to the monthly mood questionnaires. All data will be safely and anonymously transferred to our secure data store.

What will you do with this data?

All gathered data will be stored securely and accessed only by members of the research team. We will use statistical methods for our academic research. We will be looking at the effect of app usage on people's mood, taking into account other potential influences.

We'll be posting results and updates in the twitter feed, so please visit if you're interested in seeing what we find. We also hope to present our findings in academic journals and at conferences.

Is it confidential?

Yes, there is no way to track or identify you from your participation. We do not collect any personally identifiable information such as your name or contact details. We will keep all information strictly confidential and never report any responses of individual people. Information will only be used and reported at the group level. It will not be possible to identify your responses from any reports or publications.

The data will not be disclosed to third parties unless (1) we're required by law to do so, or (2) exclusively for the purposes of academic research at a recognised institution, under a strict contractual agreement with other academic researchers who also agree on the confidentiality principles outlined above.

Is it secure?

Yes. All communication between Pocket Smile and our data store is over an SSL encrypted connection. This is similar to that used in online-banking and secure shopping transactions. The data is stored in a firewalled and fully updated Linux server which can only be accessed over a secure connection.

Is it easy to quit?

Yes. Participation is completely voluntary. You are free to withdraw at any time, without giving any reason. Simply delete this app from your iPhone or take a break from the study by changing notifications per day to zero on the 'Settings' screen within the app

How much data does it use?

Sending your questionnaire responses every week uses as much data as sending a brief email (around 50KB). An unlimited data plan is not necessary to take part. The app itself is approximately 17MB to download.

Can I take part if I'm not in the UK?

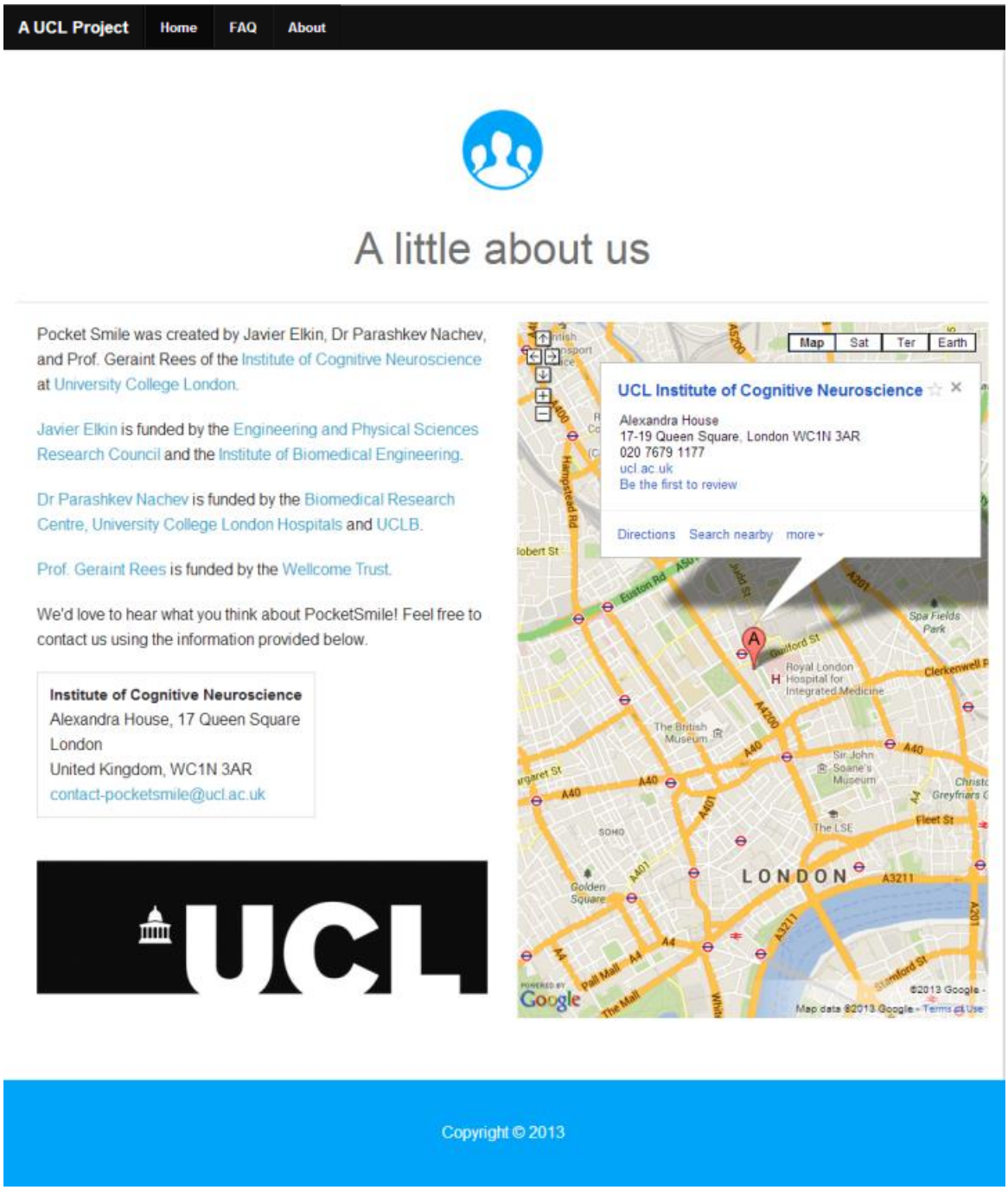
Yes. The app is available to download all over the world as long as you read and understand English!

I have another question…

If there's anything else you'd like to know, please contact us.

Copyright © 2013

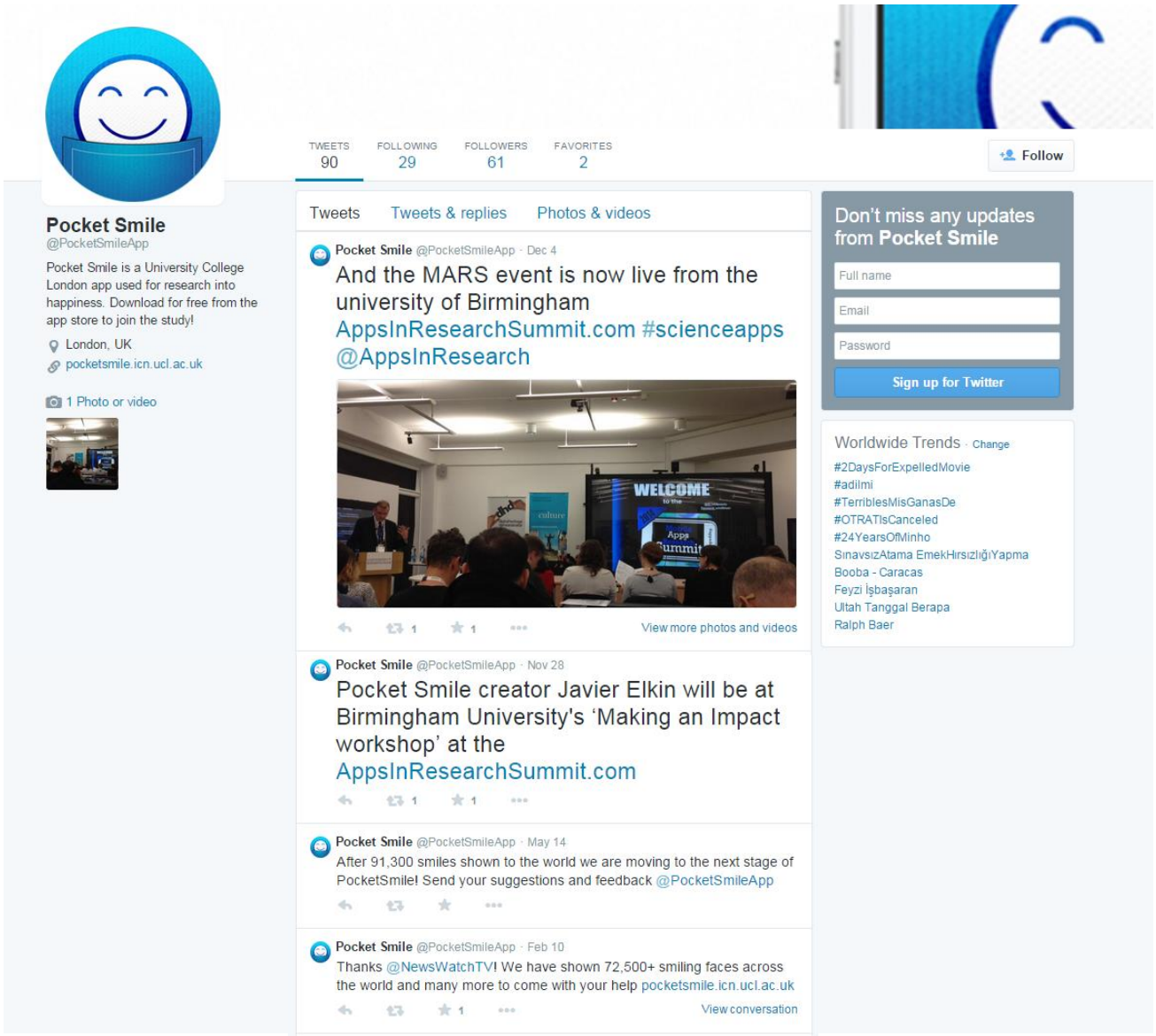
About Page


A UCL Project
Home
FAQ
About
A little about us
Pocket Smile was created by Javier Elkin, Dr Parashkev Nachev, and Prof. Geraint Rees of the Institute of Cognitive Neuroscience at University College London.
Javier Elkin is funded by the Engineering and Physical Sciences Research Council and the Institute of Biomedical Engineering.
Dr Parashkev Nachev is funded by the Biomedical Research Centre, University College London Hospitals and UCLB.
Prof. Geraint Rees is funded by the Wellcome Trust.
We'd love to hear what you think about PocketSmile! Feel free to contact us using the information provided below.
Institute of Cognitive Neuroscience
Alexandra House, 17 Queen Square
London
United Kingdom, WC1N 3AR
contact-pocketsmile@ucl.ac.uk
UCL
Map
Sat
Ter
Earth
UCL Institute of Cognitive Neuroscience
Alexandra House
17-19 Queen Square, London WC1N 3AR
020 7679 1177
ucl.ac.uk
Be the first to review
Directions
Search nearby
more
LONDON
Copyright © 2013

B.10 Social Media Presence

Twitter profile, example screenshot as of 08/12/2014 of Twitter handle @PocketSmileApp.

Link: https://twitter.com/PocketSmileApp



Facebook profile, example screenshot as of 08/12/2014.

Link: https://www.facebook.com/pocketsmileapp

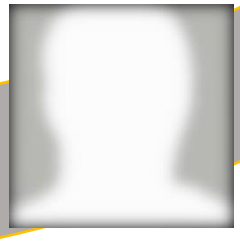
B.11 Sample image creation process

1. Selected images from internet

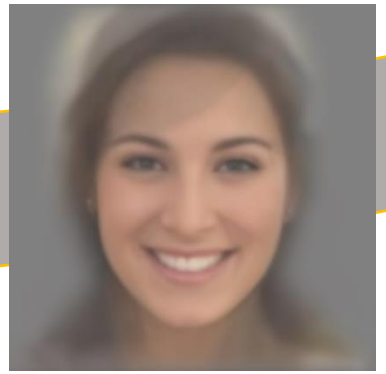
2. Features delineated in Abrasoft Facemixer 3

3. Faces averaged

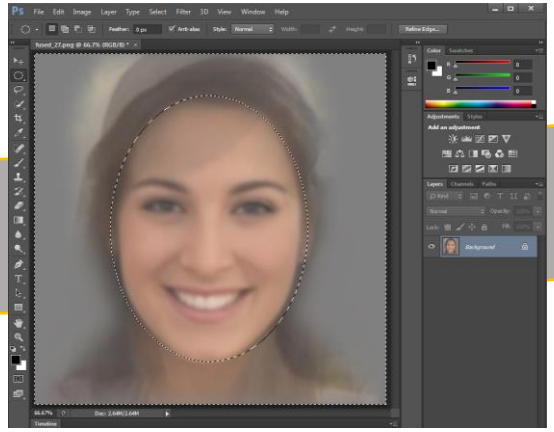
4. Masks created in MIPAV

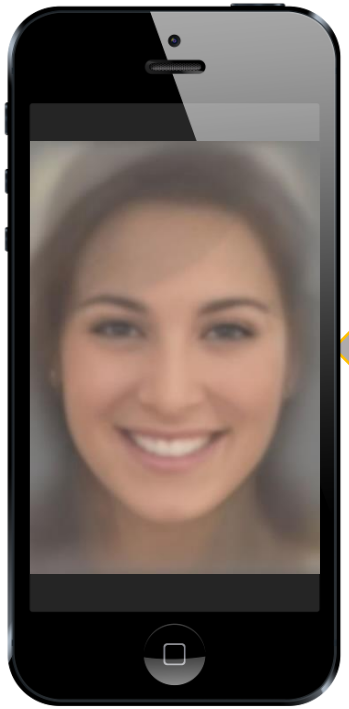


5. Masks multiplied with faces in MATLAB

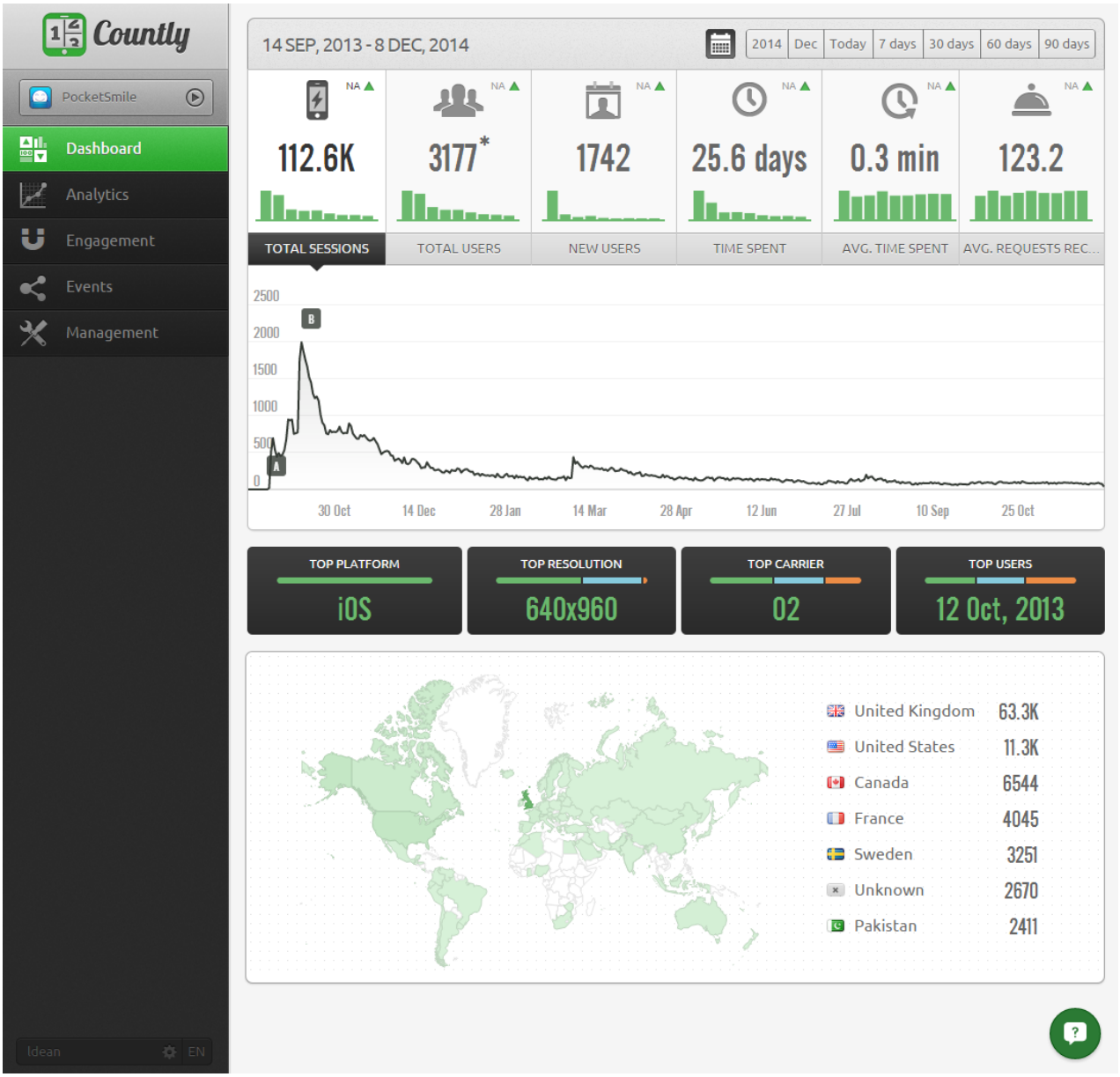


6. Background blurred and cropped in Photoshop CS6





B.12. Analytics screenshot



B.13 Mixed Effects Repeated Measures Results

TABLE B.14 – Mixed effects repeated measures results for both depression and happiness baseline measures to change in scores at every timepoint

	Test	*Coefficient*	*p*	95% CI	Test	*Coefficient*	*p*	95% CI
Timepoints								
T1	PHQ*	-1.53	<0.001	-1.88 to -1.17	SHS**	.181	<0.001	.017 to .345
T2		-1.88		-2.39 to -1.37		.463		.227 to .698
T3		-2.46		-3.09 to -1.83		.591		.301 to .880
T4		-2.15		-2.93 to -1.37		.749		.383 to 1.11
T5		-1.54		-2.51 to -.584		.545		.098 to .992
T6		-1.53		-2.58 to -.484		.398		-.090 to .886
T7		-2.00		-3.17 to -.823		.188		-.358 to .734
	* Higher PHQ score indicates higher depression				** Higher SHS score indicates higher happiness			

B.14 Source Code for Pocket Smile

The Tapparium developer provided a link to the code repository which is found at the link below:

https://bitbucket.org/arivibes/pocket-smile

Access to the full code can be provided upon request.

B.15 Source Code for Pocket Smile 2

The codes below were provided by development agency Milo Creative. This is not the full code as a React Native application consists of several repositories with multiple sub-folders and specific rules that govern them.

```
'use strict';
// Default React imports
import React, { Component } from 'react';
import { StyleSheet, View, Navigator, Text, TouchableHighlight,
   Linking, AppState, PushNotificationIOS }
 from 'react-native';

// import utilites
import colors from './styles/Colors'
import moment from 'moment'

// import services / APIs
import ServerService from './services/server/ServerService'
import Notifications from './services/notifications'

// import NavBar
import MyNavBar from './components/nav/MyNavBar'

// Import Stores & Actions
import {NavActions} from './actions/Actions';
import NavStore from './stores/NavStore';
import ProfileStore from './stores/ProfileStore';

// Import Components
import FormMenu from './components/forms'
import TabNavigation from './components/TabNavigation'
import MoodQuestionnaire from './components/forms/MoodQuestionnaire'
import RecentEvents from './components/forms/RecentEvents'
import Signup from './components/forms/Signup'
import AboutBirthYear from './components/forms/AboutBirthYear'
import AboutSatisfaction from './components/forms/AboutSatisfaction'
import AboutThanks from './components/forms/AboutThanks'
import AboutGeneric from './components/forms/AboutGeneric'
import ImageView from './components/image'
import Onboarding from './components/onboarding'


// Define main class / object
var App = React.createClass({
 getInitialState() {
  let initialRoute = this._initialRoute()
  return ({
   initialRoute: initialRoute,
   navBarHidden: initialRoute.navBarHidden,
  });
 },
 componentWillMount() {
  this.aboutJSON = require('./components/forms/json/about.json');
  console.log('componentWillMount');
 },
 componentDidMount() {
  console.log('componentDidMount');
  AppState.addEventListener('change', this._handleAppStateChange);
  this.unsubscribe  = NavStore.listen(this._buttonPress)
  Notifications.addListener(this._handleNotification)
  Notifications.requestPermissionsIfNeeded()
  this._handleNotification2IfNeeded()
  console.log('_sendJsonIfNeeded IN componentDidMount');
  this._sendJsonIfNeeded()
 },
```

```
componentWillUnmount() {
  console.log('componentWillUnmount');
  this.unsubscribe()
  AppState.removeEventListener('change', this._handleAppStateChange)
  Notifications.removeListener(this.__handleNotification)
},
_initialRoute() {
  let routes = [
    {name:'onboarding',title:' ', navBarHidden:true},
    {name:'form_menu',title:' ', navBarHidden:true,isForm:true},
    {name:'tab_navigation',title:'Settings', navBarHidden:true}
  ]
  let route = routes[0]
  if (ProfileStore.getProperty('seenOnboarding')) {
    if (!ProfileStore.getProperty('hasCompletedInitialQuestionnaire') ||
        ProfileStore.shouldDisplayQuestionnaire()) {
      route = routes[1]
      if (ProfileStore.shouldDisplayQuestionnaire()) {
        ProfileStore.setProperty('formStep',2)
      }
    } else {
      route = routes[2]
    }
  }
  return {name: route.name, title:route.title}
},
_getCurrentRoute() {
  const routes = this.refs.navigator.getCurrentRoutes()
  let route = routes[routes.length-1]
  return route
},
// this a callback gets called when notification is pressed or if the user is within the app
_handleNotification(notification) {
  console.log('_handleNotification:\n'+JSON.stringify(notification));
  this._handleNotificationInternal(notification)
},
// this gets called when app state changes or when app is reloaded (componentDidMount)
_handleNotification2IfNeeded() {
  console.log('_handleNotification2IfNeeded');
  Notifications.hasNotification(this,() =>   {
    this._handleNotificationInternal()
  })
},
_handleNotificationInternal(notification) {
  console.log('_handleNotificationInternal ');
  let isQuestionnaireNotification = notification ?
    Notifications.isQuestionnaireNotification(notification) :
    false;
  if (isQuestionnaireNotification || ProfileStore.shouldDisplayQuestionnaire()) {
    // Don't display questionnaire if already filling questionnaire
    // e.g when an image notification comes whilst filling forms
    if (this._getCurrentRoute().isForm) return
    this._buttonPress(NavActions.showQuestionnaire)
  } else {
    // Don't display image on top of forms (ie not in tab_navigation)
    let route = this._getCurrentRoute()
    console.log('_handleNotificationInternal ' + route.isForm + ' ' + route.name);
    if (route.isForm || route.name === 'image_view') return
    let imageIndex = ProfileStore.nextImageIndex()
    this._buttonPress(NavActions.showImage, imageIndex)
    console.log('- showImage')
  }
  Notifications.clearBadges();
},
_handleAppStateChange(currentAppState) {
  console.log('_handleAppStateChange:\n'+currentAppState);
```

```
    if (currentAppState == 'active') {
      this._handleNotification2IfNeeded()
      console.log('_sendJsonIfNeeded IN _handleAppStateChange');
      this._sendJsonIfNeeded()
    } else {
      const routes = this.refs.navigator.getCurrentRoutes()
      let route = routes[routes.length-1]
      if (route.name == 'image_view') {
        this._buttonPress(NavActions.leftImage)
        ProfileStore.imageDisappeared()
      }
    }
  },
  _sendJsonIfNeeded() {
    if (ProfileStore.getProperty('hasCompletedInitialQuestionnaire') &&
        !ProfileStore.getProperty('hasSentInitialQuestionnaire')) {
          this._sendInitialJson()
    }
    if (ProfileStore.getProperty('hasCompletedMidQuestionnaire') &&
        !ProfileStore.getProperty('hasSentMidQuestionnaire')) {
          this._sendMidJson()
    }
    if (ProfileStore.getProperty('hasCompletedFinalQuestionnaire') &&
        !ProfileStore.getProperty('hasSentFinalQuestionnaire')) {
          this._sendFinalJson()
    }
  },
  _sendInitialJson() {
    ServerService.send(ProfileStore.initialServerJson(),this._initialServerSuccess)
  },
  _sendMidJson() {
    ServerService.send(ProfileStore.midServerJson(),this._midServerSuccess)
  },
  _sendFinalJson() {
    ServerService.send(ProfileStore.finalServerJson(),this._finalServerSuccess)
  },
  _initialServerSuccess() {
    ProfileStore.setProperty('hasSentInitialQuestionnaire', true)
  },
  _midServerSuccess() {
    ProfileStore.setProperty('hasSentMidQuestionnaire', true)
  },
  _finalServerSuccess() {
    ProfileStore.setProperty('hasSentFinalQuestionnaire', true)
  },
  _buttonPress(action, argObj) {
    if (action == NavActions.onboardingGetstartedPressed) {
      this.refs.navigator.resetTo({name:'form_menu', title:' ',isForm:true})
      ProfileStore.setProperty('seenOnboarding', true)
    } else if (action == NavActions.tabItemPressed) {
      this.refs.navigator._navBar.setTitle(argObj)
    } else if (action == NavActions.formMenuItemPressed) {
      this.refs.navigator._navBar.setTitle(undefined)
      this.setState({navBarHidden:false})
      let onPress = () => NavActions.formThanksDonePressed()
      if (argObj == 'Sign up') {
        this.refs.navigator.push({name:'sign_up', title:'Sign up',
        leftButton:'times', isForm:true})
      } else if (argObj == 'About you') {
        this.refs.navigator.push({name:'aboutGender', title:'About you',
        leftButton:'times',isForm:true})
      } else if (argObj == 'Mood questions') {
        this.setState({rightButtonDisabled:true, animateRightButton:false})
        onPress = () => {this.refs.navigator.popToTop();
NavActions.formCheckPressed('mood_questionnaire')}
        this.refs.navigator.push({name:'mood_questionnaire', title:'Mood Questions',
```

```
      rightButton:'check', onRightPress:onPress,isForm:true})
    } else if (argObj == 'Recent events') {
      this.setState({rightButtonDisabled:true, animateRightButton:false})
      onPress = () => NavActions.questionnaireComplete()
      this.refs.navigator.push({name:'recent_events', title:'Recent Events',
      rightButton:'check', onRightPress:onPress,isForm:true})
    }
  } else if (action == NavActions.recentFormChanged ||
        action == NavActions.moodFormChanged) {
    this.setState({animateRightButton:argObj, rightButtonDisabled:!argObj})
  } else if (action == NavActions.aboutFormValueEntered) {
    this.refs.navigator.push({name:argObj, title:'About you',
    leftButton:'times'})
  } else if (action  == NavActions.aboutComplete) {
    this.refs.navigator.popToTop()
  } else if (action  == NavActions.questionnaireComplete) {
    // Fixes bug that would send questionnaireComplete consecutily when pressing on the animated
button
    const timeNow = (new Date()).getTime()
    if (this.questionnaireCompleteTimeCalled &&
      timeNow - this.questionnaireCompleteTimeCalled.getTime() < 60000) return
    this.setState({navBarHidden:false})
    if (!ProfileStore.getProperty('hasCompletedInitialQuestionnaire')) {
      ProfileStore.setProperty('hasCompletedInitialQuestionnaire', true)
      ProfileStore.setProperty('timeCompletedInitialQuestionnaire', moment())
      this._sendInitialJson()
      if (!ProfileStore.getProperty('hasCompletedMidQuestionnaire')) {
        Notifications.removeQuestionnaireNotificationJs();
        Notifications.scheduleQuestionnaireNotificationIfNeeded(1)
      }
    } else if (!ProfileStore.getProperty('hasCompletedMidQuestionnaire')) {
      ProfileStore.setProperty('hasCompletedMidQuestionnaire', true)
      ProfileStore.setProperty('hasToCompleteMidQuestionnaire', false)
      ProfileStore.setProperty('timeCompletedMidQuestionnaire', moment())
      this._sendMidJson()
      ProfileStore.switchStimulus()
      if (!ProfileStore.getProperty('hasCompletedFinalQuestionnaire')) {
        Notifications.removeQuestionnaireNotificationJs();
        Notifications.scheduleQuestionnaireNotificationIfNeeded(2)
      }
    } else if (!ProfileStore.getProperty('hasCompletedFinalQuestionnaire')) {
      ProfileStore.setProperty('hasCompletedFinalQuestionnaire', true)
      ProfileStore.setProperty('hasToCompleteFinalQuestionnaire', false)
      ProfileStore.setProperty('hasCompletedStudy', true)
      Notifications.cancelNotifications({type:'questionnaire'})
      this._sendFinalJson()
      console.log("_sendFinalJson");
    }
    this.refs.navigator.resetTo({name: 'tab_navigation',title:'Settings',
    rightButton:null, leftButton:null});
    this.questionnaireCompleteTimeCalled = new Date()
  } else if (action  == NavActions.showImage) {
    this.refs.navigator.resetTo({name: 'image_view',title:'Settings',
    rightButton:null, leftButton:null, imageIndex:argObj});
    this.setState({navBarHidden:true})
  } else if (action == NavActions.leftImage) {
    this.setState({navBarHidden:false})
    this.refs.navigator.resetTo({name: 'tab_navigation',title:'Settings',
    rightButton:null, leftButton:null});
  } else if (action == NavActions.showQuestionnaire) {
    ProfileStore.setProperty('formStep',2)
    if (!ProfileStore.getProperty('hasCompletedMidQuestionnaire')) {
      ProfileStore.setProperty('hasToCompleteMidQuestionnaire',true)
    } else if (!ProfileStore.getProperty('hasCompletedFinalQuestionnaire')) {
      ProfileStore.setProperty('hasToCompleteFinalQuestionnaire',true)
    }
```

```
    this.setState({navBarHidden:true})
    this.refs.navigator.resetTo({name:'form_menu', title:' ',isForm:true})
  } else if (action == NavActions.showAppIcon) {
    this.refs.navigator._navBar.setShowAppIcon(argObj)
  }
 },
 _renderScene(route) {
  switch (route.name) {
    case 'onboarding':
      return <Onboarding />
    case 'form_menu':
      return <FormMenu />
    case 'tab_navigation':
      return <TabNavigation />
    case 'sign_up':
      return <Signup fileName='consent' onPress={
        () => {this.refs.navigator.pop(); NavActions.formCheckPressed('sign_up')}
      }/>
    case 'mood_questionnaire':
      return <MoodQuestionnaire />
    case 'recent_events':
      return <RecentEvents />
    case 'aboutGender':
      return this._aboutGeneric('gender','aboutBirthYear')
    case 'aboutBirthYear':
      return <AboutBirthYear fieldRef='birthYear' label={this.aboutJSON['birthYear'].label}
        onPress={() => this._buttonPress(NavActions.aboutFormValueEntered, 'aboutSatisfaction')} />
    case 'aboutSatisfaction':
      return <AboutSatisfaction fieldRef='satisfaction' label={this.aboutJSON['satisfaction'].label}
      onPress={() => this._buttonPress(NavActions.aboutFormValueEntered, 'aboutHealth')} />
    case 'aboutHealth':
      return this._aboutGeneric('health','aboutRelationship')
    case 'aboutRelationship':
      return this._aboutGeneric('relationship','aboutEmployment')
    case 'aboutEmployment':
      return this._aboutGeneric('employment','aboutUse')
    case 'aboutUse':
      return this._aboutGeneric('usedBefore','aboutThanks')
    case 'aboutThanks':
      return <AboutThanks />
    case 'image_view':
      return <ImageView imageIndex={route.imageIndex} />
    default:
      console.error('Encountered unexpected route: ' + route.name);
  }
 },

 _aboutGeneric(fieldRef, nextRoute) {
  return (
    <AboutGeneric
      fieldRef={fieldRef}
      options={this.aboutJSON[fieldRef].options}
      label={this.aboutJSON[fieldRef].label}
      onPress={() => this._buttonPress(NavActions.aboutFormValueEntered, nextRoute)}
    />
  )
 },

 _configureScene : function(route){
  // this method calls for transition animation
  switch (route.name) {
    case 'webviewer':
    case 'aboutBirthYear':
    case 'aboutSatisfaction':
    case 'aboutHealth':
    case 'aboutRelationship':
```

```
      case 'aboutEmployment':
      case 'aboutUse':
      case 'aboutThanks':
      case 'form_menu':
        return Navigator.SceneConfigs.PushFromRight;
      default:
        return Navigator.SceneConfigs.FloatFromBottom;

    }
  },

  render() {
    // let initialRoute = {name: 'tab_navigation',
    //   title:'Home', rightButton:null, leftButton:null};
    return (
      <View style={styles.container}>
        <Navigator
          ref='navigator'
          initialRoute={this.state.initialRoute}
          configureScene={this._configureScene}
          renderScene={this._renderScene}
          navigationBar={
            <MyNavBar ref='navBar'
              hidden={this.state.navBarHidden}
              rightButtonDisabled={this.state.rightButtonDisabled}
              animateRightButton={this.state.animateRightButton}
            />}
          onDidFocus={this._onDidFocus}
        />
      </View>
    );
  },
});



var styles = StyleSheet.create({
  container: {
    flex: 1,
  },
});
module.exports = App;
```

```
import jstz from 'jstz'
import moment from 'moment'
import NumberService from '../services/number'

// about: gender, birthYear, satisfaction, health, relationship, employment
// consent, moodQuestionnaire, recentEvents,
// state: formStep:1,2,3,4 (Signup, About you, Mood questions, Recent Events)
//        stimulus
class Profile {
  constructor() {
    console.log('Profile::constructor')
    //this.id = md5('profile')
    this.id = window.uuid
    // forms
    this.about = {}
    const timezone = jstz.determine()
    this.about.timeZone = timezone.name()
    // faces
    this.faces = []
    //stores currently scheduled local notifications in Android (alt : iOS's getScheduledLocalNotifications)
    this.scheduledLocalNotifications = []
    // set app state
    this.formStep = 0
```

```
  this.seenOnboarding = false
  this.imageType = NumberService.randomBoolean() ? 'f' : 'l'
  this.imageIndex = 0

  this.hasCompletedInitialQuestionnaire = false
  this.hasCompletedMidQuestionnaire = false
  this.hasCompletedFinalQuestionnaire = false
  this.hasCompletedStudy = false
 }

 setFromObject(obj) {
  this.setProperty('id', obj.id)
  // forms values
  this.setProperty('about', obj.about)
  this.setProperty('moodQuestionnaire', obj.moodQuestionnaire)
  this.setProperty('midMoodQuestionnaire', obj.midMoodQuestionnaire)
  this.setProperty('finalMoodQuestionnaire', obj.finalMoodQuestionnaire)
  this.setProperty('recentEvents', obj.recentEvents)
  this.setProperty('midRecentEvents', obj.midRecentEvents)
  this.setProperty('finalRecentEvents', obj.finalRecentEvents)
  // Faces
  this.setProperty('faces', obj.faces)
  //Notifications
  this.setProperty('scheduledLocalNotifications', obj.scheduledLocalNotifications)
  // settings
  this.setProperty('settings', obj.settings)
  // app state
  this.setProperty('formStep', obj.formStep)
  this.setProperty('seenOnboarding', obj.seenOnboarding)
  this.setProperty('imageType', obj.imageType)
  this.setProperty('imageIndex', obj.imageIndex)

  this.setProperty('hasSentInitialQuestionnaire', obj.hasSentInitialQuestionnaire)
  this.setProperty('hasSentMidQuestionnaire', obj.hasSentMidQuestionnaire)
  this.setProperty('hasSentFinalQuestionnaire', obj.hasSentFinalQuestionnaire)

  this.setProperty('hasCompletedInitialQuestionnaire', obj.hasCompletedInitialQuestionnaire)
  this.setProperty('hasToCompleteMidQuestionnaire', obj.hasToCompleteMidQuestionnaire)
  this.setProperty('hasCompletedMidQuestionnaire', obj.hasCompletedMidQuestionnaire)
  this.setProperty('hasToCompleteFinalQuestionnaire', obj.hasToCompleteFinalQuestionnaire)
  this.setProperty('hasCompletedFinalQuestionnaire', obj.hasCompletedFinalQuestionnaire)
  this.setProperty('hasCompletedStudy', obj.hasCompletedStudy)
  if (obj.timeCompletedInitialQuestionnaire)
    this.setProperty('timeCompletedInitialQuestionnaire', moment(obj.timeCompletedInitialQuestionnaire))
  if (obj.timeCompletedMidQuestionnaire)
    this.setProperty('timeCompletedMidQuestionnaire', moment(obj.timeCompletedMidQuestionnaire))
 }

 static fromObject(obj) {
  let profile = new Profile();
  profile.setFromObject(obj);
  return profile;
 }

 setAboutProperty(prop, value) {
  this.about[prop] = value
 }

 getAboutProperty(prop) {
  return this.about[prop]
 }

 setProperty(prop, value) {
  this[prop] = value
 }
```

```
  getProperty(prop) {
    return this[prop]
  }

  imageAppeared() {
    let time = moment().format("DD/MM/YYYY [at] HH:mm:ss")
    this.faces.push({'time':time, action:'appeared', imageType:this.imageType})
  }

  imageDisappeared() {
    let time = moment().format("DD/MM/YYYY [at] HH:mm:ss")
    this.faces.push({'time':time, action:'disappeared', imageType:this.imageType})
  }

  isPastMidStage() {
    let diff = moment().diff(this.timeCompletedInitialQuestionnaire)
    let dur = moment.duration(diff)

    //return dur.asHours() >= 1
    //return dur.asMinutes() >= 5
    return dur.asDays() >= 10
  }

  isPastFinalStage() {
    let diff = moment().diff(this.timeCompletedMidQuestionnaire)
    let dur = moment.duration(diff)

    //return dur.asHours() >= 1
    //return dur.asMinutes() >= 5
    return dur.asDays() >= 10
  }

}

module.exports = Profile;
```

```
import React, { Component } from 'react';
import {AppRegistry} from 'react-native';

import App from './App'
import Message from './components/utils/Message'
import ProfileStore from './stores/ProfileStore'
import RNUUIDGenerator from 'react-native-uuid-generator';
import codePush from "react-native-code-push";
import GoogleAnalytics from 'react-native-google-analytics-bridge'

// AppLoader loads store and
// sets UUID if this has not been set yet

class AppLoader extends Component {
  constructor(prop) {
    super(prop)
    this.state = {
      loadCompleted:false
    }
  }

  async loadStore() {
    var val = await ProfileStore._load()
    this.setState({loadCompleted:true})
  }

  componentWillMount() {
    this.loadStore()
    RNUUIDGenerator.getRandomUUID((uuid) => {
      this.setState({uuid:uuid})
    });
```

```
    GoogleAnalytics.setTrackerId('UA-85059618-1');
    GoogleAnalytics.setDispatchInterval(30);
  }

  // componentDidMount() {
  //   codePush.sync({
  //     updateDialog: true,
  //     installMode: codePush.InstallMode.IMMEDIATE
  //   });
  // }

  render() {
    if (this.state.loadCompleted && this.state.uuid) {
      if (!ProfileStore.getProperty('id')) {
        console.log('PocketSmile::render, uuid:'+this.state.uuid);
        ProfileStore.setProperty('id', this.state.uuid)
      }
      return (<App />)
    } else {
      return (<Message label='Loading...' />)
    }
  }
}

let codePushOptions = { checkFrequency: codePush.CheckFrequency.ON_APP_RESUME };

AppLoader = codePush(codePushOptions)(AppLoader);

module.exports = AppLoader
```

Appendix C – Supporting Material for the Science Museum Experiment

Museum accessibility | Contact us | Sign up to our newsletter

Search

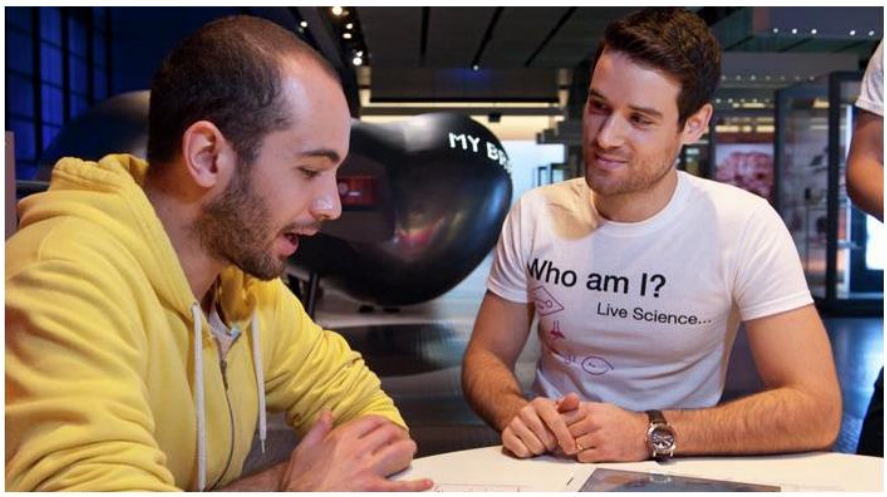
Home | Visit us | Online science | Educators | Shop online | About us

Visit us

Open seven days a week, 10.00-18.00. Entry to the Museum is free.

Plan your visit | Getting here | Tickets and prices | Eating | Shop | Museum accessibility

Live Science



Live Science is an ongoing project where scientists come into the Science Museum to carry out research using Museum visitors as volunteers. You can take part in their scientific experiments to find out more about yourself.

Nothing dangerous... just fun, interesting experiments.

Latest experiment: Can Playing a Video Game Change You?

Every day except Thursday until 18 December 2016
Times: 11:00-13.30, 14:30-17:30

Can playing a video game change how you think? Researchers from University College London are running this study to learn more about how playing video games can affect us.

Who can take part?

The experiment takes under 10 minutes. Anyone aged 6 or over can volunteer. Written consent is required for everybody. If you are under 13, your parent or legal guardian will have to sign the consent form.

Live Science

Join drop-in sessions throughout the day (except between 13.30-14.30) and take part in fun, interesting experiments.

Floor: 1
Price: **Free**

FREE

Subjects

Current science

Medicine and biology

Age ranges

Age 6-11

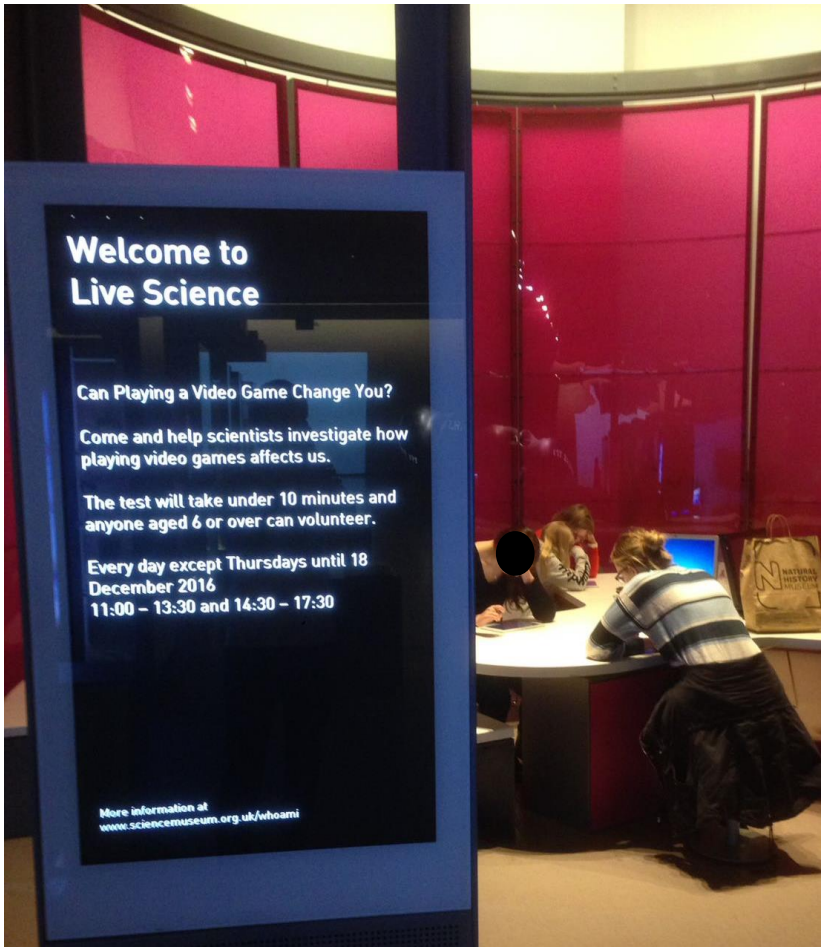
Age 12-16

Adults

Who am I?

Explore the science of who you are through objects, artwork and hands-on exhibits

C.1 Informational adverts (continued)



An advert in front of the Live Science space in the *Who Am I?* gallery of the Science Museum with participants in the background.

C.2 Videogame storyboard

Can playing a video game change you?

Designed and developed by:

Filippo Aiello
www.filippoaiello.it

Supported by:

UCL

Consent and participant information

Can Playing a Video Game Change You?

RESET

Version 5 01.11.2016

Dear Visitor,

You are invited to take part in a real research study as part of the Science Museum's Live Science programme. Live Science is an ongoing project where real scientists come into the Museum to carry out their research with Museum visitors as volunteers.

What is the experiment about?

Can playing a video game change how you think? Researchers from University College London are running this study to learn more about how playing video games can affect us. There is a surprise at the end!

What will happen in the experiment?

If you decide to take part, you will be asked play a simple card-matching game on a tablet. If two cards match they stay turned up, and you win the level when all cards are facing up. The experiment will take about 3–8 minutes to complete but there is no time limit. We will also ask for some basic information about you, such as your age and gender, and also will ask you to complete a questionnaire. This information will be kept strictly confidential and remain anonymous.

NEXT

☐ I agree

Consent and participant information

Can Playing a Video Game Change You?

RESET

will be kept strictly confidential and remain anonymous.

Why are we doing this study?
There is debate about whether video games can improve certain mental skills. We are running this study to see how our brains work while we play video games.

Who can take part in the study?
Anyone aged 6 or older is welcome to take part.

What do I need to do if I want to take part?
If you would like to take part, please ask one of the researchers. If you are aged 13 or older you will have to give your consent below to give the researcher permission to carry out the study. If you are aged of 12 or younger your parent or legal guardian will be asked to agree on this digital form. Please let the researchers know if you would like your own copy. This is standard procedure for all research with visitors.

Can I change my mind?
You can change your mind at any time during the study. It is completely voluntary and you do not have to give a reason if you want to stop at any point. If the researchers have already collected your data it will not be used in the final study and will be destroyed.

☐ I agree

NEXT

Consent and participant information

Can Playing a Video Game Change You?

RESET

What will happen to the results of this research?

Your recorded data will be stored in compliance with the Data Protection Act 1998. Any personal details recorded will be stored securely at University College London for a period of 5 years and then destroyed. Only authorised researchers will have access to this information. The study results will help us understand how our brains work in certain situations and may be published in scientific journals or presented at scientific conferences. Your name and identity will not be revealed. We can provide a summary and a final report to participants on request.

How do I find out more about the study?

You are very welcome to ask the researchers directly if you have any questions. You may also e-mail questions to Javier Elkin:
javier.elkin.10@ucl.ac.uk
For more information and to see the results of this study please visit our website:
pocketsmile.icn.ucl.ac.uk
The principal investigator in this study is Professor Geraint Rees, Dean of the UCL Faculty of Life Sciences, Institute of Cognitive Neuroscience, 17 Queen Square, London, WC1N 3AR.
If you subsequently have a complaint, please contact:
University College London Research Ethics Committee
(Project ID: 4746/002), Gower Street, London WC1E 6BT.

☐ I agree

NEXT

Consent and participant information

Can Playing a Video Game Change You?

RESET

.
The principal investigator in this study is Professor Geraint Rees, Dean of the UCL Faculty of Life Sciences, Institute of Cognitive Neuroscience, 17 Queen Square, London, WC1N 3AR.
If you subsequently have a complaint, please contact:
University College London Research Ethics Committee
(Project ID: 4746/002), Gower Street, London WC1E 6BT.

Consent

By tapping "I agree" below, you confirm that you:

- Have read and understood the information presented below detailing this study
- Had the opportunity to consider the information and ask questions
- Had your questions answered satisfactorily
- Understood that your participation is voluntary and that you can withdraw at any time without giving a reason
- Understood that your data and the personal details provided on this form will be stored securely at University College London, and only accessed by authorised researchers
- Understood that in all instances where the results are shared they will remain anon ymous
- Are aged 13 or older OR you are the legal guardian or parent of a child aged 6 or older.

☐ I agree

NEXT

About you

RESET

What is your sex?

- ○ Unselected
- ◉ Male
- ○ Female

What is your age?

- ○ Unselected
- ○ 6 - 11
- ○ 12 - 17
- ○ 18 - 24
- ◉ 25 - 29
- ○ 30 - 34
- ○ 35 - 39
- ○ 40 - 44
- ○ 45 - 49
- ○ 50 - 59
- ○ 60+

In general, how is your health?

- ○ Unselected
- ○ Very good
- ○ Good
- ○ Fair
- ◉ Poor
- ○ I prefer not to say

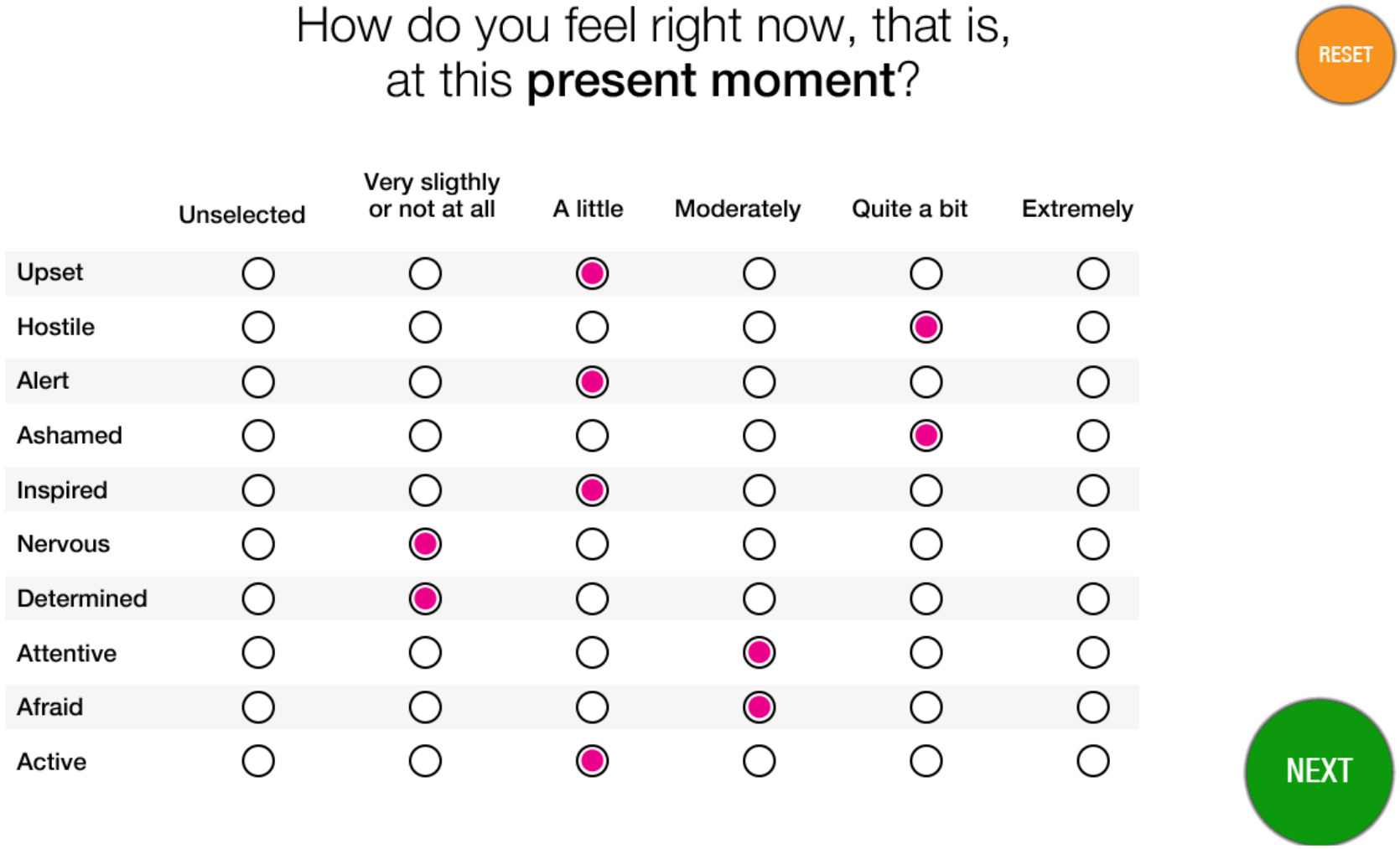
What is your relationship status?

- ○ Unselected
- ◉ Single
- ○ In a relationship
- ○ Married
- ○ Widowed
- ○ Divorced
- ○ I prefer not to say

What is your occupation?

- ○ Unselected
- ○ Student
- ◉ Employed
- ○ Retired
- ○ Unemployed
- ○ I prefer not to say

NEXT

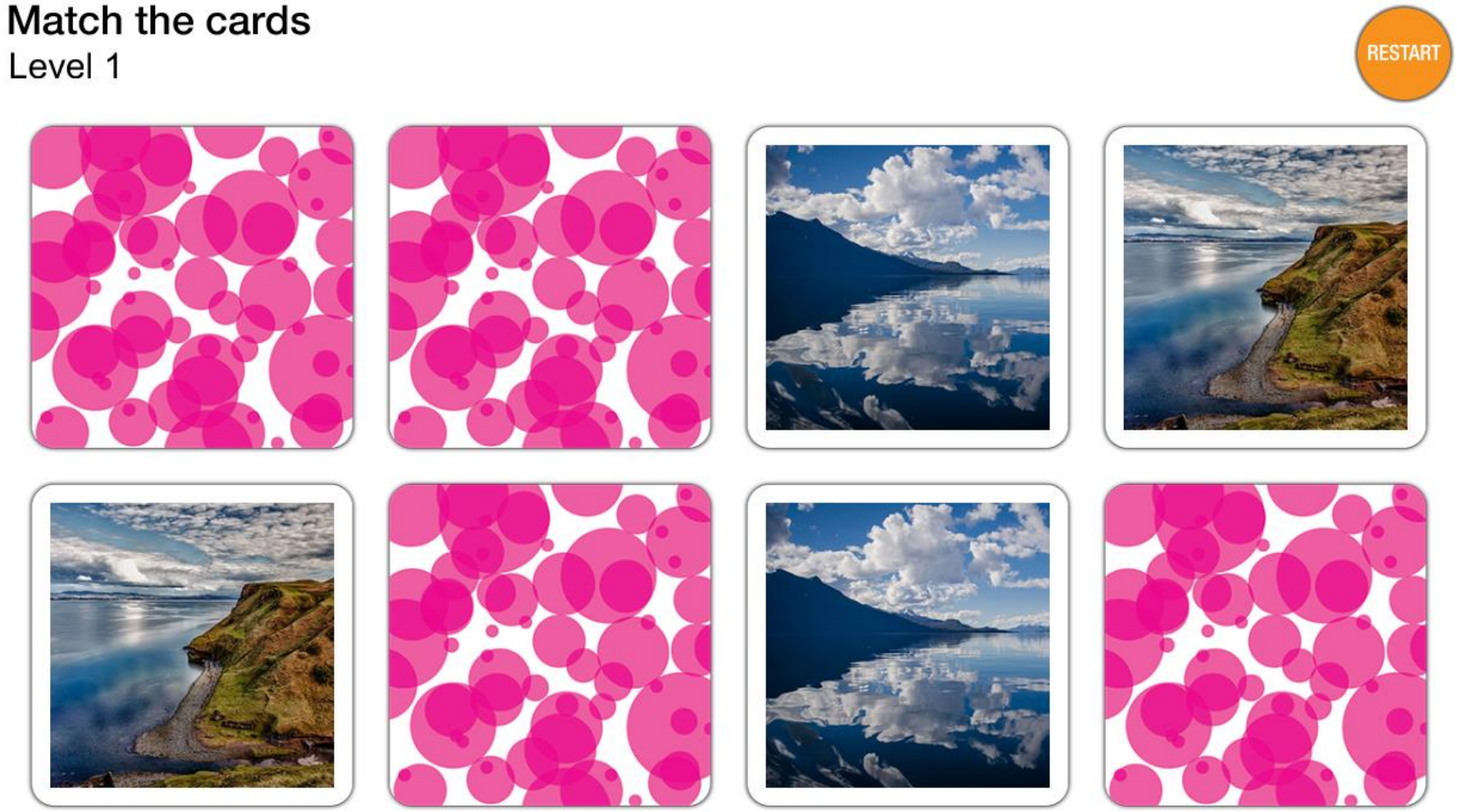

How do you feel right now, that is, at this present moment?
RESET
Unselected
Very sligthly or not at all
A little
Moderately
Quite a bit
Extremely
Upset
Hostile
Alert
Ashamed
Inspired
Nervous
Determined
Attentive
Afraid
Active
NEXT

Are you ready to play?

This is a card game where you have to pair matching images. It begins with all the cards flipped down and you can only turn them around two at a time by tapping on them. If the two cards match, they stay turned up and you win the level when all cards are facing up.

There are three levels that will increase in difficulty: easy, medium, and hard. The whole game is no longer than **8 minutes** total but there is no time limit.

Option 1: Landscapes as stimuli



Match the cards

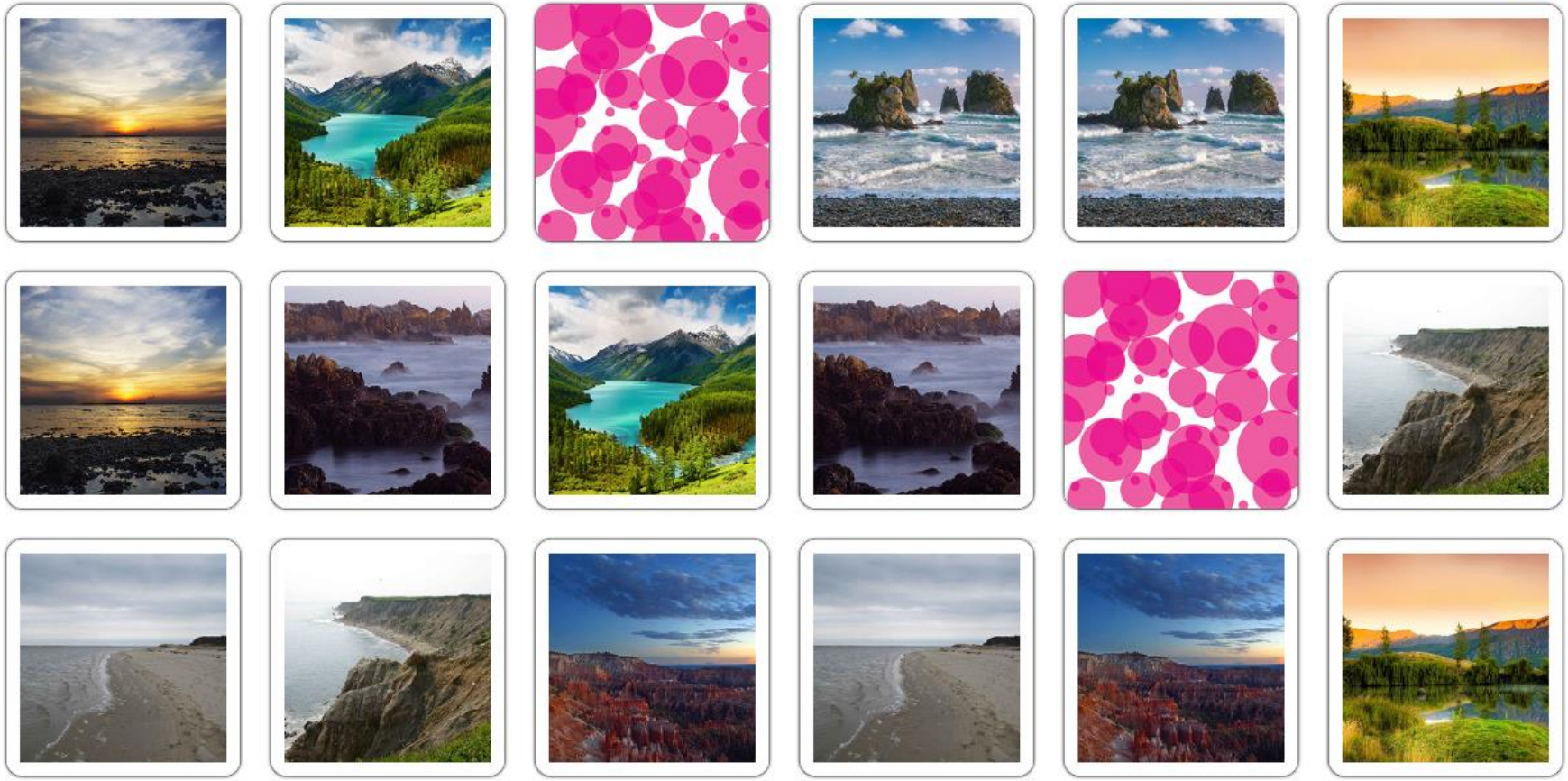
Level 2

RESET

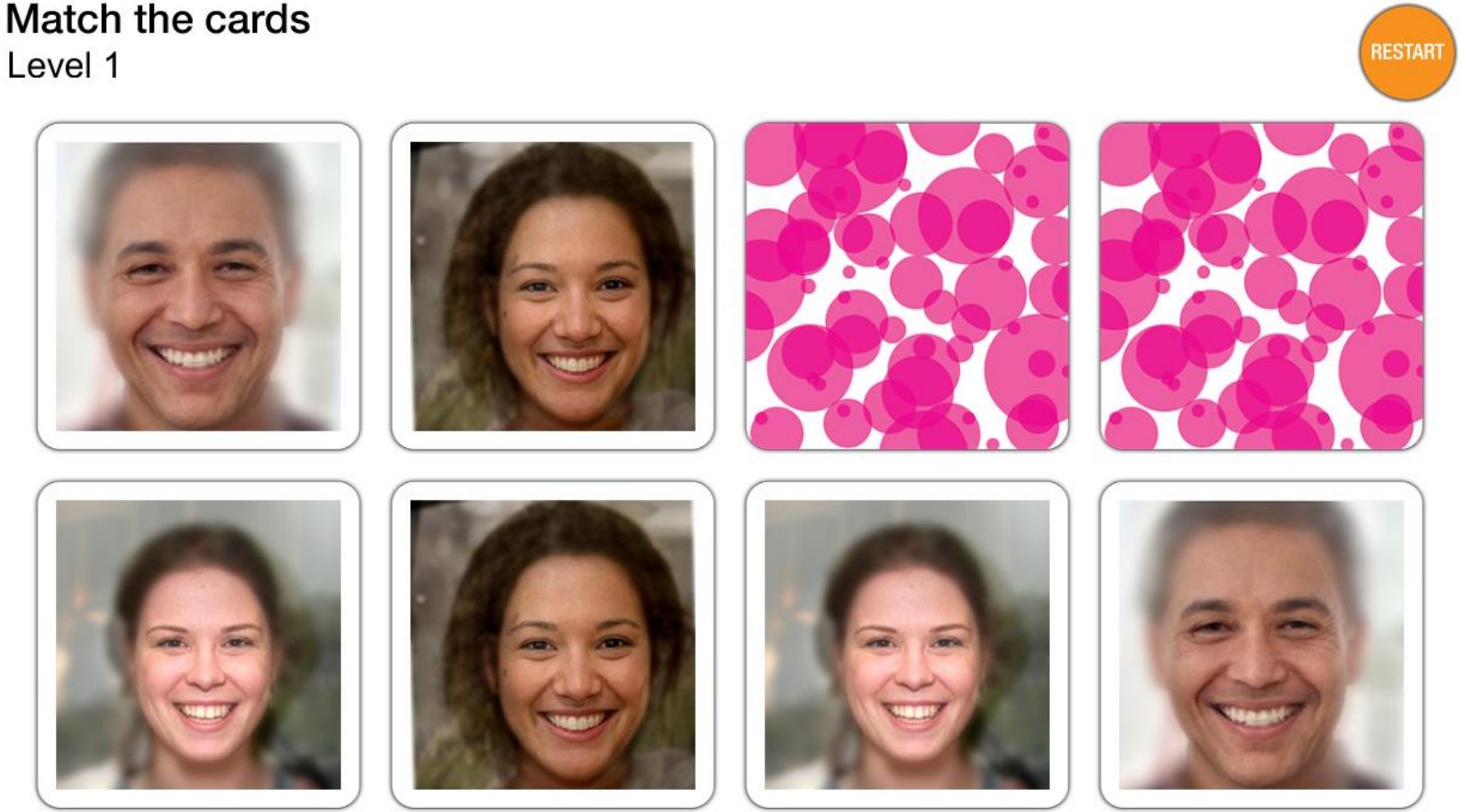
Match the cards

Level 3

RESET



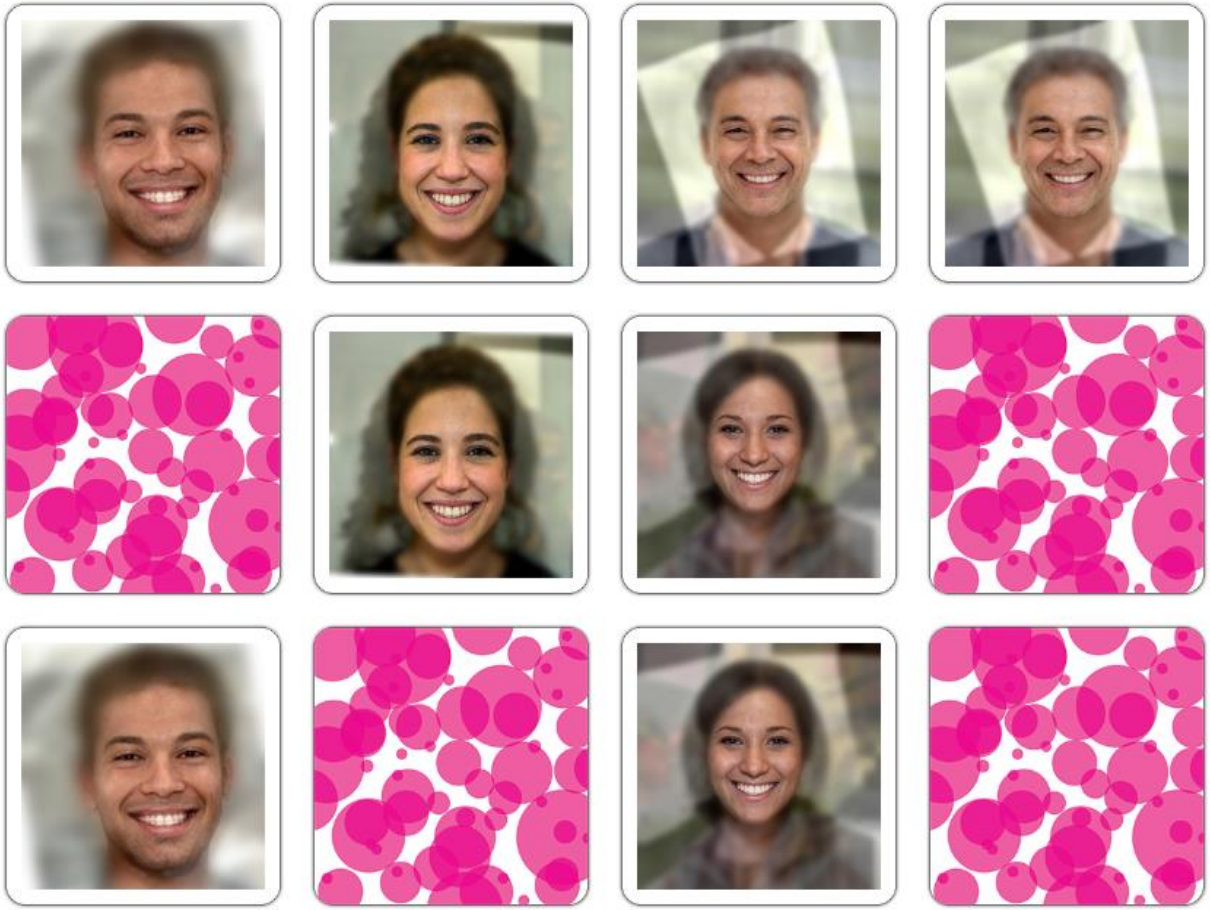
Option 2: Smiling faces as stimuli



Match the cards

Level 2

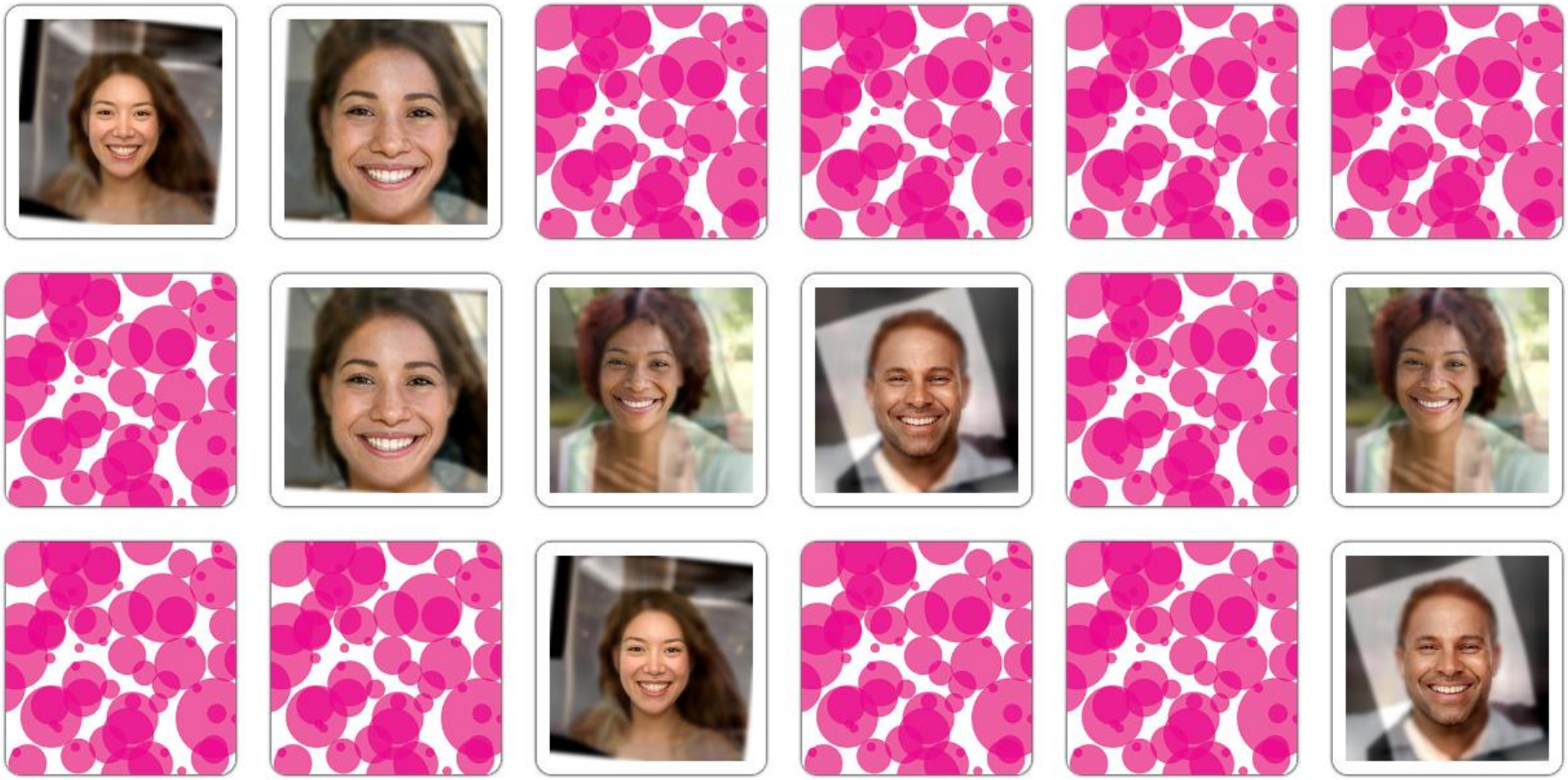
RESTART

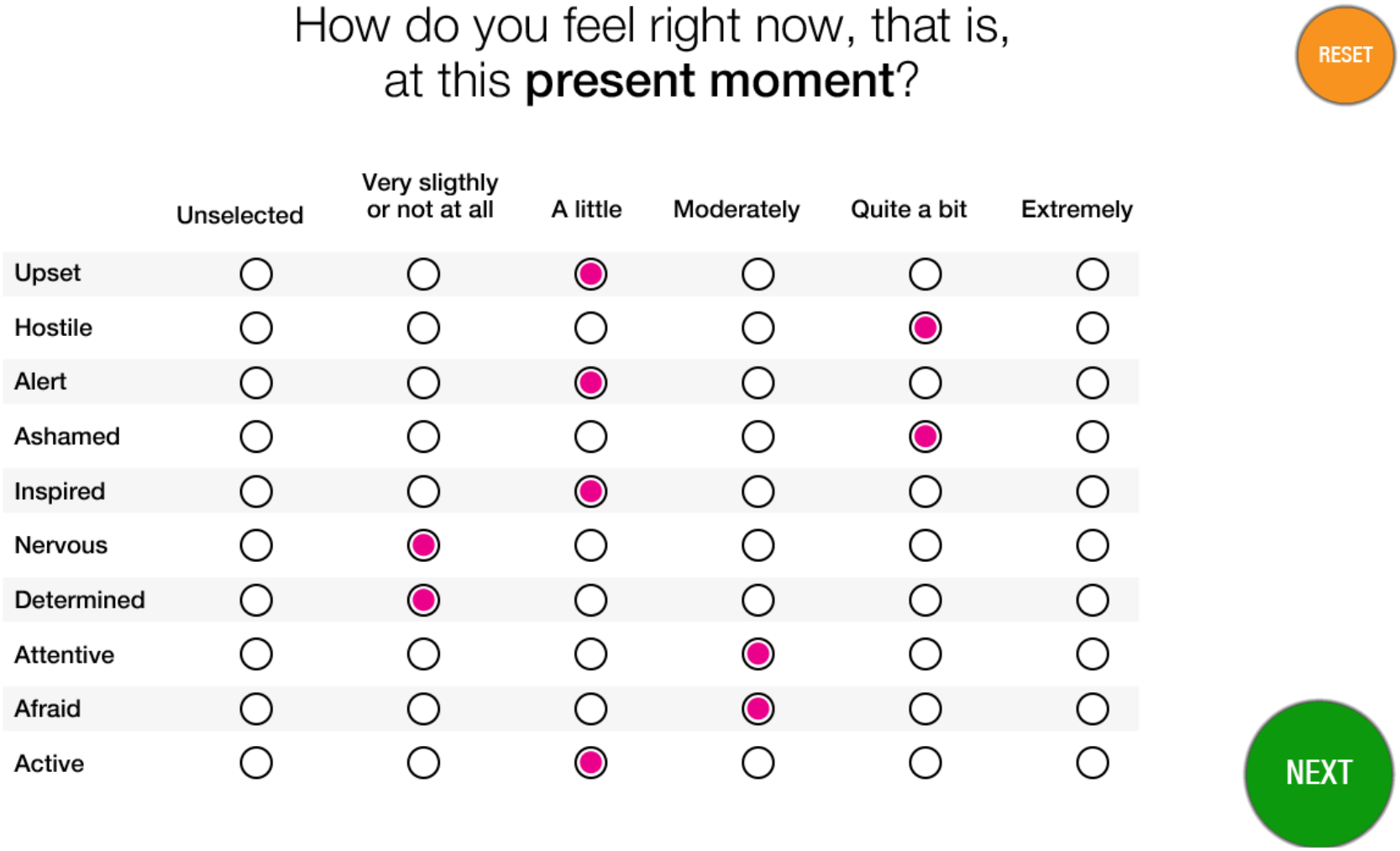


Match the cards

Level 3

RESTART




How do you feel right now, that is, at this present moment?
RESET
Unselected
Very sligthly or not at all
A little
Moderately
Quite a bit
Extremely
Upset
Hostile
Alert
Ashamed
Inspired
Nervous
Determined
Attentive
Afraid
Active
NEXT

Thank you for playing this game and taking part in Live Science!

Did you play the game with smiling faces or did you see landscapes? We were interested to find out how playing this game made you feel.

Do you feel happy when you see someone smiling? Our experiment wants to test if seeing photos of people smiling makes you happier, compared with seeing landscapes.

If something as simple as seeing a photo of someone smiling could brighten up your day, this could be used to create new types of therapy.

We'd like to share with you what we find out. If you'd like to hear about the results of this study, please visit: **pocketsmile.icn.ucl.ac.uk**

The experiment is now over, but if you'd like to play a bonus round of the game with both faces and landscapes please click below.

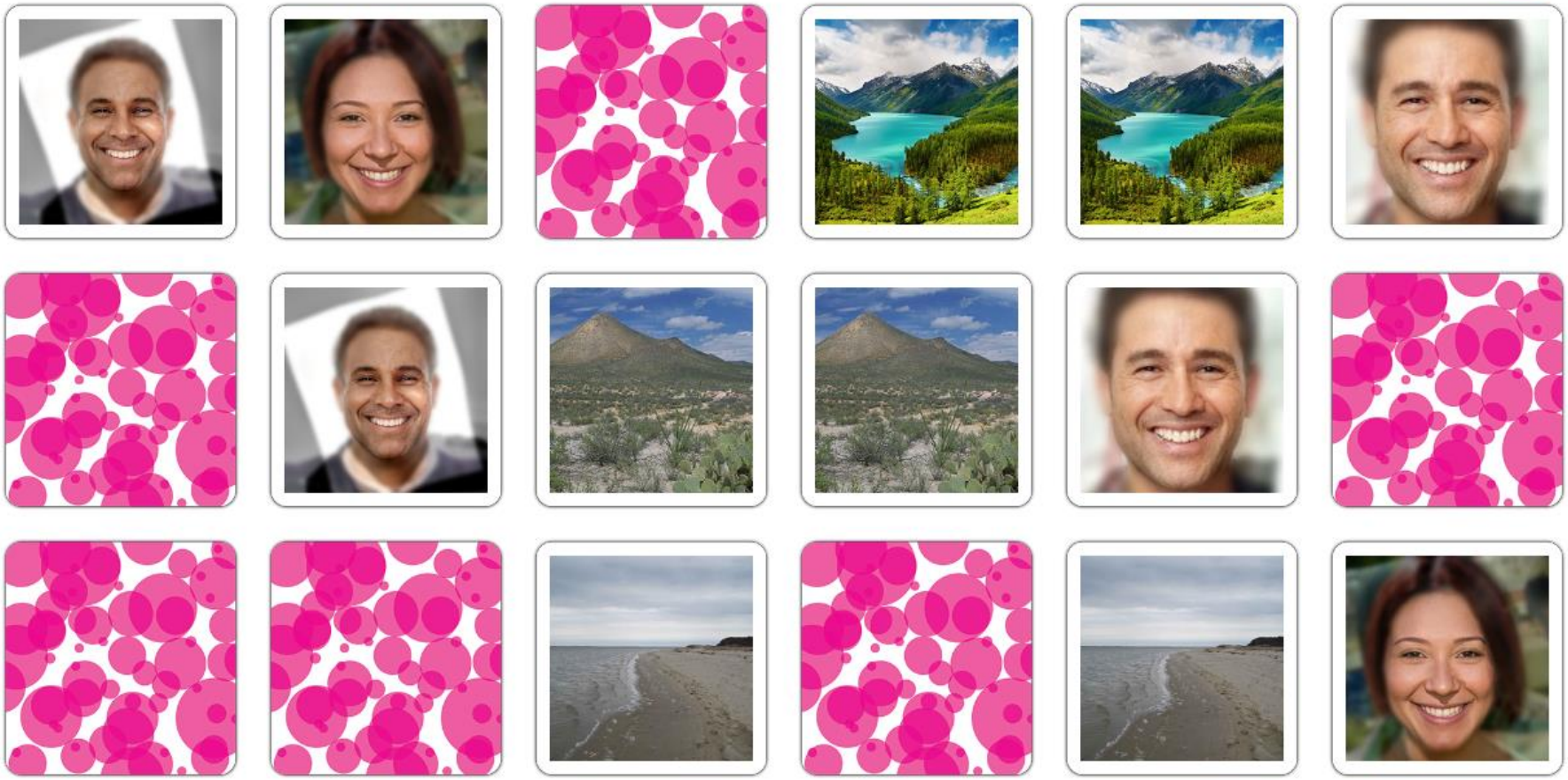
END THIS SESSION

PLAY BONUS ROUND

Match the cards

Level bonus

RESTART



Thank you for again for taking part in this study

This session has now ended

This application will restart automatically in:

14 seconds

RESTART NOW

Designed and developed by:

Filippo Aiello
www.filippoaiello.it

Supported by:

C.3 Source code

Parts of the source code provided by Filippo Aiello for the Android game experiment.

```
package
{       import flash.display.Sprite;
        import flash.display.StageDisplayState;
        import flash.display.StageAlign;
        import flash.display.StageQuality;
        import flash.display.StageScaleMode;
        import flash.events.Event;
        import com.mesmotronic.ane.AndroidFullScreen;
        import com.greensock.TweenMax;
        import org.casalib.util.StageReference;
        import starling.core.Starling;
        [SWF(width = "1280", height = "800", frameRate = "60", backgroundColor = "#FFFFFF")] //762
        public class ScienceMuseum_Mobile extends Sprite
        {       protected var _starling:Starling;
                public function ScienceMuseum_Mobile()
                {       if (stage) this.onAddedToStage();
                        else this.addEventListener(flash.events.Event.ADDED_TO_STAGE,
this.onAddedToStage);
                }
                private function onAddedToStage(event:flash.events.Event = null):void {
                        this.removeEventListener(flash.events.Event.ADDED_TO_STAGE,
this.onAddedToStage);
                        this.stage.displayState = StageDisplayState.NORMAL;
                        StageReference.setStage(this.stage);
                        stage.quality = StageQuality.BEST;
                        stage.align = StageAlign.TOP_LEFT;
                        stage.scaleMode = StageScaleMode.NO_SCALE;
                        this.visible = false;
                        TweenMax.to({}, 0.001, {});
                        this._starling = new Starling(Game, this.stage, null, null, "auto", "auto");
                        this._starling.antiAliasing = 2;
                        _starling.start();
                        setAndroidFullScreen();
                        }
                        private function setAndroidFullScreen():void
```

```
                {AndroidFullScreen.stage = stage;

                        // ANE v1.3.x

                        if (!AndroidFullScreen.immersiveMode())

                        {stage.displayState = StageDisplayState.FULL_SCREEN_INTERACTIVE;}

                        AndroidFullScreen.immersiveMode();

                        // Properties

                        /*      AndroidFullScreen.stage = stage; // Set this to your app's stage

                        AndroidFullScreen.isSupported; // Is this ANE supported?

                        AndroidFullScreen.isImmersiveModeSupported; // Is immersive mode
supported?

                        AndroidFullScreen.immersiveWidth; // The width of the screen in immersive
mode

                        AndroidFullScreen.immersiveHeight; // The height of the screen in immersive
mode

                        AndroidFullScreen.fullScreenWidth; // The width of the screen in the best
available full screen mode

                        AndroidFullScreen.fullScreenHeight; // The height of the screen in the best
available full screen mode

                        // Methods

                        AndroidFullScreen.fullScreen(); // Switch your app to the best available full
screen mode

                        AndroidFullScreen.showSystemUI(); // Show system UI

                        AndroidFullScreen.leanMode(); // Hide system UI until user interacts

                        AndroidFullScreen.showUnderSystemUI(); // Extend your app underneath the
system UI (Android 4.4+ only)

                        AndroidFullScreen.immersiveMode(); // Hide system UI and keep it hidden
(Android 4.4+ only)

                        AndroidFullScreen.immersiveMode(false); // Hide system UI until user swipes
from top (Android 4.4+ only)*/ }}}
```

Package

```
{

        import feathers.controls.text.BitmapFontTextRenderer;

        import starling.core.Starling;
        import starling.utils.AssetManager;
        import starling.textures.TextureAtlas;
        import starling.text.BitmapFont;
        import starling.textures.Texture;

        public class Assets
        {
                private static var gameTextureAtlas:Texture

                [Embed(source="../assets/textures/spritesheet.png")]
                private static const atlas:Class;

                [Embed(source="../assets/textures/spritesheet.xml", mimeType="application/octet-
stream")]
```

```
		private static const atlas_xml:Class;

		[Embed(source="../assets/textures/people.png")]
		private static const atlasPeople:Class;

		[Embed(source="../assets/textures/people.xml", mimeType="application/octet-
stream")]
		private static const atlasPeople_xml:Class;

		[Embed(source="../assets/textures/landscape.png")]
		private static const atlasLandscape:Class;

		[Embed(source="../assets/textures/landscape.xml", mimeType="application/octet-
stream")]
		private static const atlasLandscape_xml:Class;

		[Embed(source="../assets/fonts/HelveticaLight.png")]
		public static const FontTexture:Class;

		[Embed(source="../assets/fonts/HelveticaLight.fnt", mimeType="application/octet-
stream")]
		public static const FontXML:Class;

		public static var myFont:BitmapFontTextRenderer

		private static var _atlas:TextureAtlas;
		private static var _atlasPeople:TextureAtlas;
		private static var _atlasLandscape:TextureAtlas;

		private static var sAssets:AssetManager;

		public function start(assets:AssetManager):void
		{
			sAssets = assets;
		}

		public static function getFont():BitmapFont
		{
			var fontTexture: Texture = Texture.fromBitmap(new FontTexture());
			var fontXML:XML = XML (new FontXML());
			var font: BitmapFont = new BitmapFont(fontTexture, fontXML);

			return font;
		}

		public static function getTexture(name:String):Texture
		{
			init();
			return _atlas.getTexture(name);
		}

		public static function getPeopleTexture(name:String):Texture
		{
			init();
			return _atlasPeople.getTexture(name);

		}

		public static function getLandscapeTexture(name:String):Texture
		{
			init();
			return _atlasLandscape.getTexture(name);

		}

		private static function init():void
```

```
                        {
                                if (Starling.current == null)
                                        throw new Error("Initialize Starling before accessing
textures.");

                                if (_atlas == null)
                                {
                                        var atlasTexture:Texture =
Texture.fromEmbeddedAsset(atlas, false);
                                        var atlasXml:XML = XML(new atlas_xml());

                                        _atlas = new TextureAtlas(atlasTexture, atlasXml);
                                }

                                if (_atlasPeople == null)
                                {
                                        var atlasPeopleTexture:Texture =
Texture.fromEmbeddedAsset(atlasPeople, false);
                                        var atlasPeopleXml:XML = XML(new atlasPeople_xml());
                                        _atlasPeople = new TextureAtlas(atlasPeopleTexture,
atlasPeopleXml);

                                }

                                if (_atlasLandscape == null)
                                {
                                        var atlasLandscapeTexture:Texture =
Texture.fromEmbeddedAsset(atlasLandscape, false);
                                        var atlasLandscapeXml:XML = XML(new
atlasLandscape_xml());
                                        _atlasLandscape = new
TextureAtlas(atlasLandscapeTexture, atlasLandscapeXml);

                                }
                        }
                }
        }
}
<?php

/*
connect to our database
*/

$json=$_POST['dataFromApp'];
(Array)$data = json_decode($json,false);

$output="";


$dbhost = "localhost";
$dbuser = "";
$dbpass= "";
$dbname = "sciencemuseum";

$conn = mysql_connect($dbhost, $dbuser, $dbpass);

  if(! $conn ) {
    die('Could not connect: ' . mysql_error());
  }

  $output="";

  mysql_select_db($dbname);
```

```
  for ($i = 0; $i < count($data); $i++) {

                $id = $data[$i]->id;
                $idDevice = $data[$i]->idDevice;
                $photos = $data[$i]->photos;
                        $time= $data[$i]->time;
                        $totTime=$data[$i]->totTime;
                        $sex= $data[$i]->sex;
                        $age= $data[$i]->age;
                        $health= $data[$i]->health;
                        $status= $data[$i]->status;
                        $occupation= $data[$i]->occupation;
                        $upset_0= $data[$i]->upset_0;
                        $hostile_0= $data[$i]->hostile_0;
                        $alert_0= $data[$i]->alert_0;
                        $ashamed_0= $data[$i]->ashamed_0;
                        $inspired_0= $data[$i]->inspired_0;
                        $nervous_0= $data[$i]->nervous_0;
                        $afraid_0= $data[$i]->afraid_0;
                        $determinated_0= $data[$i]->determinated_0;
                        $attentive_0= $data[$i]->attentive_0;
                        $active_0= $data[$i]->active_0;

                        $upset_1= $data[$i]->upset_1;
                        $hostile_1= $data[$i]->hostile_1;
                        $alert_1= $data[$i]->alert_1;
                        $ashamed_1= $data[$i]->ashamed_1;
                        $inspired_1= $data[$i]->inspired_1;
                        $nervous_1= $data[$i]->nervous_1;
                        $afraid_1= $data[$i]->afraid_1;
                        $determinated_1= $data[$i]->determinated_1;
                        $attentive_1= $data[$i]->attentive_1;
                        $active_1= $data[$i]->active_1;


                        $id = $data[$i]->id;
                        $output.=$id;

                        //$sql = "SELECT TOP 1 1 FROM users WHERE id=$id ORDER BY num DESC limit 250
";
                        //$sql = "SELECT EXISTS (SELECT * FROM users WHERE id=$id ')";

                        $sql = " SELECT * FROM users WHERE id=$id LIMIT 1 ";
                        $result = mysql_query( $sql );
                        //if(mysql_num_rows(mysqli_query($sql)) > 0) {
      if (mysql_fetch_row($result)) {
                        //if(mysql_num_rows($result) > 0) {
                                $output.="true";
                        }
                        else{
                                $query = "INSERT INTO users (id, idDevice, photos, time, totTime, sex,
age, health, status, occupation, upset_0, hostile_0, alert_0, ashamed_0, inspired_0, nervous_0, afraid_0,
determinated_0, attentive_0, active_0, upset_1, hostile_1, alert_1, ashamed_1, inspired_1, nervous_1,
afraid_1, determinated_1, attentive_1, active_1) VALUES ('$id','$idDevice', $photos,'$time','$totTime',
$sex,$age,$health,$status,$occupation,$upset_0,$hostile_0,$alert_0,$ashamed_0,$inspired_0,$nervous_0,
$afraid_0,$determinated_0,$attentive_0,$active_0,$upset_1,$hostile_1,$alert_1,$ashamed_1,$inspired_1,$n
ervous_1,$afraid_1,$determinated_1,$attentive_1,$active_1)";

                                mysql_query($query);
                        }
}

echo "registered=true"."&output=".$output;

  mysql_close($conn);
?>
```

```
package
{
	import flash.events.Event;
	import flash.events.IOErrorEvent;
	import flash.events.SecurityErrorEvent;
	import flash.filesystem.File;
	import flash.filesystem.FileMode;
	import flash.filesystem.FileStream;
	import flash.net.URLLoader;
	import flash.net.URLLoaderDataFormat;
	import flash.net.URLRequest;
	import flash.net.URLRequestMethod;
	import flash.net.URLVariables;

	public class SaveData
	{
		private var internetConnectionStatus:Boolean = false;

		private var request:URLRequest = new URLRequest();
		private var fileXML:XML;
		private var lastRecordedNumber: int = 0;
		private var newRecordedNumber: int = 0;
		private var i: int;
		private var fileStream:FileStream;
		private var message:Array = new Array();
		private var ldr:URLLoader = new URLLoader();
		private var myVariables:URLVariables = new URLVariables;
		private var sendToPHPJson:String;



		public function SaveData()
		{
			request.method	= URLRequestMethod.POST;
			//request.url	= "";
			request.url	= "";
		}


		public function updateXMLfile(value:XML):void{
			var file:File =
File.desktopDirectory.resolvePath("ScienceMuseumsResults");
			file.createDirectory();

			file = file.resolvePath("scienceMuseumResults.xml");
			fileStream = new FileStream();
			if(!file.exists){
				trace("doesn't exist");
				fileStream.open(file, FileMode.WRITE);
				fileStream.writeUTFBytes('<?xml version="1.0"
encoding="UTF-8"?><users>');
				fileStream.writeUTFBytes(value.toXMLString());
				fileStream.writeUTFBytes('</users>');
			}
			else{
				trace("modify");
				fileStream.open(file, FileMode.UPDATE);
				//8 is the length of </users> to overwrite with the new XML
block.
				fileStream.position = fileStream.bytesAvailable-8;
				fileStream.writeUTFBytes(value.toXMLString());
				fileStream.writeUTFBytes('</users>');
			}
			readXML();
		}
```

```
                        public function readXML():void{
                                var file:File =
File.desktopDirectory.resolvePath("ScienceMuseumsResults");
                                file = file.resolvePath("scienceMuseumResults.xml");
                                fileStream = new FileStream();
                                if(file.exists){
                                        trace("exists");
                                        fileStream.openAsync(file, FileMode.READ);
                                        fileStream.addEventListener(Event.COMPLETE,
processXMLData);
                                        fileStream.addEventListener(IOErrorEvent.IO_ERROR,
errorHandler);
                                }
                        }

                        private function processXMLData(event:Event):void
                        {
                                fileXML = XML(fileStream.readUTFBytes(fileStream.bytesAvailable));
                                fileStream.close();
                                createJsonFile();
                        }

                        private function errorHandler(event:flash.events.IOErrorEvent):void {
                                trace("Error found")//not final. Just for testing
                        }

                        private function createJsonFile():void{
                                message.length = 0;
                                newRecordedNumber+= fileXML.child("user").length();
                                for( i = lastRecordedNumber ; i < fileXML.child("user").length() ; i++ ){
                                        //trace("id"+": "+String(fileXML.child("user")[i].child("id")));
                                        message.push ({
                                                id:String(fileXML.child("user")[i].child("id")),
                                                idDevice:
String(fileXML.child("user")[i].child("idDevice")),

        photos:String(fileXML.child("user")[i].child("photos")),
                                                time:String(fileXML.child("user")[i].child("time")),

        totTime:String(fileXML.child("user")[i].child("totTime")),
                                                sex:String(fileXML.child("user")[i].child("sex")),
                                                age:String(fileXML.child("user")[i].child("age")),

        health:String(fileXML.child("user")[i].child("health")),

        status:String(fileXML.child("user")[i].child("status")),

        occupation:String(fileXML.child("user")[i].child("occupation")),


        upset_0:String(fileXML.child("user")[i].child("upset_0")),

        hostile_0:String(fileXML.child("user")[i].child("hostile_0")),

        alert_0:String(fileXML.child("user")[i].child("alert_0")),
                                                ashamed_0:
String(fileXML.child("user")[i].child("ashamed_0")),
                                                inspired_0:
String(fileXML.child("user")[i].child("inspired_0")),
                                                nervous_0:
String(fileXML.child("user")[i].child("nervous_0")),
                                                afraid_0:
String(fileXML.child("user")[i].child("afraid_0")),
                                                determinated_0:
String(fileXML.child("user")[i].child("determinated_0")),
```

```
                                                                attentive_0:
String(fileXML.child("user")[i].child("attentive_0")),
                                                                active_0:
String(fileXML.child("user")[i].child("active_0")),


                upset_1:String(fileXML.child("user")[i].child("upset_1")),

                hostile_1:String(fileXML.child("user")[i].child("hostile_1")),

                alert_1:String(fileXML.child("user")[i].child("alert_1")),
                                                                ashamed_1:
String(fileXML.child("user")[i].child("ashamed_1")),
                                                                inspired_1:
String(fileXML.child("user")[i].child("inspired_1")),
                                                                nervous_1:
String(fileXML.child("user")[i].child("nervous_1")),
                                                                afraid_1:
String(fileXML.child("user")[i].child("afraid_1")),
                                                                determinated_1:
String(fileXML.child("user")[i].child("determinated_1")),
                                                                attentive_1:
String(fileXML.child("user")[i].child("attentive_1")),
                                                                active_1:
String(fileXML.child("user")[i].child("active_1"))
                                                        })
                                        }
                                        trace(JSON.stringify(message));
                                        sendToPHPJson = JSON.stringify(message);
                                        //firstProperty name in PHP
                                        myVariables.dataFromApp = sendToPHPJson;

                                        request.data = myVariables;

                                        ldr.dataFormat = URLLoaderDataFormat.VARIABLES;
                                        ldr.addEventListener(Event.COMPLETE,onComplete);
                                        ldr.addEventListener(IOErrorEvent.IO_ERROR, onError);
                                        ldr.addEventListener(SecurityErrorEvent.SECURITY_ERROR
,onSecurityErr);

                                        ldr.load(request);
                                }

                                private function onComplete(e:Event):void
                                {
                                        trace("onComplete: " +ldr.data);
                                        if(ldr.data.registered =="true"){

                                                lastRecordedNumber = newRecordedNumber;
                                        }
                                        if(ldr.data.registered =="false"){
                                                trace("registered - false: "+ldr.data.error);
                                        }
                                }

                                private function onSecurityErr(e:SecurityErrorEvent):void
                                {
                                        trace("error: " + e.text );
                                }

                                private function onError(e:flash.events.IOErrorEvent):void
                                {
                                        trace("error: " + e.toString());
                                }
                }
}
```