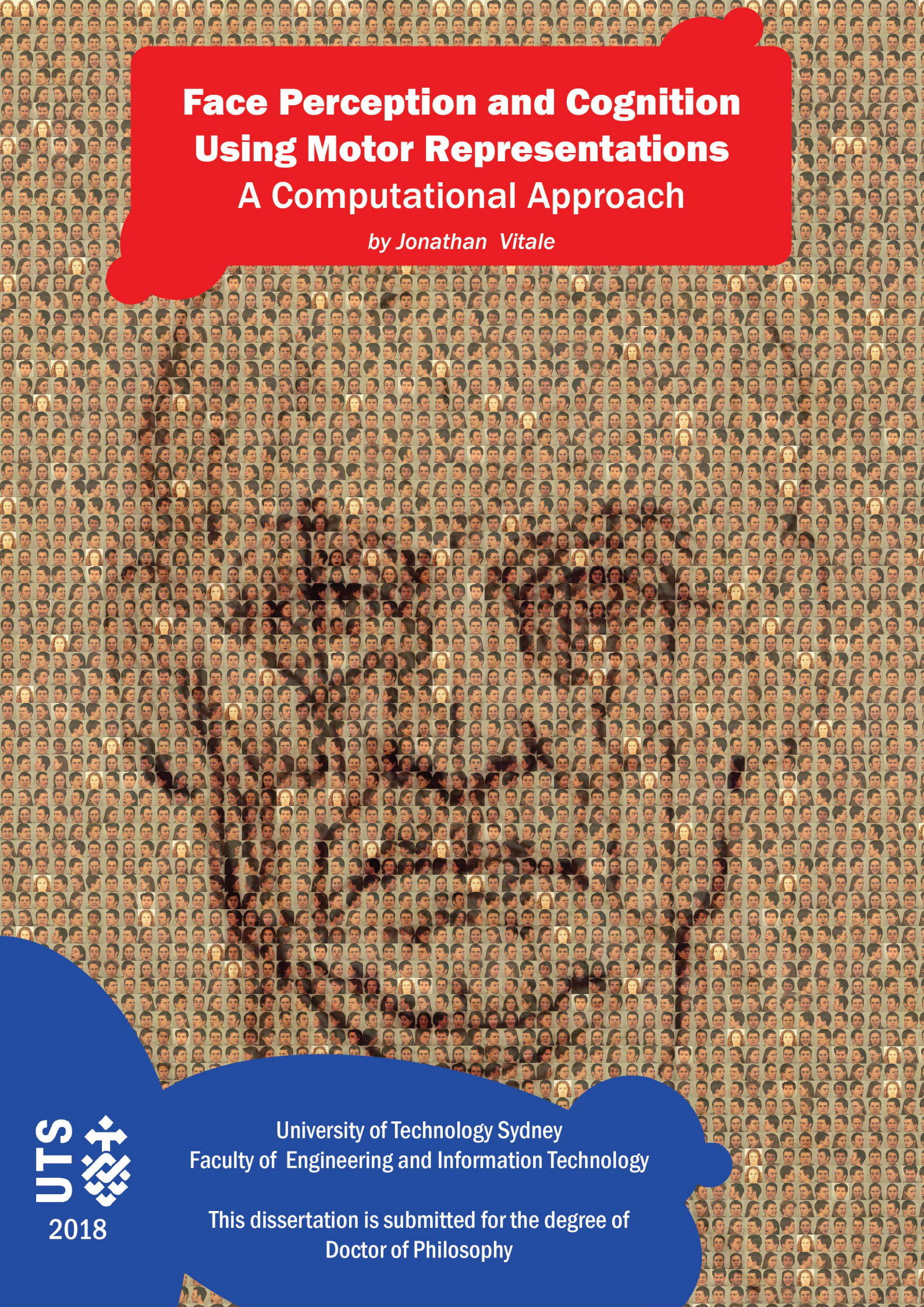


Face Perception and Cognition Using Motor Representations

A Computational Approach

by Jonathan Vitale

University of Technology Sydney
Faculty of Engineering and Information Technology

This dissertation is submitted for the degree of
Doctor of Philosophy

UTS

2018

Face Perception and Cognition Using Motor Representations

A Computational Approach

Jonathan Vitale

Faculty of Engineering and Information Technologies

University of Technology Sydney

This dissertation is submitted for the degree of

Doctor of Philosophy

Supervisor:
Prof. Mary-Anne Williams

Co-supervisor:
Dr Benjamin Johnston

2018

Certificate of Original Authorship

I, Jonathan Vitale declare that this thesis, submitted in fulfilment of the requirements for the award of a Doctor of Philosophy degree, in the Faculty of Engineering and Information Technology (School of Software) at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

Signature of Student:

Production Note:
Signature removed
prior to publication.

Jonathan Vitale

Date: 24/01/2018

I want to dedicate this dissertation to all the people who gave me the opportunity and the necessary motivation to achieve my desired goals, in particular to my parents, my relatives and my close friends.
I also want to make a special dedication to my beloved Niki, who suddenly passed away while I was completing my doctoral degree in Australia, thus breaking my heart for not being able to give her a last goodbye in Italy. Niki, you have been a good companion and an emotional lifesaver during most of my adulthood.

Acknowledgments

I started my PhD research with the ambitious objective of modelling mind-reading capabilities in machines. Obviously, the limited time available to complete my research required me to focus on a particular aspect of social cognition and to narrow down the research to investigate a limited set of research gaps identified in literature. Thus, in this dissertation, I focused the attention on face-to-face social interactions and I explored how face processing capabilities can be explained by embodied mechanisms lying at the core of a mind-reading process. This led to findings advancing both face processing and embodied cognition research.

At the end of this PhD program I still find myself passionate and highly interested in solving one of the most challenges mysteries of human mind, namely how we understand others. However, now I have a better understanding of what is my research community and what is the methodology suitable for my research. I will never neglect my original computer science background, but this PhD program helped me to acquire more confidence as a cognitive science researcher. Computer science remains my preferred methodology to assess the advanced hypotheses the reader will find in this dissertation. Furthermore, having being part of a social robotics lab offered nice opportunities to test models of cognition on robots and it provided a duplex benefit: on the one hand it is possible to enrich cognitive research community with new insights on how human social cognition might work, on the other hand it is possible to contribute the society with a novel disruptive technology like social robots. My hope is that the reader will enjoy this dissertation as I enjoyed my journey through this research program. I wish that findings and discussions pro-

vided by this dissertation can help the reader to draw inspiration for innovative theories on human cognition, thus contributing to explain what makes us humans.

This work was not possible without the people who believed in me and gave me the opportunity to join this doctoral research degree. I came in Australia after my Master degree for a one year experience to challenge myself by learning how to live independently in a foreign country. Like most of the young people coming to Australia, I found a job in the kitchen of a busy restaurant few weeks after I landed. As a sign of fate, one morning I found the advertisement of a research scholarship for a PhD program fitting my profile. I applied knowing it would have been a competitive process and the chances to get accepted would have been low. However, after few weeks, I received the good news from Prof. Mary-Anne Williams and Dr Benjamin Johnston: I was going to stay in Australia for at least other four years. Prof. Mary-Anne Williams likes to say she 'rescued' me from that kitchen, and, although I do not regret that experience, I am glad she 'rescued' me! Therefore, special thanks go to these two great supervisors, but also to the rest of the lab who supported my ideas and contributed to my research outcomes.

In addition, I thank Prof. Giuseppe Boccignone and Dr Simone Bassis. In fact, if it were not for them, I never would have had the necessary preparation to successfully complete a PhD program.

Finally, I also thank the people who did not believe in me and in my ideas. These people are perhaps the most important since they provide the necessary determination to accomplish your goals. I am sure that if it were not for them, I would not have achieved so many successes.

Thank you all.

Author's Research Contributions

Herse, S., Vitale, J., Ebrahimian, D., Tonkin, M., Ojha, S., Sidra, S., Johnston, B., Phillips, S., Gudi, C., Clark, J., Judge, W., and Williams, M.-A. (2018). Bon appetit! robot persuasion for food recommendation. In *Proceedings of 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, March 5-8, 2018 (HRI '18)*. ACM.

Novianto, R. and Vitale, J. (2014). (Eds) Proceedings of the 1st workshop on attention for social intelligence. 6th International Conference of Social Robotics (ICSR '14).

Ojha, S., Vitale, J., and Williams, M.-A. (2017). A domain-independent approach of cognitive appraisal augmented by higher cognitive layer of ethical reasoning. In *39th Annual Meeting of the Cognitive Science Society*, pages 2833–2838.

Tonkin, M., Vitale, J., Herse, S., Williams, M.-A., Judge, W., and Wang, X. (2018). Design methodology for the ux of hri: A field study of a commercial social robot at an airport. In *Proceedings of 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, March 5-8, 2018 (HRI '18)*. ACM.

Tonkin, M., Vitale, J., Ojha, S., Clark, J., Pfeiffer, S., Judge, W., Wang, X., and Williams, M.-A. (2017a). Embodiment, privacy and social robots: May i remember you? In *Proceedings of the 9th International Conference on Social Robotics (ICSR '17), Tsukuba, Japan, November 22-24, 2017*, pages 506–515, Cham. Springer International Publishing.

Tonkin, M., Vitale, J., Ojha, S., Williams, M.-A., Fuller, P., and Judge, W. (2017b). Would you like to sample? Robot engagement in a shopping centre. In *The 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '17)*.

Vitale, J. (2014). Attention for the development of empathic bonds: From facial expressions to interoception. In *1st Workshop on Attention for Social Intelligence*. Oral presentation.

Vitale, J. (2016). Towards embodied face processing theories: a computational view. In *International School of Human-Centred Computing Bremen*. Doctoral colloquium presentation.

Vitale, J., Tonkin, M., Herse, S., Ojha, S., Clark, J., Williams, M.-A., Wang, X., and Judge, W. (2018). Be more transparent and users will like you: A robot privacy and user experience design experiment. In *Proceedings of 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, March 5-8, 2018 (HRI '18)*. ACM.

Vitale, J., Tonkin, M., Wang, X., Ohja, S., Williams, M.-A., and Judge, W. (2017a). Privacy by design in machine learning data collection: A user experience experimentation. In *Symposium on Designing the User Experience of Machine Learning Systems*. AAAI Spring Symposia 2017.

Vitale, J., Williams, M.-A., and Johnston, B. (2014a). Socially impaired robots: Human social disorders and robots' socio-emotional intelligence. In *6th International Conference on Social Robotics (ICSR '14)*, pages 350–359.

Vitale, J., Williams, M.-A., Johnston, B., and Boccignone, G. (2014b). Affective facial expression processing via simulation: A probabilistic model. *Biologically Inspired Cognitive Architectures*, 10:30–41.

Vitale, J., Williams, M.-A., and Jonhston, B. (2016). The face-space duality hypothesis: A computational model. In *38th Annual Meeting of the Cognitive Science Society*, pages 514–519.

Vitale, J., Williams, M.-A., and Jonhston, B. (2017b). Facial motor information is sufficient for identity recognition. In *39th Annual Meeting of the Cognitive Science Society*, pages 3447–3452.

Wang, X., Williams, M.-A., Gardenfors, P., Vitale, J., Abidi, S., Johnston, B., Kuipers, B., and Huang, A. (2014). Directing human attention with pointing. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '14)*, pages 174–179.

Williams, M.-A., Johnston, B., Vitale, J., Tonkin, M., Judge, W., and Pandey, A. K. (2017). (Eds) Proceedings of the 1st workshop on human-robot engagement. 26th International Joint Conference on Artificial Intelligence.

Table of Contents

List of figures

List of tables

Abstract

Face perception and cognition skills are critically needed by humans to be proficient in social cognition. Social cognition is defined as the ability to make sense of others' actions and react appropriately to them. For example, determining the identity of an interaction partner is an essential precondition to engaging socially with people. In addition, recognising facial expressions contributes to regulating human social exchanges. In fact, it assists in determining the mental state of the interaction partner and selecting the best subsequent behavioural response.

Humans show a preference for faces at a very early stage. This preference is maintained throughout their lives and it contributes to the acquisition of face recognition skills, which develop with time and experience. However, newborns have the ability to process face stimuli and imitate observed facial expressions from birth. This early imitation behaviour is a plausible way to collect sensory-motor information about the configuration of observed facial muscles. If recognising people is acquired by encountering new faces, how do humans acquire such a skill? Are there any interactions between face recognition and facial motor information processing? If so, how do these mechanisms possibly interact?

I provide answers to these research questions by looking at theories of embodied cognition. Embodied cognition research suggests that cognition extends beyond the brain to include body parts. I argue that mechanisms interacting with physical or mental aspects of the body provide sensory-motor information of the observed facial stimuli. This motor information, in turn, is sufficient for the acquisition of face identity recognition capabilities. I validate this thesis by providing mathematical models

and computational simulations describing face perception and cognition. Furthermore, I show that altering the motor representations of facial configurations leads to significant deficits in face processing capabilities. The computationally simulated dysfunctions resemble the impairments observed in clinical populations affected by social disorders, namely autism, schizophrenia and psychopathy. Hence, I argue that the bodily processes modelled in this dissertation not only have causal relationships to social cognition, but they profoundly shape it. This work is a contribution to a better computational understanding of face perception and cognition and it provides initial evidence supporting embodied social cognition theories.

Scientific knowledge is in perpetual evolution;
it finds itself changed from one day to the next.

— Jean Piaget —

1 Introduction

Does the body shape the mind? Is there something beyond the brain in developing cognition? Although these questions have a long history from classic thinkers in Philosophy of Mind, providing definitive answers remains an open matter (Gallese and Sinigaglia, 2011). The topic is gaining significant attention from more technical fields too, such as cognitive science, artificial intelligence and robotics (Anderson, 2003; Chrisley, 2003; Metta et al., 2008). This stream of research can be condensed inside the research program of "Embodied Cognition" (Shapiro, 2007).

Embodied cognition theories were introduced to overcome limitations presented by standard cognitive science. Cognitive science investigates several domains, such as perception, memory, attention, language, problem-solving, and learning. The main idea of classic cognitivism is that cognition involves algorithmic processes *upon symbolic representations* (Shapiro, 2010). Hence, cognition is a form of computation based on *symbolic manipulation* according to general rules. These are applied to the symbols' shape, and not to their meaning, thus not requiring an interpretation of what the symbols are meaning or referring to.

For example, suppose that Mary sees a cat and that she knows that cats are mammals, then Mary can conclude that she saw a mammal. It is possible to represent this scenario through propositional logic, by using simple variables to

refer to Mary, the cat and mammal. These symbolic representations do not have any sensory-motor reference in the world; they are simply abstract and amodal labels. The deduction process used to come to a conclusion that 'Mary saw a mammal' is based on the nature of these symbols, without the need of interpreting what the labels 'cat' or 'mammal' are really meaning. Indeed, changing 'cat', with another mammal like 'dog' would not make any difference to the inferential process, although dramatically changing the situational scenario of the provided example. Thus, since the aim of psychology is to give descriptions of the mental processes, and these are described as algorithms realised in the brain and manipulating symbols, psychological investigations can limit themselves to processes occurring within the brain (Shapiro, 2010).

On the contrary, embodied cognition theorists suggest this is a critical limitation of standard cognitive science: the stimuli in this perspective are impoverished of their information on how the agent can interact successfully with them and attribute them meanings (Shapiro, 2010). Cognition still involves intelligent responses to stimuli, but since these stimuli are detected by an active exploration of the environment, cognition extends beyond the brain so to include organs and body parts involved in activities used to collect and provide sensory-motor and visceral representations for such information (Shapiro, 2010). In other words, sensory information is translated into *bodily representations*, thus being enriched with *first-person phenomenological bodily sensations*. These first-person bodily sensations, which can be conscious or unconscious (De Vignemont, 2011), provides information necessary to interpret and attribute meanings to the surrounding environment and facilitate interactions with other agents.

Accordingly, in embodied cognition literature the term *'embodied'* usually refers to body parts, bodily activities, or body representations (Gallese and Sinigaglia, 2011). Hence, embodied cognition theories present a novel and exciting perspective suggesting that many features of cognition are shaped by the physical body, its mental representations and the agent's interactions with the environment.

Despite the increasing interest in theories of embodied cognition, and the corresponding remarkable amount of related works, there is still little understanding of the *mechanisms*, if any, responsible for the embodiment of mind (Gallese and Sinigaglia, 2011). Furthermore, it is still under discussion whether and to what extent embodied cognition theories can explain cognitive processes (Gallese and Sinigaglia, 2011).

In this dissertation, I propose and develop models inspired by psychological theories of *social cognition* and *face processing*. I will build these computational

tools and theories in agreement with embodied cognition research. By using these models I will provide evidence advancing explanations on how embodied mechanisms can crucially shape face perception and cognition and, in turn, social cognition capabilities. In particular, I will show that by employing embodied mechanisms realising mental representations of facial motor configurations exhibiting a body format (*e.g.* a motor map) is sufficient to develop mechanisms for discriminating among facial identities. This contribution might motivate the development of more efficient algorithms and machines inspired by embodied cognition theories. In addition to this primary outcome, I will also show that the proposed models and embodiment theories can be useful contributions to provide plausible explanations advancing the understanding of some widely studied social disorders, namely autism, schizophrenia and psychopathy. These findings can provide new insights to researchers investigating such disorders.

As I will show in the rest of this work, my aim is not to completely neglect standard cognitive science work or to deny the existence of a layer of cognition where a symbolic reasoning might indeed occur. Rather, I suggest that reviewing cognition as built on top of modal bodily representations can benefit and enrich the understanding of human mind, especially in the domain of social cognition and face processing. Therefore, I will show how standard cognitive science and embodied cognition theories can complement each other without necessarily competing.

This dissertation provides innovative and significant contributions advancing a general understanding of face perception and social cognition. While doing so, it also adds computational evidence fostering embodied cognition research. In fact, this work aim to:

(i) provide computational evidence favouring embodied cognition theories and to enrich the research program with new tools suggesting how embodied mechanisms promoting social cognition can be realised from a computational level of analysis;

(ii) provide innovative plausible hypotheses able to connect embodied cognition research with face perception and cognition studies, thus promoting a novel embodied understanding of face processing. The proposed insights and evidence will validate my hypothesis from a computational perspective;

(iii) discuss new hypotheses exploring the links between embodiment and social disorders in human clinical populations. In particular, I aim to use the computational tools proposed throughout this dissertation to assess

the plausibility of the offered argument, thus promoting an embodiment understanding of the discussed social disorders.

1.1 Motivations

Humans effortlessly deal with most social situations. We have excellent skills to attribute others' mental states by simply looking at what is happening, even with still images, especially by looking at faces. In fact, as I will discuss later in Chapter 2, the face is a primary medium for developing social cognition.

Attributing mental states to others by observing their facial expressions may seem an easy task. However, not every task we judge easy to achieve has an easy explanation in terms of mental processes (Shapiro, 2010). Therefore, having a better understanding of social cognition mechanisms underlying face-to-face social exchanges would advance cognitive science research.

In addition, curiosity can be simply enough in motivating investigations: knowing more about how our social mind works would result in a better awareness of what makes us human. However, there are at least two more practical motivations driving my research.

First, a subset of human population suffers from dysfunctions in social and emotional capabilities. These populations could have severe deficits in managing social interactions, and some of these clinical individuals exhibit dangerous antisocial behaviours. As I will show in Chapter 3, these people exhibit also impaired face processing mechanisms. Thus, I am motivated to investigate the relationships between face processing dysfunctions and embodiment impairments, so to advance new plausible hypotheses favouring a better understanding of such disorders.

Secondly, advances in artificial intelligence led to the development of new forms of machines designed to coexist and interact with humans. Social robots are an example of such machines. These intelligent social agents need to exhibit social and emotional capabilities, so to resemble natural human interactions and safely cooperate with people (Vitale et al., 2014; Williams, 2012). Therefore, investigating the mechanisms plausibly shaping social cognition development and providing computational accounts modelling them would positively enrich the field of artificial intelligence.

1.2 Contributions

The broad contribution of the current dissertation is *to advance knowledge in social cognition* by offering an innovative perspective focusing on embodied cognition theories. This outcome can significantly advance embodied cognition research program, currently still at its infancy. An interdisciplinary methodology able to foster both human and artificial intelligence studies further benefits the offered broad contribution.

Contribution 1.1 (Broad). *Providing a better understanding of social cognition's core mechanisms and proposing plausible hypotheses connecting social cognition to embodied cognition theories.*

As a primary contribution of this work, I provide a computational *understanding of the mechanisms underlying face perception and processing capabilities.* I will demonstrate, from a computational perspective, that facial configurations represented in an embodied motor space suggested to be shared among people can promote the correct development of facial expression and identity recognition.

Contribution 1.2 (Primary). *Providing a computational understanding of face perception and processing mechanisms and investigating the inter-dependencies between facial expression and identity processing.*

As a secondary contribution of this dissertation, I add computational evidence to advance hypotheses *suggesting embodied mechanisms are at the core of social cognition.* Furthermore, by linking embodied cognition research to clinical studies investigating social disorders, I will show that *deficient embodied mechanisms can plausibly explain face processing impairments* observed in these clinical populations.

Contribution 1.3 (Secondary). *Providing computational evidence supporting embodiment of social cognition and introducing novel hypotheses explaining how this embodiment can potentially affect face processing capabilities.*

While developing the main argument of this dissertation, I will provide computational implementations promoting existing theories in cognitive science literature. These computational tools will support my thesis argument and the suggested secondary contributions with sound experimental results. Thus, the suggested models and their implementations are also of independent value beyond the general argument of this dissertation. This work, therefore, offers the following additional contributions:

1. **A probabilistic account for understanding others through our body.**
   Gallese (2016) and Goldman (2013) suggest that we can understand others because we share similar bodies and brains. Thus, we can use our body to simulate in ourselves internal states very similar to the ones of our interaction partners and employ this phenomenological evidence to interpret others (see Section 2.3.4 and in particular Chapter 4 for a discussion). However, at present, no contributions are proposing explanations of the computational mechanisms plausibly underlying this process and connecting to other aspects of cognition, such as face processing (Oztop et al., 2006). To the best of my knowledge, the recent work of (Boccignone et al., 2018) is the only available computational model linking face perception to embodied cognition research. However, the model proposed by (Boccignone et al., 2018) is a more elegant and sophisticated extension of the model presented in Chapter 4 of this dissertation. In addition, the connection with other face processing mechanisms, such as identity recognition, is still missing. In this dissertation, I fill this gap by offering a computational account explaining these mechanisms, limited to the scenario of face-to-face interactions. I will provide the necessary links to embodied cognition theories and argue how embodied mechanisms are plausibly more advantageous than non-embodied and purely cognitive mechanisms.

2. **A novel computational framework in agreement with modern face processing studies.**
   Valentine (1991; 2001; 2015) proposed the face-space framework as a way to explain how we perceive, represent and discriminate invariant features of the face (*e.g.* sex, attractiveness, face shape, *etc.*). It has been suggested that faces are represented in a multidimensional space accordingly to their perceived features. Thus, for example, a component (*i.e.* an axis or dimension) of this space might account for the size of eyebrows, whereas another one the roundness of the faces. Together, these components help in discriminating among identities. Modern literature in face perception and processing strongly demonstrated the interdependent nature of invariant and dynamic features of face stimuli (Ganel and Goshen-Gottstein, 2004; Ganel et al., 2005; Kadosh et al., 2016; Pell and Richards, 2013; Rhodes et al., 2015, but see also Sections 2.2.4 and 5.1 for an additional discussion). However, at present the face-space framework is limited in explaining invariant features limiting the scope of the investigations to identity recognition

mechanisms. To the best of my knowledge, there are no computational models available able to unify invariant and dynamic features of the face in a single multi-dimensional representation facilitating both facial identity and expression recognition (Vitale et al., 2016, 2017). Therefore, in this dissertation, I propose a new hypothesis suggesting that this framework might exhibit a twofold structure able to integrate both invariant facial features (as in its original implementation) and dynamic facial features (*i.e.* facial expression and pose) in a single multidimensional space. This innovative integral explanation would help in linking embodied cognition theories to face processing studies.

1.3 Significance

Face-to-face social exchanges happen very early in life and are therefore extremely crucial for social cognition development (Grossmann, 2015, but see also Chapter 2 and, in particular, Section 2.2 for more details). Thus, enriching the overall understanding of the mental processes governing episodes of face-to-face social interactions is beneficial for human cognition research and can advance our society. For example, it has been argued that a better understanding of the early mechanisms of social cognition will be relevant for the development of better social, educational, and clinical policies (Grossmann, 2015).

This contribution is even more significant if the considered methodology makes use of an interdisciplinary approach, as proposed in this work. Indeed, by using computational models to assess the introduced hypotheses, it would be possible to inexpensively simulate many controlled experimental conditions otherwise challenging and expensive to realise with human participants. When the computational results are satisfying enough to advance the plausibility of new hypotheses, it would be possible to design similar experiments with human participants, compare the results, and provide definitive conclusions validating, rejecting or pivoting the provided hypotheses. Therefore, computational accounts of human cognition are of imperative value for advancing cognitive science research.

Furthermore, the provided computational tools can be extended to enhance their performance and used to advance artificial intelligence research. It is not always convenient to implement artificial intelligence algorithms exactly resembling human cognition (*e.g.* if birds flap their wings to fly it does not necessary mean that we need to build machines flapping their wings to fly). Nevertheless, knowing

the computational implementations of the milestone mechanisms shaping human cognition would definitely help to advance the development of novel and more sound artificial intelligence algorithms (*e.g.* if we have a better understanding of aerodynamic laws we can build machines able to fly without flapping their wings and in a better way than birds do). Hence, this work translates in methodological and practical advantages, which I will discuss in the remainder of this section.

1.3.1 Methodological Significance

Promoting integration of cognitive science findings with embodied cognition research. Instead of completely neglecting traditional cognitive science results favouring embodied cognition theories, this dissertation aims to complement the results gathered from cognitive science research with plausible modern theories of embodied cognition. I will offer crucial links between the introduced computational accounts inspired by embodied cognition theories and theoretical frameworks widely used in face processing cognitive studies. Therefore, this work is of vital importance to offer theories and tools able to foster a positive and constructive collaboration between cognitive science and embodied cognition communities.

Facilitate investigations of social cognition embodiment theories. Providing computational tools and theories advancing an embodied understanding of social cognition mechanisms would significantly push forward human cognition research. It would contribute to the development of additional theories eventually assessed by novel experimental methodologies. Indeed, computational experimentation can provide supplementary evidence to assist cognitive science researchers in designing human experiments investigating salient phenomena observed in the proposed computational simulations.

Allow falsifiability of embodied cognition theories by mean of computational frameworks and theories. Whereas present research in embodied cognition provides theories compellingly supported by neuroscience and physiological evidence, these theories find just limited support from computational accounts. In particular, the embodied mechanisms suggested to be at the core of social cognition are yet to be fully explained by a definitive computational theory (Oztop et al., 2006). Therefore, this work is a significant step towards a concrete assessment of embodied cognition theories.

1.3.2 Practical significance in human cognition research

Advancing understanding of links between mirroring mechanisms and face processing. More than two decades ago Rizzolatti et al. (1996) found in human brain a population of interesting neurons: the mirror neurons. These neurons were first discovered in various cortical areas of primates (Di Pellegrino et al., 1992; Rizzolatti and Craighero, 2004) and later in other animal species like birds (Prather et al., 2008). Since then, research in the human mirror neuron system is still active (Rizzolatti, 2005; Rizzolatti et al., 2014). All of them present the same function: they translate sensory information describing motor acts done by others into a motor format similar to that the observers themselves generate when they perform those acts (Rizzolatti and Fabbri-Destro, 2010).

Evidence from neuroscience studies compellingly advance the understanding of the neurological basis of such mechanisms. Nevertheless, currently available computational models are often exploring motor imitation processes of body acts limited to arms and hands (Oztop et al., 2006), thus neglecting one of the most salient body parts for social cognition development, namely the face.

In this work, I will show that mirroring mechanisms processing observed face stimuli are not only crucial for understanding the intentions of others (Iacoboni et al., 2005, and as suggested by the mainstream in mirror neurons research). Rather, these mechanisms are also vital for the development of face processing skills, such as facial identity discrimination. This contribution is particularly useful to demonstrate that embodiment is plausibly at the core of many social cognition processes.

Thus, this work supports the recently advanced *reuse hypothesis* (Gallese, 2005; Gallese and Caruana, 2016). According to this hypothesis, specific neuro-mechanisms that originally serve sensory-motor functions are reused in the service of social cognition (Gallese, 2005). This work will provide new insights on *how* mirroring mechanisms can be reused for promoting face processing and social cognition. These new understandings can assist the research community in addressing currently open research questions (Gallagher, 2015).

Guiding the design of new therapies for people affected by social disorders. A secondary contribution of this dissertation is to propose new plausible hypotheses suggesting impairments in embodied mechanisms to be at the core of social disorders, such as autism, schizophrenia and psychopathy.

These hypotheses would advance a significant research problem. For example, it is still not clear if face processing impairments observed in autism lead to other social cognition dysfunctions or if it is instead the other way around, with deficits in face processing resulting from earlier developing deficits in social cognition (Weigelt et al., 2012). A similar question is still open in schizophrenia research, where an understanding of the mechanisms underlying face processing impairments exhibited by this clinical population remains incomplete (Chen and Ekstrom, 2015). Finally, there is limited evidence of effective treatment of psychopathy (Salekin et al., 2010) due to the limited understanding of the causes of the exhibited dysfunctions, included dysfunctions in face processing (Dawel et al., 2012). All these social disorders have considerable social and economic costs for the individual and society (Dawel et al., 2012; Harvey, 2014; Knapp et al., 2009).

In this work, I will offer computational simulations demonstrating that dysfunctional interpretations of the sensory-motor and visceral dimensions of face stimuli leads to impairments in identity recognition, similarly to the ones observed in these clinical populations. Thus, the computational evidence provided in this work could provide helpful insights to researchers investigating the causes of social cognition dysfunctions exhibited by the considered clinical populations. This, in turn, would facilitate the development of innovative therapies.

This is not just speculation since therapies with this objective in mind are currently at hand. For example, Winkielman et al. (2015) suggest that if embodiment is part of autistic deficits, it should be possible to recover these individuals' social disorders by providing specific training enhancing their embodiment capabilities. In addition, emotional training in children with high callous-unemotional traits (precursors of adult psychopathy) administered via empathic interactions with parents has been found to be effective for reducing antisocial problematic behaviours in this population (Dadds et al., 2012). In Section 2.1.5 I report literature supporting the idea that empathy is mediated by embodied mechanisms.

1.3.3 Practical significance in computer science research

Promote collaborations between facial identity and expression recognition communities in machine learning. Machine learning community identifies facial identity and facial expression recognition as particularly distinct tasks. As a matter of fact, most of the available datasets of face stimuli focus on providing either samples assessing facial identity classification performance or

samples specifically thought for facial expression classification tasks (Jain and Li, 2011).

In this dissertation, I will show that the two tasks complement each other. The insights provided by this work can significantly advance machine learning community by suggesting new algorithms able to integrate these two capabilities, for example in deep learning computational models.

Deep learning is a promising methodology to train deep artificial neural networks, namely computational models inspired by human neurons and designed to learn complex non-linear functions mapping inputs to desired outputs. These models allow representations of data with multiple levels of abstraction able to compellingly classify new observations (LeCun et al., 2015) and models linking to embodied cognition research are currently at hand (Boccignone et al., 2018). Although the proposed models excel in many recognition capabilities, sometimes even above human performance (Taigman et al., 2014), the underlying structure and representations resulting from training these networks are difficult to interpret (Sussillo and Barak, 2013). This approach can achieve very accurate results, though still not providing any useful information about how the brain works (see Martinez, 2017, for a discussion on this topic). Promoting collaborations between facial expression and identity recognition researchers could lead to find significant interactions between representations of dynamic (*e.g.* facial expression) and invariant (*e.g.* identity) features of the face in these models. These new discoveries would likely lead to the development of models exhibiting enhanced performance.

Advancing artificial social learning. Being able to understand the social world is a necessary skill to learn from social interactions (Grossmann, 2015). Thus, a better understanding of social intelligence, supplemented by computational models, can advance artificial intelligence by developing novel algorithm promoting artificial social learning. For example, the embodied mechanisms discussed and modelled in this dissertation (see Chapter 4) can be extended and refined to allow artificial social agents to understand others' actions by using the system normally employed to perform such actions. This new cognitive architecture will provide significant advantages. For instance, it will reduce computational complexity in action representation (*i.e.* a single representation for both execution and interpretation of actions), and it will offer better extensibility of the system, namely learning on-line a new motor skill and the goal/intention associated with

it would result in a new representation for its interpretation when observed again in others.

Promote long-term human-machine social interactions. In order to develop artificial social agents able to sustain long-term interactions with humans, it is necessary to proficiently understand social and emotional intelligence (Dautenhahn, 2002). These forms of intelligence are especially necessary for social agents interacting with elderly people or hospitalised individuals, often in need of a more significant level of empathy. Thus, the insights and computational models provided by this work are the first steps towards a computational understanding of social and emotional intelligence, providing the basis for sustainable and more authentic long-term human-machine social interactions.

1.4 Thesis Argument

In this dissertation I demonstrate the following thesis:

Thesis *Sensory-motor information of face stimuli is sufficient to facilitate the acquisition of face recognition capabilities because this information is available early in life via embodiment mechanisms able to map sensory information of novel face stimuli, encountered throughout social exchanges, onto corresponding motor representations.*

In other words, the embodied mechanisms discussed and modelled in this dissertation translate observations of novel face stimuli into sensory-motor information. They do so while preserving critical aspects of the body. These embodied representations facilitate the interpretation of facial motor acts observed in others, such as facial expressions of emotion. I will demonstrate, from a theoretical and computational standpoint, that this sensory-motor information is sufficient to develop mechanisms able to classify facial identities. Therefore, I will suggest that embodied mechanisms plausibly underlie social cognition development.

In fact, face identity discrimination is an important ability for social cognition; it provides paramount information about the identity of the interaction partners, which can bias and regulate social exchanges. Therefore, facial motor information is sufficient to shaping face processing capabilities.

I will argue that face processing embodiment is a parsimonious way human brain may have evolved to allow the development of face discrimination ability from birth. Furthermore, I will propose that face processing deficits in clinical

populations may be due to dysfunctional embodied mechanisms encoding facial motor dynamics. Therefore, validating this thesis and advancing the proposed secondary contributions will add computational evidence supporting embodied cognition research.

In order to avoid misunderstandings, I want to make clear that the aim of this work is not to provide a definitive model of how facial expressions and identity recognition are *physically realised* in human brain. Rather, this work wants to provide a new plausible and parsimonious description of the *computational mechanisms* that may potentially govern an episode of face-to-face social interaction.

1.5 Proposed Methodology

The methodology advancing the proposed thesis argument will use:

1. A comprehensive literature survey of three widely investigated social disorders, namely autism, schizophrenia and psychopathy (see Chapter 3). Through this survey, I will be able to identify critical connections between social cognition, face processing and embodied mechanisms. I will use the proposed insights to advance hypotheses that will supplement and lead to the validation of my thesis argument;

2. A probabilistic account of embodied mechanisms (see Chapter 4). This account will be necessary to demonstrate that it is possible to computationally model the functions of embodied mechanisms suggested by embodied cognition literature. Furthermore, I will show that these mechanisms can be implemented to realise sensory-motor information of face stimuli, thus linking embodied cognition research to face processing studies;

3. A computational model of face processing for face expression and identity recognition (see Chapter 5). This computational framework will be used to demonstrate that dynamic (*e.g.* facial motor configurations) and invariant (*e.g.* identity) features of the face are strictly interdependent and they both contribute to the development of face processing capabilities. Therefore, I will connect facial identity recognition to facial expression classification and offer a link between this model and the previously mentioned computational account of face stimuli embodiment.

I will then use all the suggested hypotheses, theories, models and computational implementations to provide quantitative evidence in support of my thesis argument (see Chapter 6). In addition to this direct computational evidence, in Chapter 6, I will evaluate my thesis on other aspects, in agreement with a common methodology of theory validation suggested by philosophy of science research (Barsalou, 1999):

Innovation: Is the new hypothesis provocative enough to disrupt current perspectives, thus pushing face processing research forward?

Falsifiability: Is it possible to conduct an experiment which could falsify my argument that sensory-motor information is sufficient for the acquisition of face recognition skills?

Parsimony: Is the hypothesis more elegant and less complex than previous theories in face processing?

Integrability: Can the hypothesis be integrated with findings supporting other pre-existing theories in face processing?

Plausibility: Is the hypothesis free from conceptual problems?

I will answer positively to all these questions, thus providing strong conclusions in support of my thesis argument.

1.6 Acknowledged Limitations

A doctoral research project is constrained by very limited time and resources, thus imposing limitations. Social cognition, embodied cognition theories and face processing involve a wide variety of literature, require an extensive expertise and cover many domains of analysis. Given my backgrounds, and the ones of my research team, my choice oriented on computer science as a reasonable domain of analysis for my hypotheses. Despite this choice, I drew inspiration and foundations from psychological, neurological and clinical studies of social cognition.

Thus, one main limitation of this dissertation is the absence of direct evidence coming from experiments involving humans. The provided evidence is instead based on data resulting from computational simulations of psychological theories. This data is sufficient to justify my thesis argument from a computational perspective and to provide a plausible novel understanding of social cognition. However, in this dissertation, I do not aim to provide a definitive answer on how

social cognition is realised in the human brain. Rather, my objective is to provide evidence suggesting plausible and innovative hypotheses able to foster embodied cognition research program.

Another limitation is the choice of delimiting social cognition investigation to face-to-face interactions. There are obviously other forms of interactions vital for social cognition development. However, as I will discuss in Chapter 2, the human face is the most communicative channel available among humans, one of the earlier mechanisms to acquire knowledge, and one of the more studied topics in both human and computer science research.

These limitations, although potentially circumscribing the impacts that this work might have on the scientific community, are not negatively affecting the final argument of my dissertation. Indeed, this work presents a new perspective on social cognition embodiment that future works in other domains can test further.

1.7 Dissertation Overview

In Chapter 2, I will review the available literature on social cognition, face perception and embodied cognition theories to provide context framing the present research. In doing so, I will discuss some of the significant open challenges motivating this research.

In Chapter 3, I will provide a review of works in clinical psychology studying dysfunctions of human social disorders. This survey will promote the necessary connections between social cognition, embodied cognition and face processing. In addition, I will conclude Chapter 3 with hypotheses supplementing my main thesis argument.

Chapters 4 and 5 provide the computational tools necessary to validate my thesis argument. In Chapter 4, I will provide a computational model of embodied mechanisms underlying an episode of face-to-face interaction. I will show that this model can realise sensory-motor information of observed face stimuli having a bodily format and promoting the classification of facial motor configurations. In Chapter 5, I will introduce a new hypothesis on face processing suggesting that facial expression and identity recognition significantly interact. Hence, I will provide a computational model able to validate this hypothesis and advancing the main thesis argument.

Chapter 6 will serve to validate my thesis argument. By discussing the previous evidence and extending the computational tools developed earlier I will

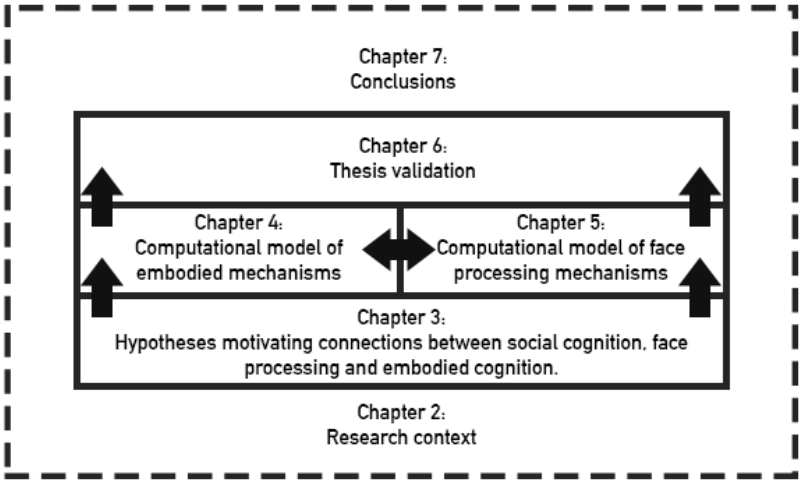


Figure 1.1 Visual representation of the core contributions of the present dissertation. Chapter 2 provides a context to frame the rest of the dissertation chapters. Chapter 3 provides the motivations for the present research, and it establishes connections between social cognition, embodied cognition theories and face processing, which will be the foundations for the following two chapters. Chapter 4 provides a computational model of embodied mechanisms employed during face-to-face social interactions, which connects to the face processing model proposed in Chapter 5. These two chapters together will validate the thesis in Chapter 6. Finally, the conclusions in Chapter 7 will provide a summary of the findings reconnecting with all the previous chapters.

present additional data in support of my argument and the proposed secondary contributions.

I will conclude in Chapter 7 with a discussion on the present contribution and their significance to advance the research gaps identified in Chapter 2. I will further reinforce my thesis argument by assessing its innovation, falsifiability, parsimony, integrability and plausibility. Finally, I will define a working agenda for future works. Figure 1.1 visually summarise the contents and structure of the remainder chapters of this dissertation.

Chapter Bibliography

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial intelligence*, 149(1):91–130.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4):577–660.

Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., and Lanzarotti, R. (2018). Deep construction of an affective latent space via multimodal enactment. *IEEE Transactions on Cognitive and Developmental Systems*.

Chen, Y. and Ekstrom, T. (2015). Visual and associated affective processing of face information in schizophrenia: A selective review. *Current Psychiatry Reviews*, 11(4):266–272.

Chrisley, R. (2003). Embodied artificial intelligence. *Artificial Intelligence*, 149(1):131–150.

Dadds, M. R., Cauchi, A. J., Wimalaweera, S., Hawes, D. J., and Brennan, J. (2012). Outcomes, moderators, and mediators of empathic-emotion recognition training for complex conduct problems in childhood. *Psychiatry Research*, 199(3):201–207.

Dautenhahn, K. (2002). *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, volume 3. Springer Science & Business Media.

Dawel, A., O'Kearney, R., McKone, E., and Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience & Biobehavioral Reviews*, 36(10):2288–2304.

De Vignemont, F. (2011). Bodily awareness. *Stanford Encyclopedia of Philosophy*.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, 91(1):176–180.

Gallagher, S. (2015). Reuse and body-formatted representations in simulation theory. *Cognitive Systems Research*, 34:35–43.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, 4(1):23–48.

Gallese, V. (2016). Finding the body in the brain. From simulation theory to embodied simulation. In McLaughlin, B. P. and Kornblith, H., editors, *Goldman and His Critics*, pages 299–314. John Wiley & Sons.

Gallese, V. and Caruana, F. (2016). Embodied simulation: Beyond the expression/experience dualism of emotions. *Trends in Cognitive Sciences*.

Gallese, V. and Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, 15(11):512–519.

Ganel, T. and Goshen-Gottstein, Y. (2004). Effects of familiarity on the perceptual integrality of the identity and expression of faces: The parallel-route hypothesis revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3):583.

Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., and Goodale, M. A. (2005). The involvement of the "fusiform face area" in processing facial expression. *Neuropsychologia*, 43(11):1645–1654.

Goldman, A. I. (2013). The bodily formats approach to embodied cognition. *Current Controversies in Philosophy of Mind*, page 91.

Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological Bulletin*, 141(6):1266.

Harvey, P. D. (2014). Clinical and cost implications of treating schizophrenia: Safety, efficacy, relapse prevention, and patient outcomes. *Journal of Clinical Psychiatry*, 75.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, 3(3):530–535.

Jain, A. K. and Li, S. Z. (2011). *Handbook of Face Recognition*. Springer.

Kadosh, K. C., Luo, Q., de Burca, C., Sokunbi, M. O., Feng, J., Linden, D. E., and Lau, J. Y. (2016). Using real-time fMRI to influence effective connectivity in the developing emotion regulation network. *NeuroImage*, 125:616–626.

Knapp, M., Romeo, R., and Beecham, J. (2009). Economic cost of autism in the uk. *Autism*, 13(3):317–336.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Martinez, A. M. (2017). Computational Models of Face Perception. *Current Directions in Psychological Science*, 26(3):263–269.

Metta, G., Sandini, G., Vernon, D., Natale, L., and Nori, F. (2008). The iCub humanoid robot: An open platform for research in embodied cognition. In *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, pages 50–56. ACM.

Oztop, E., Kawato, M., and Arbib, M. (2006). Mirror neurons and imitation: A computationally guided review. *Neural Networks*, 19(3):254–271.

Pell, P. J. and Richards, A. (2013). Overlapping facial expression representations are identity-dependent. *Vision Research*, 79:1–7.

Prather, J. F., Peters, S., Nowicki, S., and Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, 451(7176):305–310.

Rhodes, G., Pond, S., Burton, N., Kloth, N., Jeffery, L., Bell, J., Ewing, L., Calder, A. J., and Palermo, R. (2015). How distinct is the coding of face identity and expression? Evidence for some common dimensions in face space. *Cognition*, 142:123–137.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and embryology*, 210(5-6):419–421.

Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., and Rozzi, S. (2014). Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiological Reviews*, 94(2):655–706.

Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192.

Rizzolatti, G. and Fabbri-Destro, M. (2010). Mirror neurons: From discovery to autism. *Experimental Brain Research*, 200(3-4):223–237.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., and Fazio, F. (1996). Localization of grasp representations in humans by PET:1. Observation versus execution. *Experimental Brain Research*, 111(2):246–252.

Salekin, R. T., Worley, C., and Grimes, R. D. (2010). Treatment of psychopathy: A review and brief introduction to the mental model approach for psychopathy. *Behavioral Sciences & the Law*, 28(2):235–266.

Shapiro, L. (2007). The embodied cognition research programme. *Philosophy compass*, 2(2):338–346.

Shapiro, L. (2010). *Embodied Cognition*. Routledge.

Sussillo, D. and Barak, O. (2013). Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Computation*, 25(3):626–649.

Taigman, Y., Yang, M., Ranzato, M., and 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*, pages 1701–1708.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43(2):161–204.

Valentine, T. (2001). Face-space models of face recognition. *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*, pages 83–113.

Valentine, T., Lewis, M. B., and Hills, P. J. (2015). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, pages 1–24.

Vitale, J., Williams, M.-A., and Johnston, B. (2014). Socially impaired robots: Human social disorders and robots' socio-emotional intelligence. In *6th International Conference on Social Robotics*, pages 350–359.

Vitale, J., Williams, M.-A., and Jonhston, B. (2016). The face-space duality hypothesis: A computational model. In *38th Annual Meeting of the Cognitive Science Society*, pages 514–519.

Vitale, J., Williams, M.-A., and Jonhston, B. (2017). Facial motor information is sufficient for identity recognition development. In *39th Annual Meeting of the Cognitive Science Society (Under review)*.

Weigelt, S., Koldewyn, K., and Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience & Biobehavioral Reviews*, 36(3):1060–1084.

Williams, M.-A. (2012). Robot social intelligence. In *Social Robotics*, pages 45–55. Springer.

Winkielman, P., Niedenthal, P., Wielgosz, J., Eelen, J., and Kavanagh, L. C. (2015). Embodiment of cognition and emotion. In Wagner, D. and Heatherton, T., editors, *APA Handbook of Personality and Social Psychology, Volume 1: Attitudes and Social Cognition*, volume 1, pages 151–175. American Psychological Association Washington, DC.

Do not become a mere recorder of facts, but
try to penetrate the mystery of their origin.

— Ivan Pavlov —

2

Research Context

As a general target of this dissertation, I will argue *if* and *how* social cognition is plausibly embodied. To gather the necessary computational evidence, I will focus on the more narrow research field of face processing as a valuable domain where testing my hypotheses. This chapter will provide a glimpse of *social cognition*, *face processing*, and *embodied cognition theories* from human studies.

I will start by reviewing human *social cognition* literature (Section 2.1). First, I will provide the definition and the significance of social cognition by introducing the necessary backgrounds motivating the broad topic of this dissertation (Section 2.1.1). I will then review theories suggesting how humans achieve social cognition in agreement with embodied cognition standpoint (Section 2.1.2). In Sections 2.1.3, 2.1.4 and 2.1.5, I will introduce the mechanisms possibly underlying social cognition. This information would be necessary to motivate the computational model proposed in Chapter 4. Finally, I will provide additional backgrounds on social cognition development (Section 2.1.6). In this final section, I will suggest that the face is a vital medium for acquiring social cognition capabilities, therefore connecting to Section 2.2 where I will discuss the role of the *face* in social cognition.

Section 2.2 will identify the main capabilities of face processing mechanisms able to promote social cognition, namely face detection (Section 2.2.1), facial identity processing (Section 2.2.2) and facial expression processing (Section 2.2.3).

In addition, I will briefly review the interactions between facial expression and identity processing (Section 2.2.4). The insights of this review will motivate the development of the computational model proposed in Chapter 5.

In Section 2.3, I will provide the necessary background on *embodied cognition theories*. This information would be required to enrich the previously introduced topics with embodied cognition understandings. Finally, in Section 2.4, I will identify the most significant research gaps that the contributions of the present work will advance.

In this dissertation, I have deliberately chosen to review literature from different disciplines to best cover the complexity of the proposed topics. Although this approach has the advantage of providing a much broader background, it has a drawback too: the same terms might be used to explain different phenomena or mechanisms, and different terms might be used to explain the same phenomena or similar mechanisms (Gentsch et al., 2016). Thus, I will try to put an extra effort in maintaining the coherence among the terms introduced in this review and further used throughout this dissertation, by clearly providing their definitions and by listing their interchangeable terms whenever necessary.

2.1 Social Cognition

In this section, before reviewing social cognition literature, I suggest first to define the meaning of the term *cognition* in cognitive science research. Neisser (1967) gives one of the earliest definitions of cognition in his textbook on cognitive psychology first published in 1967. According to him, cognition is *"those processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used"* (Neisser, 1967, 2014, page 4). Cognition is not limited to the process of external stimuli but also extends to the process of mental images and hallucinations (Neisser, 1967, 2014). Cognition arises from the interaction of three levels (Gentsch et al., 2016):

The perceptual level, which is concerned with sensing external information (*i.e.* exteroception), internal sensations of feelings, such as hunger and pain, and organs movements (*i.e.* interoception) and sensations of own muscular-skeletal state (*i.e.* proprioception);

The cognitive level, which is concerned with processing the available information and giving rise to appropriate mental states;

The motor or behavioural level, which is concerned with executing actions based on the active mental states and motivational social factors.

The study of the interplay between perception, cognition and action has a long tradition and interest in cognitive science research (Gentsch et al., 2016).

In the following sections, I will provide the definition of social cognition and its significance in promoting society (Section 2.1.1). Then, I will discuss available explanations of social cognition functioning in agreement with embodied cognition standpoint (Section 2.1.2). Finally, I will offer the mechanisms plausibly underlying social cognition (Sections 2.1.3, 2.1.4 and 2.1.5) and its developmental process in humans (Section 2.1.6).

2.1.1 Definition and Significance of Social Cognition

Social cognition is used as an umbrella term referring to a wide range of mental processes required to perceive, process, and interpret *social information* (Brothers, 1990). Social information includes identity, the direction of movement, the category of posture, facial expressions, the quality of vocalisation, the knowledge of which other individuals are present and what their mutual relations are (Brothers, 2002).

Definition 2.1 Social cognition *is the set of mental processes required to make sense of others' actions and react appropriately to them (Langton et al., 2000).*

Social cognition requires the interaction of many mental processes extending beyond the specific social domain. In fact, to process social information it is necessary to use mechanisms having a specific social function, such as face recognition or emotional cues extraction from vocalisations, but also other processes having functions covering a more broad range of domains, such as memory and attention (Kennedy and Adolphs, 2012).

Research in social cognition attempts to understand and explain how interactions with other individuals can influence thoughts, feelings, and behaviour of people (Grossmann, 2015). Literature in psychology suggests that the precondition to correctly master social cognition is to understand that others have *different minds* (Premack and Woodruff, 1978), namely different beliefs, desires, intentions and knowledge. Indeed, we need to understand that others have diverse backgrounds and knowledge to proficiently make sense of their behaviour and promote acceptable social interactions (Brothers, 2002). This ability is known as Theory of Mind (Premack and Woodruff, 1978):

Definition 2.2 Theory of Mind *is the ability to understand that others have different minds (Premack and Woodruff, 1978).*

Theory of Mind is not available from birth; it develops gradually during the first 3-5 years of life together with other social cognition abilities (Gallese et al., 2009), thus requiring other cognitive mechanisms.

To live and survive in our complex society, it is fundamental to proficiently manage social cognition mechanisms (Brothers, 2002). In fact, social cognition enables people to cope with other individuals' actions, by recognising them, understanding them, and reacting appropriately to them (Gallese, 2001). In addition, social cognition provides mechanisms to predict future social events so to anticipate others' upcoming actions and consequently adjust one's behaviour (Gallese et al., 2009).

In the following section, I will provide some of the available theories in literature advancing an explanation on how social cognition can be realised by humans. These theories would inspire the computational account proposed in Chapter 4.

2.1.2 Theories of Social Cognition

Previously I suggested that social cognition allows people to understand others' mental states by observing their behaviour and processing observable social information. Although it is possible to provide a precise definition of social cognition and its significance in promoting society, it is not so easy to explain *how* social cognition is used to infer others' mental states. This question has long tradition in folk psychology (Davies and Stone, 1995), philosophy of mind (Braddon-Mitchell and Jackson, 2006), and it has contributed to novel theories in neuroscience (Gallese and Goldman, 1998).

Gallese (2001) proposed the *'shared manifold hypothesis'* as a way to advance an explanation of social cognition functioning. The shared manifold is a conceptual tool used to explain how people can understand others by mean of shared mechanisms resonating with others' behaviour. The author suggests that at the basis of social cognition in humans there are shared neuro-mechanisms capable of *mirroring* other minds, thus providing us experiential insights of others (Gallese et al., 2004). He proposes that the brain is capable of translating sensory information of others' behaviour into first-person representations of the very same observed behaviour. Importantly, this mirroring capability is not limited to vision, but it includes other senses like hearing (Rizzolatti and Craighero, 2004). Indeed, Ricciardi et al. (2009) demonstrated that the mirror system can develop even in blind people.

Hence, the mirror system provides supramodal[1] sensory representations of actions that can be used by people to proficiently interact with each others (Ricciardi et al., 2009).

Therefore, at the core of social cognition are suggested mechanisms able to translate sensory information into phenomenological evidence promoting an understanding of others by *feeling like them* (Goldman, 1993). This idea is not novel in philosophy of mind, where it finds Simulation Theory as a suitable theoretical account (Gallese and Goldman, 1998). Traditional literature in philosophy of mind offers two main approaches leading to Theory of Mind and enabling social cognition: Theory-Theory and Simulation Theory (Goldman, 1992).

Definition 2.3 Theory-Theory *account proposes that the mind-reader*[2] *deploys a naïve psychological theory to infer mental states in others by observing the behaviour of the interaction partners and by considering the current environment and context where the interaction is happening (Goldman and Sripada, 2005).*

Definition 2.4 Simulation Theory *suggests that the mind-reader selects a mental state to attribute to others after reproducing or enacting within himself the very state in question, or a relevantly similar one (Goldman and Sripada, 2005).*

By using a simulation-based approach, people do not need a set of rules or theories to infer others' mental states; instead, they use their own body as a model of others (see Chapter 4 and in particular Section 4.1 for in-depth discussion). Thus, to deploy Simulation Theory it is necessary that very same *phenomenological mechanisms* are shared among individuals, as suggested by Gallese (2001). As I will show in Chapter 4, providing a computational theory of these shared mechanisms is not a trivial task.

Both Simulation Theory and Theory-Theory suggest valid approaches describing how people can master social cognition capabilities. However, the two theories have a significant difference: Simulation Theory requires an *embodiment* and experiential insights in order to *"put the mind-reader in other's shoes"* (Goldman and Sripada, 2005), whereas Theory-Theory does not. Simulation Theory contributes in *resonating* and *empathising* with others by means of shared brains and bodies, whereas Theory-Theory enacts a 'cold' cognitive appraisal process requiring to define and manage a set of complex propositional rules.

[1]This term refers to a representation that is not specific of a single sensory input, but it efficiently integrates several of them in a single more abstract multimodal representation.

[2]In this work, I will use the term 'mind-reader' to denote the subject trying to infer the mental state of the interaction partner

Therefore, among the possible hypotheses attempting an explanation of social cognition development, the shared manifold hypothesis, together with Simulation Theory, are the most relevant ones for this work, since supplemented by compelling neuroscientific evidence (see Section 2.1.4) and fits really well from an embodied cognition standpoint. Thus, these hypotheses will be the foundations of my computational model presented in Chapter 4.

In Sections 2.1.3, 2.1.4 and 2.1.5, I will provide a set of mechanisms, and their definitions, plausibly underlying social cognition and in agreement with the theories suggested in this section.

2.1.3 When Emotions Come into Play: Emotional Contagion

The definitions in Section 2.1.1 suggested a strictly cognitive nature of social cognition, namely capabilities employed to *reason* about others. Nevertheless, in Section 2.1.2 I suggested that social cognition can be achieved by mean of simulation-based processes realising phenomenological states in people. Indeed, it has been suggested that social cognition is not limited to reasoning, but it also subsumes phenomenological and emotional capabilities (Salovey and Mayer, 1989). These capabilities are crucial to shaping *emotional intelligence*.

Definition 2.5 Emotional Intelligence *is the ability to perceive, manage, and reason about emotions, within oneself and others (Ermer et al., 2012).*

This form of intelligence contributes to the accurate appraisal, expression and regulation of emotion in oneself and in others, and the use of feelings to motivate, plan and achieve in one's life (Salovey and Mayer, 1989).

One of the core mechanisms of emotional intelligence is *Emotional contagion* (Coplan and Goldie, 2011).

Definition 2.6 Emotional contagion *is "the tendency to rapidly mimic and synchronise facial expressions, vocalisations, postures, and movements with those of another person and, consequently converge emotionally with others" (Coplan and Goldie, 2011, page 68).*

To converge to another's emotional state, emotional contagion requires the reproduction of a behaviour similar to the one observed in others. In the literature we find two distinct terms defining a reproduced behaviour: *imitation* (or true imitation) and *mimicry* (or automatic imitation) (Vivanti and Hamilton, 2014).

Definition 2.7 Imitation *involves a conscious mechanism*[3] *having the aim of copying both the means and the goals of an observed behaviour with high fidelity (Vivanti and Hamilton, 2014).*

Definition 2.8 Mimicry *occurs when the observer automatically and unconsciously matches an observed behaviour (Vivanti and Hamilton, 2014).*

As I will further discuss in Section 2.1.6, emotional contagion is particularly important for learning through social interactions, since it allows infants to associate perceptual (*e.g.* a visual image of an object) or symbolic (*e.g.* a word) information to an experienced bodily state communicated by an emotional reaction of the caregiver (De Vignemont and Singer, 2006).

Although emotional contagion is often suggested to occur automatically and outside of conscious awareness by means of mimicry mechanisms (Coplan and Goldie, 2011), in this dissertation I will use the term emotional contagion whenever the observer's body *overtly* replicate others' body activity (motor and/or visceral) to converge to a shared emotional state, independently of how the body activity is replicated. In other words, I will use this term either if the body activity is mediated by conscious mechanisms (*i.e.* imitation) or if the body activity is mediated by unconscious ones (*i.e.* mimicry).

2.1.4 Mirroring as Core Mechanism of Social Cognition

In the previous section, I suggested that emotional contagion is mediated by mimicry mechanisms reproducing an observed behaviour. However, to overtly replicate the observed behaviour it is necessary to have mental representations of the motor potentials underlying the realisation of that behaviour (Nummenmaa et al., 2008). Are there mechanisms in the brain having such function?

While investigating properties of the prefrontal cortex of monkeys, Di Pellegrino et al. (1992) found a population of neurons with interesting properties. The peculiar feature of these neurons was that they discharged both when the monkey performed a certain motor act and when it observed another individual (either another monkey or a human) performing that or a similar motor act. Given the property of these neurons to mirror observed actions in the observer brain, they were named *mirror neurons*. Importantly, the sensory information is not limited to visual information. For example, some of these neural cells also spike when

[3]In this context, the term 'conscious mechanism' refers to any cognitive process reaching subject's awareness.

the subject hear a sound correlated to a particular action and, similarly, when this action is performed by the subject (Rizzolatti and Craighero, 2004). Neurons with similar functions were also found in human brain (Rizzolatti et al., 1996), in various cortical areas of other primates (Rizzolatti and Craighero, 2004), and in birds (Prather et al., 2008). Therefore, the mirroring function of these neurons is shared among all species.

Definition 2.9 Mirroring *is a process mapping the sensory representation of actions, emotions or sensations of others' onto the observer's own motor, viscero-motor or somatosensory representation of the observed actions, emotions or sensations (Gallese and Sinigaglia, 2011).*

In this dissertation I will use the terms mirroring, mirroring mechanisms or inner imitation interchangeably to define a process coupling perceptual observations to *covert* mental representations of the bodily activity associated with such observation.

Gallese (2001) suggests that these neurons are the neural substrate of the proposed shared manifold hypothesis. In other words, similar brain structures, shared among people, allows individuals to activate similar neural patterns and, therefore, share similar phenomenological states. This mechanism guarantee an 'access' to others minds, thus promoting social cognition.

Mirroring can be seen as precursors of emotional contagion. In fact, to overtly replicate an observed behaviour, mental representation of such motor act must be activated (Gallese, 2016). Therefore, mirroring mechanisms plausibly situate at the core of social cognition. Indeed, mirror activity is often correlated with mind-reading abilities, thus indicating its underlying social dimension (Goldman and de Vignemont, 2009). Furthermore, a study of Enticott et al. (2008) showed that a marker of mirror neuron activity in the premotor cortex correlates with performance on a measure of social cognition. In particular, the accuracy in recognising facial expression of emotion from static images is associated with the motor-evoked potential amplitude during action observation.

So far, the mechanisms proposed to underlie social cognition (mirroring and emotional contagion) are sufficient to elicit in the observer a phenomenological state isomorphic to the one of the interaction partner. However, they are not enough to reach the *attribution stage*, in which the inferred mental state is ascribed to others. In the following section, I will review literature suggesting empathy to be a valid ability for the attribution of emotional states to others (De Vignemont and Singer, 2006). As I will demonstrate from the literature review, mirroring

mechanisms are plausibly situated at the foundations of empathy (Goldman, 1993).

2.1.5 Empathy: Understanding Others through Inner Imitation

In Section 2.1.3, I demonstrated that social cognition is not only a process involving cognitive mechanisms, but it also includes emotional mechanisms. These emotional mechanisms facilitate shared phenomenological states in people, thus promoting mutual understanding. In Section 2.1.4, I reviewed literature suggesting that the shared phenomenological states can be realised via mirroring mechanisms since their function is to map perceptual information of motor acts onto first-person bodily representations of these motor acts. Therefore, with these motor representations available, the subject can activate a similar body state via mimicry mechanisms and consequently converge to an emotional state akin to the one experienced by the interaction partner. However, how can this emotional state be attributed to others? The literature suggests empathy be the answer to this question.

Definition 2.10 Empathy *is broadly described as an understanding of another person's feelings (Coplan and Goldie, 2011).*
Empathy requires that the subject (De Vignemont and Singer, 2006, page 435):

(i) is in an affective state;

(ii) this state is isomorphic to another person's affective state;

(iii) this state is elicited by the observation or imagination of another person's affective state;

(iv) the subject knows that the other person is the source of one's own affective state.

As suggested in Sections 2.1.3 and 2.1.4, points (i), (ii) and (iii) of Definition 2.10 are promoted by mirroring and emotional contagion mechanisms. Indeed, Lipps (1935), in his first definition of empathy, suggested that its elicitation happens via a process of *inner imitation*. Replicating the observed behaviour (either covertly or overtly) can make us *feeling like him/her* (Iacoboni, 2009).

On the other hand, point (iv) of Definition 2.10 distinguishes an empathic process from a simple emotional contagion episode. When the emotional state

elicited by an emotional contagion process is consciously *attributed to others*, the process is called *emotional empathy*, *mirror empathy* or *low-level empathy* (Coplan and Goldie, 2011):

Definition 2.11 Emotional empathy *is the process of consciously attributing an emotional state currently experienced by the subject to his interaction partner (Coplan and Goldie, 2011).*

Since the emotional state to attribute to others is determined by a *feel* experienced within the subject via mirroring and emotional contagion, emotional empathy does not require to cognitively assess a given situation in order to identify and attribute such emotional state (Coplan and Goldie, 2011).

Nevertheless, the experienced feeling can be enhanced, suppressed or updated after assessing the current event (De Vignemont and Singer, 2006). For example, consider Mary observing Luke smiling. Mary may activate mirroring mechanisms underlying bodily representations of happiness and attribute this state to Luke. However, if Luke is smiling, embarrassed, after having almost tripped in front of a crowd of individuals, the feeling mirrored by Mary can be replaced by embarrassment, even before reaching awareness (Bargh, 1994; Singer and Lamm, 2009), thus promoting correct empathy. This cognitive assessment can be realised through a cognitive appraisal process.

Definition 2.12 Cognitive appraisal *is the process of consciously assessing the current situation in reference to one's own personal well-being. (Scherer and Ellgring, 2007).*

The appraisal of the events can be done in terms of appraisal variables as suggested by Scherer (1999), so to facilitate the generalisation over similar situations. This appraisal process selects the type of emotional activation of the subject given specific factors of the event, such as novelty, pleasantness and significance to own goals. Cognitive appraisal can be self-directed (first-person perspective) or world-directed (third-person perspective) (Lambie and Marcel, 2002). However, during an empathic episode, this process is limited in assessing the situation from another's point-of-view (as per requirement (iv) of Definition 2.10 introduced on page 31).

In addition, cognitive appraisal can be done without necessarily experiencing an emotional state. For example, if a subject sees a person crying he can still

attribute to him an emotional state of sadness[4], without necessarily feeling sad (*i.e.* without becoming emotionally affected via emotional contagion). In this case, we are speaking of *cognitive empathy*, *reconstructive empathy* or *high-level empathy* (Coplan and Goldie, 2011).

Definition 2.13 Cognitive empathy *is the process of consciously attributing an emotional state to others by using a third-person perspective to cognitively assess a given situation, without physically experiencing that emotional state (De Vignemont and Singer, 2006).*

However, cognitive empathy does not meet requirements (i) and (ii) of the Definition 2.10 of empathy introduced on page 31. Therefore, real empathy is limited to the emotional empathy process (De Vignemont and Singer, 2006).

In summary, in this section, I reviewed literature suggesting that the attribution of an emotional state to others can be realised through empathy. Mirroring mechanisms assist empathy by providing the necessary representations of the bodily state observed in others. To this process, it can or cannot follow the experience of an emotional state isomorphic to the one of the interaction partners. When the subject experiences an isomorphic emotional state we are speaking of emotional empathy. The underlying emotional state is realised via emotional contagion mechanisms and potentially updated or replaced with another one by a cognitive appraisal process. When the subject does not experience an isomorphic emotional state we are speaking of cognitive empathy. In this case there is no underlying emotional state, but only a cognitive assessment of the situation via a cognitive appraisal process.

Given the discussed literature, I argue that mirroring can be placed at the core of social cognition, and for this reason, in my dissertation, I will focus on modelling these mechanisms and their potential interaction with other cognitive processes. Figure 2.1 presents the taxonomy of the processes underlying social cognition discussed so far. This figure serves as a summary of the social cognition mechanisms, limited to the ones discussed in the present chapter.

Overall, in Section 2.1.2, I proposed the shared manifold hypothesis and Simulation Theory as valid accounts to explain how people can make sense of others and attribute them mental states. I then explored the potential mechanisms underlying such processes in Sections 2.1.3, 2.1.4 and 2.1.5, proposing emotional contagion,

[4]This knowledge of 'crying', and similarly other behaviours, can be still provided by mirroring mechanisms, likely in a non-propositional form (*i.e.* acquired 'by doing') (see Gallese, 2005, for a discussion)

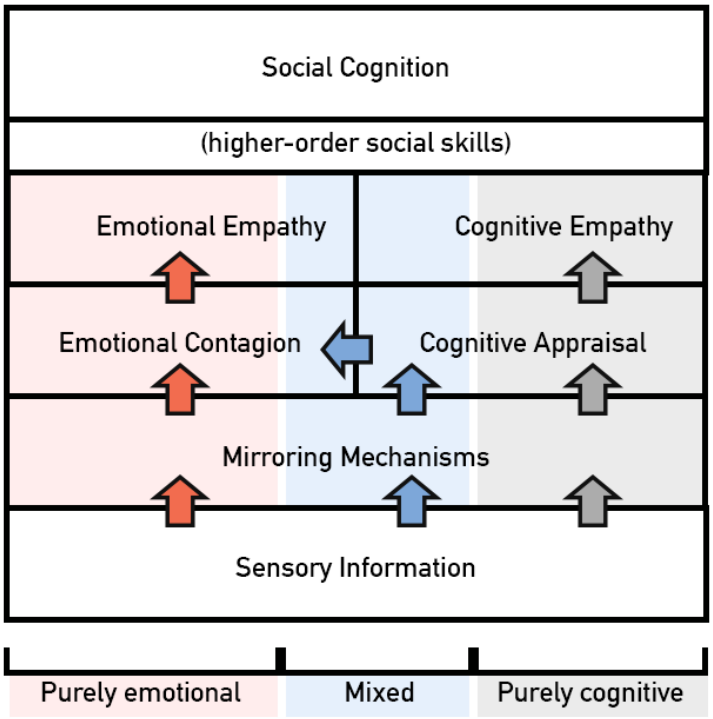


Figure 2.1 Taxonomy of processes underlying social cognition. The arrows indicate the flow of information (input/output) among the considered processes. Sensory information (external or internally generated via mental imagery) enables mirroring mechanisms to provide bodily representations of the interaction partner's phenomenological state. This mirrored state may elicit emotional contagion via an automatic route and be consciously attributed to the interaction partner via emotional empathy (emotional route). The mental state attribution can also be done without eliciting the very same state in the observer and by mean of a cognitive appraisal process, consciously attributing the mental state via cognitive empathy (cognitive route). Finally, the mirroring mechanisms may elicit emotional contagion, but the realised state can be updated or replaced by another state following a cognitive appraisal process (mixed route).

mirroring, and empathy as valid explanations. In the following Section 2.1.6, I will investigate the developmental process of social cognition via the proposed mechanisms, suggesting face-to-face interactions to be vital for promoting social cognition development.

2.1.6 Social Cognition Development

The acquisition of social and emotional competencies rely heavily on culture, ethics, social norms and common-sense (Eisenberg et al., 1998). Thus, human

development plays a crucial role in shaping social cognition capabilities. Human infants born tuned to their social environment and immediately prepared for social interaction (Grossmann, 2015). In fact, infants present from birth several biases preferentially orienting them to relevant social stimuli. Examples are a visual preference to faces, auditive preference to voices and a preference for biological motion over other types of motion (Grossmann, 2015).

The necessary social exchanges between infant and caregiver are supported by effective communication channels (Trevarthen, 2006). Humans, differently from other species, can use language to provide instructions and assist learning. However, learning the meaning of words happens only around 18 months (Halliday, 1975). Yet, even without this crucial ability, infants are able to learn vital skills, including elementary social and emotional capabilities (Hamlin et al., 2007). Therefore, infants are able to extract *social information* from social interactions even without being proficient in language (Reissland, 2013) or necessarily attributing mental states to others (Gallese et al., 2009). But how can it be possible?

Developmental studies suggest that this early social understanding is hardwired in human (Knapp et al., 2013). Imitative behaviour (*i.e.* imitation or mimicry) and mirroring are fundamental for enabling social interactions, since they allow relating, interacting and having a representation of others (Rochat et al., 2009), and they are available from birth at least in a rudimentary form (Meltzoff, 2007; Meltzoff and Moore, 1983; Trevarthen, 2006). As suggested in Sections 2.1.3 and 2.1.4, these tools bridge self to others' minds, thus enabling social understanding (Gallese, 2001). Since these capabilities are available from birth, are grouped under the concept of *innate intersubjectivity* (Trevarthen, 2006).

Obviously, without social partners, there cannot be social interactions and shared minds (Gallese, 2001). Thus, to promote social exchanges between infant and caregivers, humans are naturally inclined to raise their offspring. This is possible because of the peculiar biological design of humans' brain (and in more in general mammals' brain), which transform aversion to infant stimuli to an irresistible attraction (Rilling and Young, 2014). These social exchanges are further promoted by infant's *self-relevance* and *joint engagement* early capabilities.

Definition 2.14 Self-relevance *is the sensitivity to any sign of or change in an agent's action that indicates that the interaction or communication is directed at the infant (Grossmann, 2015).*

Definition 2.15 Joint engagement *is the ability of the infant to detect an external object or event as shared during the interaction with the caregiver (Grossmann, 2015).*

These capabilities are suggested to be *primary* for early social-cognitive development *"as they serve as a vital basis for learning from (and collaborating with) others"* (Grossmann, 2015, page 1280). Learning from social exchanges with caregivers and pairs is not made by accident, but it is *selective* and *intentional* (Rochat et al., 2009). The ability to become an intentional agent during social interactions constitutes the *primary intersubjectivity* stage of social development (Trevarthen, 2006).

Since language is not available from birth, the face is a valid alternative to creating social opportunities for the newborn and caregivers (Klinnert et al., 1986). For example, facial expression effectively communicates emotional values and facilitates learning (Niedenthal et al., 2014). Hence, during interactions with the infant, the caregiver speaks with a particular intonation and uses exaggerated facial expressions (Reissland, 2013). Self-relevance capability allows the infant to be aware of these interactions and it promotes engagement between the newborn and the caregiver. These emotional facial exchanges enable mechanisms in the infant advancing the extraction of social information via mirroring and emotional contagion, accordingly to what is suggested in Sections 2.1.3 and 2.1.4.

Facial exchanges are particularly crucial to learn the emotional valence of novel stimuli offered by the environment. For example, Zarbatany and Lamb (1985) showed that infants exposed to a novel toy look at the facial expression of the caregiver to attribute to the toy a positive or negative connotation. This is an example of *triadic exchange* between the infant, the caregiver and the target of the emotional evaluation (in this case the toy). This early behaviour provides knowledge of the target of others' attention and contributes to the development of *social referencing ability*.

Definition 2.16 Social referencing *is the ability to "use one's perception of other persons' interpretations of the situation to form their own understanding of that situation" (Feinman, 1982, p. 445).*

Importantly, this ability is not limited to the infant's parents, but it extends to a broader group of adults with whom the infant previously familiarised with (Klinnert et al., 1986).

Triadic exchanges in social interactions sign the passage from primary to *secondary intersubjectivity*: the infant begins to pay attention to how others act

and what they do with objects in everyday contexts, thus becoming aware of their intentions and contextualised actions (Gallagher and Hutto, 2008). Newborns, during second intersubjectivity, can appraise only limited concept to associate with the objects of their attention. Their world is essentially divided into either good and bad things or events, promoting respectively approach and avoidance (Rochat et al., 2009). To master cognitive appraisal processes, the child has to acquire a sense of self much more elaborated than the one necessary for mastering primary and secondary intersubjectivities (Lewis, 2008). Thus, infants necessarily pass from secondary to tertiary and later intersubjectivities when able to represent what others perceive of themselves and to use this representation to negotiate the values of events and things under their attention (Rochat et al., 2009). This ability is continuously refined, and enriched with the development of more complex emotions, such as pride, embarrassment and guilt, reaching almost the full range of adult emotions around the third year of life (Lewis, 2008). This richer sense of self allows the child to acquire Theory of Mind, thus promoting adult social cognition (Lewis, 2008).

Mirroring provides an effective and efficient way to learn from birth without the need of a developed language. These mechanisms are mediated by social exchanges with the caregiver, and the face has an especially important role in social interactions. By using triadic exchanges, infants can learn the emotional valence of the surrounding world, enrich their emotional repertoire and promote a more elaborated sense of self. This in turn allows the child to master Theory of Mind and acquire adult social cognition capabilities.

Given the particularly significant role of face in shaping social cognition, in Section 2.2 I will review literature on the fundamental functions of face processing. The resulting insights will motivate the computational model proposed in Chapter 5.

2.1.7 Summary

In Section 2.1.1, I reported the definition and significance of social cognition to offer background on the broad topic of this dissertation. In Section 2.1.2, I presented literature supporting the shared manifold hypothesis and Simulation Theory being valid explanations of how people understand others and attribute them mental states. These accounts are crucially aligned with embodied cognition standpoint, therefore motivating their investigation in this dissertation. In Sections 2.1.3, 2.1.4 and 2.1.5 I reviewed the possible mechanisms underlying

social cognition, identifying mirroring as a foundation of social cognition and emotional contagion and empathy as additional crucial processes promoting social cognition development. Finally, in Section 2.1.6, I discussed developmental studies to investigate how mirroring can promote social cognition development. I argued that social exchanges between the infant and the caregiver are extremely crucial for a correct development of social cognition capabilities and that the face is the most salient communication channel available from birth and promoting social interactions. Therefore, in this first section of the present literature review I introduced three main aspects that will motivate the rest of the work:

1. Simulation Theory and the shared manifold hypothesis provide a valid explanation of social cognition fitting embodied cognition standpoint. This is a desirable feature given the proposed broad contribution of this dissertation suggested in Section 1.2;

2. Mirroring mechanisms are likely to be at the core of this embodiment understanding of social cognition, thus motivating the development of appropriate computational tools able to describe their mechanisms;

3. The face is one of the most salient communication channel promoting social cognition development. The motor information displayed by face stimuli is processed by mirroring mechanisms, therefore linking face processing to embodied cognition research, which is again a desirable feature for the broad contribution proposed by this dissertation.

Section 2.2 will investigate face processing capabilities. These capabilities are significantly crucial for cognition. Therefore, connecting the discussed embodied mechanisms to the cognitive capabilities presented in Section 2.2 will necessarily enrich the argument of this dissertation.

2.2 The Face in Social Cognition

As discussed in Section 2.1.6, being able to direct attention to faces and process the provided social information is of paramount importance for promoting social cognition development. In order to gather information from faces, the baby has to visually *detect* them, *discriminate* over different identities and *attribute mental states* conveyed through facial expressions.

In particular, identifying other social agents is a critical capability necessary from birth. Distinguishing other humans from inanimate objects (Simion and

Di Giorgio, 2015), and discriminating a friend from a foe (Palermo and Rhodes, 2007) facilitate our survival. Faces are salient social stimuli, regardless what they are currently expressing through facial behaviour, since they convey critical social information, such as identity, race, sex, attractiveness and direction of eye gaze (Palermo and Rhodes, 2007). Therefore, this social information is of particular importance to support social exchanges.

When a face is detected and the interaction partner is identified, facial expressions further promote social cognition development. In fact, as I reviewed in Section 2.1.6, facial expressions are important regulators of social interactions during social referencing episodes (Matsumoto et al., 2008). Infants can use the facial expression of others to interpret the emotional valence of the target object of their attention (Zarbatany and Lamb, 1985). The social and emotional information extracted by faces is vital for solving social problems (Matsumoto et al., 2008).

According to neurological models of Haxby et al. (2000), human face processing recruits a complex and distributed neural system of multiple regions. As part of the core system of face processing, the author suggests three functionally distinct systems: the inferior occipital region, contributing to early stage of face perception, the lateral fusiform gyrus, processing invariant characteristics of faces (*e.g.* identity-related features), and the superior temporal sulcus, processing their dynamic aspects (*e.g.* facial expression and pose-related features). Despite this understanding of the neurological basis of face processing, the mechanisms underlying such capabilities are yet to be fully discovered (Simion and Di Giorgio, 2015).

In the following sections, I will provide a brief review of the current understanding of face processing. The insights gathered from this review will motivate the development of the computational model provided in Chapter 5.

2.2.1 Face Detection

The most basic aspect of face processing is the detection of a face in a visual scene. This process requires the extraction of common features of face stimuli facilitating their detection (Tsao and Livingstone, 2008). In this dissertation I will not model aspects of face detection, which I assume given when collecting formatted face stimuli. Nevertheless, this section is crucial to provide a more comprehensive understanding of literature in face processing.

Face stimuli are processed in a special way compared to other stimuli. Behavioural studies demonstrate that using modified face stimuli produces particular perceptual effects not reproduced when undergoing the same manipulations on other objects (Robbins and McKone, 2007). One peculiar characteristic of face stimuli, for example, is that they are processed in a more holistic way than stimuli representing other objects (Piepers and Robbins, 2012; Simion and Di Giorgio, 2015).

Definition 2.17 *The term* **holistic** *is used to suggest properties of a stimulus that are only apparent when considering it as a whole and they cannot be derived by only considering properties of its parts (Wagemans et al., 2012).*

A pure holistic understanding implies processing the object as a series of templates that cannot be described into subparts and interrelations between them (Piepers and Robbins, 2012). This can be a problem for face stimuli, since for taking into account variations of viewpoint, pose, expression and in general changeable features of the face it would be necessary a large set of templates for each individual identity (Piepers and Robbins, 2012). For this reason, although faces are processed more holistically than other objects, it seems more plausible that face processing requires at least a certain amount of configural and relational information of the perceived facial regions (*e.g.* mouth, nose, eyes) (Piepers and Robbins, 2012). Hence, in face perception literature the terms holistic and configural are sometimes considered within the same process (McKone and Yovel, 2009) and referred as *holistic/part-based model* (Piepers and Robbins, 2012).

There are two types of configural information in faces: *first-order configuration* and the *second-order configuration* (Piepers and Robbins, 2012).

Definition 2.18 First-order configuration of face *refers to the basic geometrical configuration of the main facial features, such as eyes above the nose and nose above the mouth (Piepers and Robbins, 2012).*

Definition 2.19 Second-order configuration of face *indicates the variations of distances between such features and their absolute positioning (Piepers and Robbins, 2012).*

Whereas first-order configuration is thought to be crucial for detecting faces and the sensitivity to it seems to be hard-wired from birth (Johnson et al., 1991), second-order configuration significantly facilitates discrimination between individual faces (Tsao and Livingstone, 2008), and it is refined throughout development (Sangrigoli

and De Schonen, 2004). Newborns show a preference for orienting toward simple schematic face-like patterns (T-shaped schematic face) respecting the geometry of basic facial features (Johnson et al., 1991).

The Binocular Correlation Model (Wilkinson et al., 2014) suggests that this neonatal bias for face stimuli is the result of a visual filtering mechanism depending on the limited binocular integration owned by infants that facilitate the detection of face-like patterns. Hence, preference for faces at birth can rely on the intrinsic perceptual structural properties of faces (and other similar objects), such as symmetry and highly contrasted areas (*i.e.* the eyes), which attract newborns' attention (Rochat et al., 2009).

As an alternative to the Binocular Correlation Model, Morton and Johnson (1991) propose a two-process model of face detection and processing suggesting that infants posses two separate mechanisms: *Conspec* and *Conlearn*. Conspec is a subcortical mechanism selectively tuned to geometrical features of the face. This mechanism is present at birth, and it guides the second cortical mechanism to acquire knowledge about faces and to specialise in face recognition. Thus, after development, Conlearn would be used to recognise faces (Johnson et al., 2015).

Simion and Di Giorgio (2015) reviewed evidence (Simion et al., 2003, 2001, 2006; Turati, 2004; Turati et al., 2002; Viola Macchi et al., 2004) supporting the hypothesis that the preference for face stimuli is the consequence of a more general bias preferring the structural and configural properties of the face and other similar stimuli, rather than a more specific preference for face stimuli only. Nevertheless, the debate is still open, thus requiring further investigations. As previously mentioned, I will not investigate processes of face detection. Instead, my focus will be on facial expression and identity processing, which I will introduce in the following Sections 2.2.2 and 2.2.3.

2.2.2 Face Identity Processing

Once a face has been detected, it may be processed to recognise the exhibited identity. Facial identity recognition requires a sensitivity to invariant visual information of the detected face in order to attribute an identity to it (Fitousi and Wenger, 2013).

Discriminating the identity of an individual is a vital social ability. It allows humans to establish long-term interactions, and above all, the crucial mother-infant attachment, as discussed in Section 2.1.6. Furthermore, this ability is also essential to assist social behaviour since it provides access to quick judgements. For instance,

an adult identifying the interaction partner as a child would result in different decisions compared to identifying the child as another adult. Thus, recognising the identity of the interaction partner enhances decision-making outcomes (Grossmann, 2015).

The primary cues assisting newborns in discriminating a face are its species, race and gender (Grossmann, 2015). The sensitivity to these cues is closely related to aspects of the social environment in which they are situated, characterising facial identity recognition mechanisms of the infants throughout their development (Grossmann and Vaish, 2009). An example supporting this claim is the difficulty adults have in identifying faces of races differing from the ones usually encountered during early years of development (Sangrigoli and De Schonen, 2004). This effect is known in the literature as the *"other-race effect"*, and it shows that facial identity recognition, contrarily to face detection, does include mechanisms able to adapt and develop with experience (Nelson, 2001). This effect diversified with respect to the social context in which the infant is situated (Pascalis et al., 2005; Sangrigoli et al., 2005).

The other-race effect is an example of *perceptual narrowing* process, observed in several perceptual modalities and domains in human development (Lewkowicz and Ghazanfar, 2009).

Definition 2.20 Perceptual narrowing *is the developmental process during which the brain uses perceptual information gathered from the environment to shape perceptual abilities (Scott et al., 2007).*

Hence, humans improve their ability in perceptually discriminating things often experienced during development. At the same time, they decrease the perceptual discrimination of things to which they are not often exposed. For instance, only 6 months old children can discriminate between identities of another species (*i.e.* monkeys), and this capability narrows down to discriminating identities of humans only when the infants reach 9 months of age (Pascalis et al., 2002). However, the brain may maintain the ability to perceptually discriminate things not often perceived during development, but this discrimination must be recruited through specific experience (Scott et al., 2007).

Overall, facial identity processing is a particularly vital ability promoting social cognition. Whereas it is still not explained how people discriminate identities, it is well supported by evidence that face identity recognition is not entirely an innate capability, but it includes mechanisms developing with experience, and its development highly depends on the social context in which the subject is situated.

2.2.3 Face Expression Processing

Recognising facial expression is a key process operating alongside identification. This additional information can be used as evidence to attributing mental states to others. Facial expression recognition requires a sensitivity to overt changeable feature of the detected face that helps to attribute certain states likely indexing its dynamics and/or behaviour (*e.g.* facial expression of emotion, pose and viewpoint) (Fitousi and Wenger, 2013).

Facial expressions arise from the visual impact that the contraction of facial muscles have on the face skin. The human face is a complex system of muscles, restraining in total 43 of them, also called mimetic muscles. Ekman and Friesen (1976) measured visual realisation of facial muscles activation with its Facial Action Coding System (FACS) in order to more easily describe the affect[5] displayed by facial expressions. The methodology suggests describing facial expressions as combinations of Action Units. The Action Units are descriptors of the movements of facial muscles (Ekman and Friesen, 1976), indicating if the corresponding muscle is contracted or not, and the level of contraction. Given the dynamic nature of muscle contractions, when processing facial expression (differently from facial identity processing), temporal dynamics play a crucial role in promoting better classification performance (Ambadar et al., 2005).

Although newborns have not enough cognitive capabilities to provide an interpretation of observed facial expressions, infants can at least match the observed facial expressions, even well before the development of early cognitive capabilities (Meltzoff and Moore, 1983) (but see Meltzoff et al., 2017; Oostenbroek et al., 2016; Simpson et al., 2016, for a recent discussion on the topic). In fact, Meltzoff and Moore (1983) showed that human infants of about 40 minutes old are able to mimic simple facial expressions, such as mouth opening and tongue protrusion. Therefore, facial expression recognition can be mediated by rudimentary imitative mechanisms available since the very beginning of life (Iacoboni, 2009). This important early ability can be further refined through reciprocal social exchanges between the caregiver and the baby (Iacoboni, 2009). Importantly, Meltzoff and Moore (1992) showed that this early behaviour cannot be simply explained as a reflex matching the observed action with the enacted one, but it encompasses a broader psychological framework.

[5]Affect is a basic sense of *feeling* coming from the body, a non-conscious experience of intensity (Shouse, 2005). This differentiates from emotion, that is a much more complex construction the subject can express through language.

Second-order configuration of face crucially assists facial expression recognition (Calder et al., 2000). Information about this second-order configuration of face can be gathered from the analysis of key points surrounding basic features of the face (McKone and Yovel, 2009). These markers are referred to as facial landmarks or fiducial points. Measuring the distances between each of these landmarks can facilitate facial expression recognition. Conversely, this information can be implicitly provided by representations encoding the face holistically (McKone and Yovel, 2009). In other words, the face stimulus can be processed holistically and represented by appropriate encodings able to implicitly communicate crucial information of face second-order configuration (see Chapter 5 for examples of this alternative understanding).

Overall, facial expression processing is a crucial capability promoting social cognition. Differently from facial identity processing, innate mirroring mechanisms can assist facial expression processing. This mirroring mechanisms can be refined through experience (Iacoboni, 2009).

2.2.4 Face Expression and Identity Interactions

In Sections 2.2.2 and 2.2.3 I reported studies to suggest the importance of face identity and expression recognition for everyday life and their links to social cognition skills. Although these two aspects of face processing are extremely important, researchers are still debating if there are interactions between them and, if so, how these interactions are shaped (Yankouskaya et al., 2014).

In face processing studies are proposed at least three possible views about this topic (Yankouskaya et al., 2014): (1) a parallel and complete separation of face identity and expression computations; (2) asymmetric processing of face identity and expression, with facial expression processing depending on facial identity coding, but not the vice versa; and (3) complete interactions between face identity and expression processing.

Traditional studies in face processing supported the first understanding (1). Under this perspective, facial identity processing is distinct from processes of other aspects of the face, such as facial expression. These models propose distinct pathways in the brain dealing with either invariant or dynamic features of the face (Bruce and Young, 1986; Bruyer et al., 1983). Nevertheless, modern understanding in face perception provides alternative models proposing an inter-relationship between invariant and dynamic features of the face (2,3) that are recently gathering

more consensus (Pell and Richards, 2013, but see Section 5.1 of this dissertation for more details).

There are at least two hypotheses supporting the idea that facial dynamics facilitate identity recognition: the supplementary information hypothesis and the representation enhancement hypothesis (O'Toole et al., 2002; O'Toole and Roark, 2010; Xiao et al., 2014). The supplementary information hypothesis suggests that facial movements provide facial characteristics distinctive of the individual in addition to static facial information, thus facilitating the identification of the observed face. The representation enhancement hypothesis proposes that facial movements facilitate the creation of more robust three-dimensional face representations, which better support face recognition.

In addition, the literature includes studies suggesting that invariant facial features affect facial expression recognition. For example, face stimuli of women and younger individuals appear to increase cues associated with happiness, whereas face stimuli of men and older individuals those of anger (Becker et al., 2007). Furthermore, Hess et al. (2009) demonstrated that facial configurations that create impressions of dominance and affiliation are the same that make a face appear to show respectively anger and happiness.

Additional interactions between facial expression and identity processing come from infants studies. Meltzoff and Moore (1992) suggested that one of the psychological functions that early imitation serves is to identify people. They argued that infants use the nonverbal behaviour of people to identify who they are and they use imitation as a tool to verifying the identity. In their study, Meltzoff and Moore (1992) exposed 6-weeks-old infants to two gestures: mouth opening and tongue protrusion. Each infant was exposed to two actors, mother and stranger. For each infant, one of the actors demonstrated one type of gesture, and the other actor demonstrated the other gesture. During this pilot, the infants were not controlled in visually tracking the entrance and exit of the two actors. The results showed that during the interaction with the second actor the newborns significantly responded more to the second gesture with the previously observed one. During a more in-depth analysis of the recordings, the authors were able to identify the possible reason for that. Newborns keeping visual track of the entrance and exit of the actors matched more the second observed gesture, suggesting that the adult gestures are not simple stimuli automatically triggering the infant's behaviour, but rather the infant willingly uses imitation to help discriminate among identities. The authors concluded that "(a) young infants do not have a fully developed system for determining person identity; and (b) infants use actions,

including facial gestures, as part of assimilating, 'knowing', and communicating with a person" (Meltzoff and Moore, 1992, page 17).

The study by Meltzoff and Moore (1992) is particularly relevant to the argument presented in this dissertation. Indeed, I will support a similar standpoint in this dissertation (see Chapter 6 and conclusions in Chapter 7), although revising it. In particular, in my conclusions I will suggest that infants may use imitative behaviour and/or mirroring mechanisms to facilitate facial expression recognition, and the retrieved motor information would be enough to develop cognitive mechanisms facilitating facial identity recognition skill. Importantly, the work by Leo et al. (2018) provides additional evidence in support to this hypothesis. The authors investigated the role of dynamic information for the acquisition of face recognition capabilities in newborns. They found that, whereas during habituation to static face stimuli infants show a preference for familiar stimuli (*i.e.* same identity), the newborns develop a preference for unfamiliar face stimuli (*i.e.* different identity) when habituated to dynamic face stimuli exhibiting happy and fearful facial expressions. These results suggest that motor movements of face stimuli facilitate identity recognition in infants. In addition, the authors found that newborns were not able to recognise identities when habituated to multistatic face stimuli, namely the very same stimuli used during the dynamic condition, but presented frame by frame in a non-fluid and static way. Therefore, the amount of pictorial information provided to the infants cannot be the explanation for the acquisition of face recognition abilities, which finds a more likely justification in the interpretation of the observed facial motor configuration.

2.2.5 Summary

In summary, in this review on face processing, I again stressed the importance of face for promoting social cognition. Face processing includes three main functions: face detection, facial identity processing and facial expression processing. Face detection is suggested to be an innate capability in humans, due to innate features of the visual system and to the specific configuration of face stimuli. Facial identity processing develops with time, and different social context leads to different discrimination abilities. Finally, facial expression processing has been suggested to rely on mechanisms ready from birth, but likely refined throughout experience. Knowing which mechanisms are innate and which ones develop with time is important to set assumptions and constraints for the present thesis argument.

In this literature review I have also reported contrasting hypotheses on the separation or integration of facial identity and expression processing mechanisms. Whereas traditional studies in face processing suggested a complete separation between facial expression and facial identity processing, recent studies promote understandings where processes of facial expression and identity interact. Both the two streams provide compelling evidence in support of their understanding and it is still unclear whether or not these mechanisms rely on separate brain structures and representations. In my dissertation, I will discuss a model that can provide an understanding able to unify these two contrasting views.

So far I proposed most of the connections between the covered topics. It is not a case that I decided to introduce this review on face processing between the literature review on social cognition and the one on embodied cognition research. The reason for doing so is that facial expression processing provides useful connections between social cognition, embodied cognition and cognition in general. In fact, on the one hand, facial expression processing connects to social cognition because necessary for social cognition development. Furthermore, facial expression processing connects to mirroring mechanisms and consequently to embodied understanding of social cognition (*i.e.* the shared manifold hypothesis and Simulation Theory). Finally, facial expression processing has been suggested to interact with facial identity processing, which is an important cognitive task, thus providing a valid connection with high-level cognition. Since the desired broad contribution of this dissertation is to advance embodied cognition theories, it is necessary to identify connections between embodied mechanisms (*i.e.* mirroring) and (social) cognitive processes (*i.e.* facial identity processing). In Section 2.3, I will further enrich the present discussion by providing the necessary background in embodied cognition theories.

2.3 Theories of Embodied Cognition

In Section 2.1 I provided a review of social cognition studies in order to give the reader a background of the broad topic presented in this dissertation, whereas Section 2.2 provided insights on the role of the face in social cognition. The intention of this section is to provide a more extensive background on embodied cognition theories. This section will provide the methodology used to assess the plausibility of the hypotheses discussed in this dissertation from the embodied cognition standpoint.

2.3.1 An Alternative to Classic Cognitivism

Embodied cognition is a research program born as a reaction to classical cognitive science comprising a set of methods from several fields, such as philosophy, neuroscience, psychology, and robotics (Leitan and Chaffey, 2014). Traditional cognitive studies suggest that cognition relies on mental representations having symbolic structures and quasi-linguistic properties (Fodor and Pylyshyn, 1988). Hence, intelligence is identified with higher-order reason and language (Goldman, 2013). The main argument of classical cognitivism is that thought is the manipulation of abstract symbols according to some rules (Goldman, 2013), going beyond the information contained in the input stream (Shapiro, 2010). This view is still prevalent in cognitive science studies, although concerns were expressed about its viability (Cowart, 2004).

The representations are symbolic in the sense that a single concept can relate to many different referents in the world (Wilson and Foglia, 2011). For example, the symbolic concept "chair" can refer to many actual types of chairs. These representations are amodal, in the sense that the interpretation of the representation is not constrained to a particular modality (Wilson and Foglia, 2011). Therefore, these representations do not offer any explicit relationship to physical and functional features of world referents (Wilson and Foglia, 2011). In addition, traditional cognitive studies suggest that cognitive processes begin in the brain with symbolic inputs, and they end with symbolically encoded outputs (Shapiro, 2010). Hence, classic cognitive science suggests internal representations employed in language, concept formation, and memory to be distinct from those processed by sensory-motor systems (Goldman, 2013, page 93): "*cognition occupies a level entirely segregated from perception and motor execution* (Goldman, 2013)".

Embodied cognition questions this view, by suggesting that cognition is not independent of the agent's bodily experience and its interactions with the environment: *cognition is strongly affected by aspects of the agent's body beyond the brain* (Wilson and Foglia, 2011). Thus, the body becomes directly involved in cognition and not secondary to it (Leitan and Chaffey, 2014). For instance, as an agent learns to control its own movements while performing goal-directed actions, it develops an understanding of its own perceptual and motor abilities, which lead to the acquisition of more complex cognitive processes (Cowart, 2004).

Thought and language becomes grounded in low-level cognition through sensory-motor processes (Goldman, 2013). Hence, the key assumption of this research program is that the body acts as a constituent of the mind rather than

being a passive perceiver and motor actor serving the mind (Leitan and Chaffey, 2014). It is not necessary to precisely quantify in how many cognitive tasks the body plays this role, but it is assumable that embodied cognition thesis suggests embodiment having a dominant role in cognition (Goldman and de Vignemont, 2009).

Despite proposing a contrasting view, embodied cognition does not necessarily mean to depart from classical cognitive science completely. Embodied cognitive science could still accommodate a level of symbolic representation. However, what makes it to differ from classical approaches is the replacement of a pure propositional encoding with a modality-specific representation, which can bind the symbol to the actual world referent, thus enabling its interpretation (Wilson and Foglia, 2011).

An example of how symbolic representations can be integrated with modality-specific features is Barsalou's perceptual symbols theory (Barsalou, 1999). This theory holds on the assumption that human cognition does not consist of amodal representations having arbitrary relations to their world referents, but rather representations whose activation patterns include information from various sensory modalities. Hence, an implication of this theory is that symbols are not independent of the biological system that embodies them, and the content conveyed would be likely to vary if intelligent systems varied physically (Wilson and Foglia, 2011). Table 2.1 summarises the main points of the two perspectives.

Classic cognitivism	**Embodied cognition**
Computer metaphor of mind: manipulation of symbols based on rules.	Coupling metaphor of mind: embodiment, environment and action shape mind.
Cognition can be understood in isolation by focusing on the agent's internal processes.	Cognition needs to be understood with the analysis of the interplay between mind, body and environment.
Cognition is primarily achieved by computational mechanisms.	Cognition is primarily achieved by goal-directed actions.
Cognition passively retrieves information.	Cognition actively constructs information, depending on the agent's embodiment and its interactions with the environment.
Symbolic and amodal representations	Sensory-motor representations

Table 2.1 Comparative analysis of classical cognitivism and embodied cognition (table replicated and adapted from Cowart, 2004).

2.3.2 Research Methodologies in Embodied Cognition Studies

Although the central claim of embodied cognition is generally that sensory-motor capabilities of an agent and its interactions with the environment enable the development of particular cognitive capacities (Cowart, 2004; Winkielman et al., 2015), there are different conceptions of embodiment in the literature, leading to different accounts and degrees of how strongly cognition is believed to be shaped by the body and its interactions with the environment (Shapiro, 2010). Furthermore, since embodied cognition research program is still in its early stages and researchers are yet to define an exact definition of what is meant by embodied cognition (Cowart, 2004), the task of unifying this research program under a single methodology becomes harder.

Shapiro (2010) suggests to organise current accounts of embodied cognition theories into three streams of research: the *conceptualisation* hypothesis, the *replacement* hypothesis, and the *constitution* hypothesis. In the following paragraphs, I will briefly review the proposed accounts (but see Shapiro, 2010, for a more exhaustive review). I will provide conclusions for each of the reviewed accounts in relationship to the aimed broad contribution of this thesis presented in Section 1.2, namely providing evidence supporting an embodiment understanding of social cognition. This will help to identify the best methodology to assess the plausibility of the hypotheses introduced in the remainder of this dissertation.

The Conceptualisation Hypothesis

The conceptualisation account of embodied cognition argues that the experiences and concepts an organism possesses are constrained by the nature of the oganism's body (Shapiro, 2010). An example that can be included in this stream is Varela et al. (1991) conception of cognition as *"embodied action"*. Under this view, perception and action becomes inseparable aspects of cognition. In fact, Varela et al. (1991) suggest that a successful interaction with the environment will require a tight loop between sensory and motor processes; the agent's motion would allow perceiving new features, whereas perceiving such novel features would influence future motions, and so on (Shapiro, 2010). Since cognition is affected by sensory-motor processes, the world becomes perceiver-dependent, and it is not predetermined.

For instance, opportunities to interact with the environment through an action (*i.e. affordances*) are not predetermined in the world, but are the result of the interaction between the perceived environment and the motor cortex (Maranesi

et al., 2014). These potential motor acts depend on the organism's embodiment (Glenberg and Kaschak, 2002): a branch affording a resting place for a bird, and so activating a potential motor act of lying on such branch, would not have the same effect for a pig (Shapiro, 2010). Different embodiments would lead to the creation of different concepts and the development of different cognitive capabilities (Shapiro, 2010).

Other notable examples of works falling into the conceptualisation research stream are the studies dealing with the "symbol grounding problem' 'Barsalou (1999); Searle (1980). This problem asserts that symbols cannot acquire meaning simply in virtue of bearing relationships to other symbols, but they need to be grounded to external world referents. Perceptual symbols theory proposed by Barsalou (1999) is a milestone study proposing a theoretical solution to the symbol grounding problem. Traditional cognitive science suggests that the stimulation of several sensory systems is translated into a code that is then processed further into amodal symbols used in language and thought. Barsalou (1999), on the contrary, proposes that this step is unnecessary. Modal information provided by the perceptual system does not need to be detached from cognitive manipulation of the corresponding representation. Perception and cognition share systems at both the cognitive and neural levels (Barsalou, 1999). During a perceptual experience, association areas in the brain capture bottom-up patterns of activation in sensory-motor areas. When recalling the same perceptual symbol, the same association areas partially reactivate sensory-motor areas to implement it. Continuous interactions with a perceptual component (*e.g.* chair, green, cake) would extract schematic representations stored in memory. Memories of the same component can be used to implement a simulator able to produce limitless simulations of the component. These established simulators can implement a basic conceptual system representing types, supporting categorisation and allowing inferences and reasoning (Barsalou, 1999). Hence, perceptual symbols are simply reconstructions of representations as they appeared in their original perceptual coding (Barsalou, 1999), and a different embodiment would result in different perceptual symbols leading to different concepts and cognition (Shapiro, 2010).

This hypothesis can potentially provide the methodology necessary to address the aimed broad contribution of this dissertation. In fact, Shapiro (2010) suggests that the conceptualisation hypothesis does provide a competitive explanation of cognition with respect to standard cognitive science. However, it fails in providing a case against symbolic representation (Shapiro, 2010), thus making standard cognitive science account preferable. Also, some of the experimental results

included into the conceptualisation hypothesis stream can find an explanation in standard cognitive science, thus making the hypothesis less plausible (Shapiro, 2010). Therefore, I suggest that the conceptualisation approach of embodied cognition is not the most desirable methodology to assess the hypotheses I will discuss in this dissertation.

The Replacement Hypothesis

The replacement hypothesis argues that traditional methods and concepts of cognitive science should be abandoned in favour of alternative methods and conceptual foundations (Shapiro, 2010). Cognition cannot be explained through computational models consisting of symbols manipulation and a linear input-output processing system. In addition, representations are not necessary to achieve intelligence.

In this account, cognition is viewed as an *emergent* phenomenon. In other words, it does not exist a single module of cognition, but the dynamical interactions of local components give rise to intelligent behaviour. Mind, body, and world work as a system changing over time and developing cognition (Leitan and Chaffey, 2014). Hence, dynamical system theory, an area of mathematics studying the evolution of models consisting of several interacting components, provides valid descriptions fitting under the replacement account (Leitan and Chaffey, 2014). In an emergent system, no top-down rules are governing the behaviour of the parts in realising outcomes, but rather parts self-organise according to constraints and opportunities offered by the environment. An example of emergent cognition without representations is the "Subsumption architecture" designed for robotics systems proposed by Brooks (1990). The main idea of this architecture is that instead of designing the robot planner as components transmitting representations through the classic sense-model-plan-act paradigm, the decomposition should be made in terms of components directly connecting sensing with behaviour. In this way, sensing feeds directly into action without the need of using a representational stage (Shapiro, 2010). Cognition emerges from the interaction between the agent and the environment (Brooks, 1990).

An additional concept for the replacement account of embodied cognition is *coupling* (Leitan and Chaffey, 2014; Shapiro, 2010). Coupling is the idea that parts of a system, when described through mathematical formalism, must include a term referring to other parts of the same system. Thus, every part of the system

is always in relation to other components and never a discrete entity (Leitan and Chaffey, 2014).

The replacement account is definitely a real competitor of traditional cognitive science (Shapiro, 2010). In fact, it offers a complete departure from traditional methodologies in cognitive science. Furthermore, the works included in this stream provide compelling alternatives that do not make use of concepts or representations to achieve cognition. Cognition is suggested to be a phenomenon emerging from dynamical systems based on simple coupled components. Therefore, contrary to conceptualisation hypothesis research, this approach can more plausibly win a competition against traditional cognitivism (Shapiro, 2010), thus being a suitable methodology to assess the hypotheses discussed in this dissertation.

However, the replacement hypothesis proposes a too strong perspective completely departing from traditional cognitive science research. Since one of the aimed benefits of this work is to promote integration of traditional cognitive science research and embodied cognition research (Section 1.3), the replacement hypothesis is not the most suitable methodology to assess the hypotheses introduced in this work. Therefore, I will not use this research methodology in my dissertation, thus focusing on another plausible and more gentle methodological approach.

The Constitution Hypothesis

The constitution hypothesis argues that *constituents* of cognitive processes extend beyond the brain (Shapiro, 2010). In order to have a better understanding of what suggested by this account, it is important first to define what it means being a constituent of the mind, instead of a simple cause. Given two processes **X** and **Y** showing some interactions, these interactions are either causal or constitutional based on the following definitions (Shapiro, 2010, page 170):

Definition 2.21 Cause: *a process* **X** *is a cause or it affects a process* **Y** *if* **X** *is separable from* **Y** *in the sense that it happens before* **Y** *exists, or takes place in a location distinct from* **Y***'s.*

Definition 2.22 Constituent: *a process* **X** *is a central or important constituent of a process* **Y** *if* **Y** *would fail or be something else without* **X***'s presence;*

Thus, in order to fit under the methodology suggested by the constitution hypothesis research approach, the advanced hypotheses have to present evidence of a *constitutive* role of *embodiment* for the correct functioning of a *cognitive task*.

A mere causal relationship between the two processes would not be enough to support the case in favour of the constitution hypothesis.

An example of work under the constitution hypothesis research stream is Clark's coupling argument (Clark and Chalmers, 1998). This argument suggests that certain objects in the external world are utilised by the mind in such a way that they couple with cognition thus being extensions of the mind itself into the environment. For instance, consider the task of solving a multiplication. This tasks can be achieved solely in the head, or by making use of paper and pencil. In both the cases, the obtained result is the same. However, in the second instance the paper and pencil couple with the mind of the mathematician, thus extending his mind into the environment. This is one of the strongest views of embodiment, named "extended mind thesis". This thesis suggests that cognition extends not only beyond the brain into the body, but beyond that as well, thus extending into the surrounding world (Clark and Chalmers, 1998).

However, there are other weaker perspectives of embodiment included under the constitution hypothesis methodology. These works limit embodiment interpretation to aspects of the body, thus not necessarily extending to the environment (Rauscher et al., 1996). In addition, some other works in constitution hypothesis research limit the meaning of embodiment further, by proposing an interpretation based on mental representations (Goldman, 2013). In this section, I will introduce the former standpoint, namely embodiment as an extension into aspects of the body, while I will dedicate a full section about bodily mental representations since I will make use of them to support the hypotheses offered by this dissertation.

An example of body as constituent of cognition is gesturing. Rauscher et al. (1996) showed that preventing subjects from gesturing when asking them to describe situations with spatial content significantly increases dysfluencies in their speech. Nevertheless, subjects free to move did not show significant dysfluencies. Importantly, the effort required to prevent gesturing did not explain the observed dysfluencies. In fact, when speaking about non-spatial situations this group of subjects showed greater fluency than subjects in the group allowed to gesture. Thus, gesturing has not only a communicative role, but it facilitates lexical access (Rauscher et al., 1996).

This work fits the constitution hypothesis methodology since in this experiment gesturing was observed to structure information and assist cognition (Shapiro, 2010), similarly to the paper and pencil of the coupling argument (Clark and Chalmers, 1998). Tversky (2009) tested the ability of subjects in solving a set of problems. The authors divided the subjects into two groups and only to the

first group was provided pencil and paper. Surprisingly, the individuals without paper and pencil made use of gestures to solve the same problems solved with paper and pencil by the other group. Gestures may be used to assist off-loading and organising spatial working memory, as paper and pencil did, thus potentially being considered constituent of cognition (Tversky, 2009).

Differently from the conceptualisation and the replacement hypotheses, constitution hypothesis does not present a methodology completely departing from standard cognitivism (Shapiro, 2010). In fact, most of the arguments and works under this stream can be explained by knowledge depending on sensory-motor and environmental contingencies, which is represented in the brain, as standard cognitive science would agree (Shapiro, 2010). However, it is still possible to discuss which kind of knowledge and how this knowledge is represented in the brain (*i.e.* propositional *vs.* non-propositional/bodily), thus possibly advance an embodied understanding of cognition.

Constitution hypothesis research provides a valid methodology able to extend standard cognitivism without completely neglecting its previous findings (Shapiro, 2010). In this methodology, traditional cognitive science theories are not rejected, but only re-fitted (Shapiro, 2010). It is possible to maintain what is good in traditional theories of cognition and extend them with new insights coming from embodied cognition theories.

Furthermore, constitution hypothesis provides a clearer methodology for the validation of its theories than conceptualisation and replacement hypotheses do. In fact, to validate this hypothesis is enough to demonstrate that an embodied process constitutes another cognitive process, namely that without the embodied process the cognitive process fails or becomes something else. Finally, the considered embodiment interpretation should not be only physical but it can be limited to mental representations resembling bodily features (as I will discuss in the following Section 2.3.3).

As Shapiro (2010, page 210) concludes: "If I'm right that pursuit of Constitution comes without a cost to standard cognitive science, then there's no harm in trying, and, perhaps, tremendous benefit". Hence, given all these convenient reasons, in this dissertation I will make use of the constitution hypothesis methodology to assess my hypotheses, thus adding computational evidence fostering embodied cognition research. In the following section, I will investigate the topic of *bodily mental representations*. I will add significant insights motivating the development of the computational account proposed in Chapter 4.

2.3.3 Embodiment via Bodily Representations

In the previous section, I introduced some interpretations of embodiment proposed in embodied cognition literature. For example, one interpretation suggested embodiment be the body anatomy, while other interpretations extended the concept to bodily actions, and others further extended it to a coupling between mind and environment. However, there is an additional interpretation of embodiment suggesting *bodily mental representations*. This interpretation offers a more moderate perspective of cognition embodiment, but it also gives an opportunity to integrate standard cognitive science research with embodied cognition theories (Goldman, 2012).

Goldman and de Vignemont (2009) introduced the concept of bodily mental representations (from now on abbreviated with b-reps) after reviewing different notions of embodiment in embodied cognition literature and providing their taxonomy. They argued that the currently existing interpretations of embodiment were not satisfying the constraint for a fruitful definition of embodied cognition. Goldman and de Vignemont (2009) offered a list of features recommended for a suitable definition of embodiment (Goldman and de Vignemont, 2009, page 154):

Definition 2.23 *An interpretation of* **Embodiment:**

(i) should assign central importance to the body (understood literally), not simply to the situation or environment in which the body is embedded;

(ii) should concentrate on the cogniser's own body, not the bodies of others. Perception of another person's body should not automatically count as embodied cognition.

Furthermore, an embodied cognition thesis should present the following features (Goldman and de Vignemont, 2009, page 154):

Definition 2.24 *An* **Embodied Cognition Thesis:**

(iii) should be a genuine rival to classical cognitivism;

(iv) should also make a clear enough claim that its truth or falsity can be evaluated by empirical evidence.

Goldman and de Vignemont (2009) suggest that an interpretation based on body anatomy is not really competitive with classic cognitive science. In fact, traditional cognitive science similarly recognise the crucial role of physical constraints

the body has on our perception and cognition. Thus, an interpretation of embodiment in terms of body anatomy is not desirable, since it violates requirement (iii) for a fruitful embodied cognition thesis as defined in Definition 2.24.

Similarly, an interpretation of embodiment in terms of bodily actions is not really competing with classic cognitivism, since it is recognised the role of body movements in influencing perception and cognition. In addition, the available evidence from studies on embodied cognition demonstrated only a causal role of sensory-motor contingencies on perceptual experience and not a constituitive role. Therefore, embodiment defined in terms of bodily actions violates requirement (iii) of Definition 2.24.

Hence, Goldman and de Vignemont (2009) recommend an interpretation of embodiment in terms of b-reps. This interpretation can be further divided into two competitive sub-interpretations: b-reps in terms of bodily contents, and b-reps in terms of bodily formats.

The bodily content interpretation stands for mental representations having contents depending on body anatomy or somatosensory and visceral activity of the body. However, this interpretation is not enough to depart from classical cognitivism. For instance, it is possible to provide a bodily content representation of 'waving my hand' in a purely symbolic format, although having contents related to the body. This representation would not depart from classic cognitivism, thus violating requirement (iii) of Definition 2.24 defined previously on page 56. On the contrary, the bodily formats interpretation put a stronger constraint on the structure of the representations. Here is the format itself that has to be related to the body, and having only a bodily content is not sufficient.

Definition 2.25 *A* **bodily formatted representation** *is a mental representation satisfying the requirements of a genuine embodiment interpretation as per Definition 2.23, and posing constraints on what that representation can represent (Goldman and de Vignemont, 2009). These constraints are determined by the specific configuration of the human body (Gallese, 2016).*

The difference between content and format is subtle but significant. For instance, an address can be provided through natural language, such as "15 Broadway, Ultimo NSW 2007", or through a set of coordinates. In both cases, the content of the representations is the same (*i.e.* the same address), but formatted in two different ways (the first amodal, the second visually via a two-dimensional map).

Importantly, an embodiment interpretation based on bodily formatted representations (from now on abbreviated with b-formats) provides a competitive

approach to cognitive science, without completely departing from it. In fact, cognitive science and neuroscience widely support the idea that information in the brain is organised into representations with specific formats (Goldman, 2012). For instance, the neural substrate of experiental representations of the body are encoded onto topographically mapped region of the somatosensory cortex (Gazzaniga, 2004), whereas activation of areas in the motor cortex is topographically organised for representing bodily effectors and enable movement commands to be sent to those effectors (Goldman, 2012).

The integration of b-formats with standard cognitive science is a convenient feature able to satisfy one of the aimed benefits of this work discussed in Section 1.3. In addition, Goldman and de Vignemont (2009) suggest that b-formats interpretation of embodiment is the most likely interpretation to support social cognition embodiment. These representations are suggested to be used to form interoceptive or directive representations of one's own bodily states and activities (Goldman, 2012), thus becoming crucial in mediating social cognition capabilities. Given the suitability of the discussed features to meet the desired objectives of the present dissertation, the computational account of embodied mechanisms proposed in Chapter 4 will make use of b-formats interpretation of embodiment.

The following section will review the debates in favour and against social cognition embodiment. This will conveniently complement Section 2.1.2, thus closing the literature review with a link back to the broad topic of this dissertation, namely social cognition.

2.3.4 Social Cognition Embodiment

In Sections 2.1.3, 2.1.4 and 2.1.5 I provided a summary of the core processes underlying social cognition, and I discussed how mirroring can plausibly be at the heart of social cognition (see Figure 2.1 on page 34). Through mirroring mechanisms a 'cognitive continuity' can exist within the domain of intentional state attribution in humans, and the mirror neuron system represents its neural correlate (Gallese, 2016). Mirroring allows to detect and match actions and goals of others in own mind (and body), thus granting their understanding. Therefore, mirroring is an implicit and functional way for attributing a given mental content to others (Gallese, 2016).

In order to provide an additional modern understanding of mirroring mechanisms functions, Gallese introduced the new *'embodied simulation theory'*. This theory suggests (a) to extend the function of mirroring, so far limited to promote

attribution of mental states to others (as discussed in Section 2.1.2), to other cognitive capabilities, and, as consequence, (b) it promotes an embodiment understanding of cognition (Gallese, 2005, 2016; Gallese and Sinigaglia, 2011). In other words, differently from Simulation Theory, embodied simulation theory does not limit the function of simulation mechanisms to promote the attribution of mental states to others, but it suggests that simulation mechanisms provide broad functionalities shaping several aspects of cognition (Gallese, 2005). Therefore, embodied simulation theory does not present a threat for Simulation Theory in the broad sense (*i.e.* attribution of mental states to others), but it becomes complementary to it (Gallese, 2016).

The simulation process enabled by embodied simulation mechanisms is embodied because its function is realised by mirror neurons modulating sensory-motor information. This neural system uses a pre-existing body-model available in the brain of the subject, and it does not require a propositional form of knowledge (Gallese, 2005). Indeed, Goldman and de Vignemont (2009) suggest that this non-propositional knowledge can be likely shaped in b-formats. Therefore, mental states and processes enacted during embodied simulation episodes and represented by common b-formats can be *reused* for other abilities, thus extending their use beyond mind-reading tasks (Gallese, 2016); mirroring mechanisms and b-formats interpretation of embodiment become central concepts within embodied cognition research.

The reuse of the realised bodily representations for other aspects of cognition finds its origins in the *reuse hypothesis*. This hypothesis suggests that mental simulations occurring for one purpose are reused for another purpose (Gallese and Caruana, 2016). This is different from the approach suggesting simulation as *resemblance*, which advocates that simulation mechanisms occur to simply copy mental states of others for getting an understand of them, as suggested by classical Simulation Theory models (Gallese and Caruana, 2016).

Definition 2.26 Simulation as reuse. *Mental simulations occurring for one purpose are reused for another purpose (Gallese and Caruana, 2016).*

Definition 2.27 Simulation as resemblance. *Simulation mechanisms occur to simply copy mental states of others for getting an understanding of them (Gallese and Caruana, 2016).*

Simulation as reuse denies that the simulation process is strictly used for attributing mental states to others, but instead, it suggests that simulation

mechanisms provide more basic and general functions that can be employed to shape other cognitive capabilities. Gallese (2005) suggests that mind-reading capability can be just the result of an evolutionary reuse of the mirror neuron system, which was originally employed by the brain to achieve other capabilities (Gallese, 2005). The mirroring mechanisms could have been originally selected for solving basic sensory-motor matching but adapted for shaping mind-reading capabilities assisting human social behaviour (Gallese, 2005).

Gallagher (2015) contests embodied simulation theory by suggesting that it does not promote a valid embodied interpretation. In fact, the theory lies on top of *mental* representations that cannot really account for the body *per se*, thus offering a too weak interpretation of embodiment insufficient to support embodied cognition theories. However, it is important to note that body and brain cannot be dissociated (Shapiro, 2010); a standpoint where cognition is suggested to be purely mental and symbolic without considering the body as one of its constituents is limited as much as a standpoint considering only the *physical* body as mean of cognition. Rather, it is the body-brain system that makes possible cognition development. In fact, at a certain stage, the body has to be coded somehow in the brain in order to be used for achieving cognitive capabilities. Without a body, there cannot exist bodily representations. Similarly, without bodily representations, there cannot be an active body. Thus, completely dissociating the brain from the body and the body from the brain will limit an understanding of cognition.

Importantly, the body is not only the physical matter but also the mental representations of it (Goldman, 2013). For this reason, this view is not reductive of a full-bodied account as suggested by Gallagher (2015). Bodily representations are vital for social interactions with other social agents (Gallese and Caruana, 2016). Given this argument, people with paralyses, body disabilities or lack of sensory input should exhibit deficits in social cognition skills. It is important to mention that a person can be physically paralysed or lacking sensory input, but still able to mentally activate at least partial sensory-motor representations of the considered actions (Conson et al., 2008; Ricciardi et al., 2009). For example, Bate et al. (2013) studied subjects affected by Möbius sequence, a condition characterised by congenital bilateral facial paralysis. The authors found that the majority of the clinical subjects participating to the study did not exhibit deficit in mental simulation of facial expressions and they exhibited only partial deficits in facial expression and identity recognition tasks. In addition, Matsumoto and Willingham (2009) found that congenitally and non congenitally blind individuals did not differ

from sighted individuals in the way they produced spontaneous facial expressions of emotions. Their findings suggest that emotional expression might originate from an evolved and potentially genetic source common to all humans, regardless of their disability (Matsumoto and Willingham, 2009). Indeed, Ricciardi et al. (2009) found that blind patients and sighted participants activated similar brain motor areas when presented with actions either aurally presented, in the case of blind people, or visually pantomimed, in the case of sighted subjects. Their finding suggests the presence of a supramodal sensory representations of motor actions that allow individuals with no visual experience to interact effectively with others. Therefore, it is possible that people having physical dysfunctions can still have (partially) preserved mental motor representations that are enough to allow them succeeding in social skills.

In Chapters 3 and 6, I will suggest that dysfunctional facial motor representations lead to face processing deficits. In fact, I will argue that impaired bodily representations do not allow a correct simulation of phenomenological states in the impaired individual and, for this reason, these subjects cannot fully understand others' motor actions because there are no tools (or 'broken' tools) to make sense of it: *"who you are and what you can experience thus affect the way you perceive others"* (De Vignemont, 2009, page 464).

2.3.5 Summary

In summary, in this review on embodied cognition theories I reported the following main insights:

(i) Embodied cognition theories provide a valid perspective to enrich traditional cognitive science research since it suggests that the world is not represented in the mind by amodal propositional knowledge, but this knowledge is strictly connected to other levels of cognition, such as the perceptual level and the motor level. In this way, for example, it is possible to attribute meanings to objects in the world;

(ii) Whereas conceptualisation and replacement research approaches provide valid methodologies to advance embodied cognition research, constitution research approach is the most favoured one for this dissertation. In fact, it presents an interpretation of embodied cognition that can still integrate well with traditional cognitive science research, although enriching it with new perspectives. Furthermore, this account proposes a methodology to validate

embodied cognition thesis, which is preferable than the ones suggested by conceptualisation and replacement approaches;

(iii) The interpretation of embodiment necessary to validate embodied cognition thesis does not have to be necessarily physical, but it can be limited to mental representations having bodily formats;

(iv) Mirroring mechanisms and bodily formatted representations can extend their function to promote other cognitive capabilities, thus not limiting to the attribution of mental states to others.

In this dissertation (i) I will provide hypotheses advancing embodied cognition, but at the same time not depart completely from traditional cognitive science studies. (ii) I will assess the plausibility of such hypotheses by using the methodology proposed by the constitution hypothesis research approach. (iii) I will make use of bodily representation with bodily format as the interpretation of embodiment. In this way (iv) I will be able to suggest that mental simulation function promoted by mirroring mechanisms is not only crucial for attributing mental states to others, but it shapes other vital aspects of cognition.

2.4 Research Gaps and Dissertation Scope

In this section, I will define the extent of this dissertation by summarising the identified research gaps and their significance in addressing the aimed broad contribution of this thesis. In Sections 2.1.2 and 2.3.4, I argued that simulation mechanisms can plausibly explain human social cognition development. Within this perspective, I presented studies showing that mirroring mechanisms, having their neural correlate in the mirror neuron system, are the catalysts for social capabilities.

Although the literature presents computational theories explaining such mechanisms (Oztop et al., 2006), these models often limit their investigation to map sensory-motor information in order to achieve motor control capabilities or to infer mental states from others. To the best of my knowledge, there are no computational models of embodied simulation connecting to other cognitive skills (but see Boccignone et al., 2018, for a more recent and extended version of the theory and model presented in Chapter 4), and suggesting the reuse of mirroring mechanisms for other capabilities. This is a significant limitation for embodied

cognition research since **it does not provide computational evidence suggesting that embodied mechanisms can be *reused* for several cognitive capabilities**, perhaps even unexpected capabilities traditionally proposed to have strictly cognitive development, like facial identity recognition (Nelson, 2001). This dissertation will address this gap, thus fostering embodied cognition research.

However, the more important question is not whether embodied mechanisms are related to cognition, but *how they are related* to it (Kilner et al., 2007). Therefore, I reviewed the available methodologies to assess embodied cognition theories in Section 2.3.2, and I identified constitution hypothesis research as the best methodology to assess embodied cognition theories. Unfortunately, **the current state of knowledge is still far from providing definitive evidence in favour of a constitutive role of embodiment for cognition development** (Gallese and Sinigaglia, 2011). Therefore, this work will provide computational evidence suggesting the plausibility of the constitution hypothesis.

In Section 2.2, I reviewed findings in human face perception and processing. I showed how faces are an important and universal communication channel, crucial for the development and learning of social skills (Palermo and Rhodes, 2007). Face identity and expression processing are dependent on task and experience, but their computational mechanisms are not yet well understood (Yankouskaya et al., 2014). For example, we do not know yet how face identity and expression processing interact (Yankouskaya et al., 2014), how this interaction is affected by experience (Yankouskaya et al., 2014) and where in the face processing hierarchy representations of invariant and dynamic facial features interact (Calder and Young, 2005; Yankouskaya et al., 2014). Advancing understandings of 'how' face processing is performed and its underlying mechanisms is an inherently more theoretical and harder task than testing 'where' in the brain specific capabilities are performed (Redcay, 2008). Therefore, providing computational theories answering these questions is a significant contribution to enrich cognitive science research (Simion and Di Giorgio, 2015). In this dissertation I will provide these necessary computational theories so to provide valid explanations of the mechanisms underlying face identity and expression recognition processing. Specifically, I will provide a computational model of face processing mechanisms that can be used for computational simulations explaining face processing mechanisms. I will use this model to suggest how face identity and expression coding interacts, how face recognition can be acquired using motor representations, and how dysfunctional motor representations can impact face processing skills. Finally, I will comment how the proposed model integrates traditional and modern understandings of

face perception and cognition and suggest at what processing level invariant and dynamic facial features may interact.

Chapter Bibliography

Ambadar, Z., Schooler, J. W., and Cohn, J. F. (2005). Deciphering the enigmatic face the importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, 16(5):403–410.

Bargh, J. (1994). The four horsemen of automaticity: Intention, awareness, efficiency, and control as separate issues. *Philosophical Explorations*, 15(3):255–275.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4):577–660.

Bate, S., Cook, S. J., Mole, J., and Cole, J. (2013). First report of generalized face processing difficulties in möbius sequence. *PLoS One*, 8(4):e62656.

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K., and Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, 92(2):179.

Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., and Lanzarotti, R. (2018). Deep construction of an affective latent space via multimodal enactment. *IEEE Transactions on Cognitive and Developmental Systems*.

Braddon-Mitchell, D. and Jackson, F. (2006). *Philosophy of Mind and Cognition: An Introduction*. Wiley-Blackwell.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, 6(1):3–15.

Brothers, L. (1990). The neural basis of primate social communication. *Motivation and Emotion*, 14(2):81–91.

Brothers, L. (2002). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Foundations in Social Neuroscience*, pages 367–385.

Bruce, V. and Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77:305–327.

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., and Rectem, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. *Brain and Cognition*, 2(3):257–284.

Calder, A. J. and Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6(8):641–651.

Calder, A. J., Young, A. W., Keane, J., and Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2):527.

Clark, A. and Chalmers, D. (1998). The extended mind. *Analysis*, 58(1):7–19.

Conson, M., Sacco, S., Sarà, M., Pistoia, F., Grossi, D., and Trojano, L. (2008). Selective motor imagery defect in patients with locked-in syndrome. *Neuropsychologia*, 46(11):2622–2628.

Coplan, A. and Goldie, P. (2011). *Empathy: Philosophical and Psychological Perspectives*. Oxford University Press.

Cowart, M. (2004). Embodied cognition. *The Internet Encyclopedia of Philosophy*.

Davies, M. and Stone, T. (1995). *Folk Psychology: The Theory of Mind Debate*. Blackwell.

De Vignemont, F. (2009). Drawing the boundary between low-level and high-level mindreading. *Philosophical Studies*, 144(3):457–466.

De Vignemont, F. and Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, 10(10):435–441.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, 91(1):176–180.

Eisenberg, N., Cumberland, A., and Spinrad, T. L. (1998). Parental socialization of emotion. *Psychological Inquiry*, 9(4):241–273.

Ekman, P. and Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1):56–75.

Enticott, P. G., Johnston, P. J., Herring, S. E., Hoy, K. E., and Fitzgerald, P. B. (2008). Mirror neuron activation is associated with facial emotion processing. *Neuropsychologia*, 46(11):2851–2854.

Ermer, E., Kahn, R. E., Salovey, P., and Kiehl, K. A. (2012). Emotional intelligence in incarcerated men with psychopathic traits. *Journal of Personality and Social Psychology*, 103(1):194.

Feinman, S. (1982). Social referencing in infancy. *Merrill-Palmer Quarterly*, pages 445–470.

Fitousi, D. and Wenger, M. J. (2013). Variants of independence in the perception of facial identity and expression. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1):133.

Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2):3–71.

Gallagher, S. (2015). Reuse and body-formatted representations in simulation theory. *Cognitive Systems Research*, 34:35–43.

Gallagher, S. and Hutto, D. (2008). Understanding others through primary interaction and narrative practice. In Zlatev, J., Racine, T. P., Sinha, C., and Itkonen, E., editors, *The Shared Mind: Perspectives on Intersubjectivity*, volume 12, pages 17–38. John Benjamins Publishing.

Gallese, V. (2001). The 'shared manifold' hypothesis. From mirror neurons to empathy. *Journal of Consciousness Studies*, 8(5-6):33–50.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, 4(1):23–48.

Gallese, V. (2016). Finding the body in the brain. From simulation theory to embodied simulation. In McLaughlin, B. P. and Kornblith, H., editors, *Goldman and His Critics*, pages 299–314. John Wiley & Sons.

Gallese, V. and Caruana, F. (2016). Embodied simulation: Beyond the expression/experience dualism of emotions. *Trends in Cognitive Sciences*.

Gallese, V. and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12):493–501.

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8(9):396–403.

Gallese, V., Rochat, M., Cossu, G., and Sinigaglia, C. (2009). Motor cognition and its role in the phylogeny and ontogeny of action understanding. *Developmental Psychology*, 45(1):103.

Gallese, V. and Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, 15(11):512–519.

Gazzaniga, M. S. (2004). *The cognitive neurosciences*. MIT press.

Gentsch, A., Weber, A., Synofzik, M., Vosgerau, G., and Schütz-Bosbach, S. (2016). Towards a common framework of grounded action cognition: Relating motor control, perception and cognition. *Cognition*, 146:81–89.

Glenberg, A. M. and Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3):558–565.

Goldman, A. and de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences*, 13(4):154–159.

Goldman, A. I. (1992). In defense of the simulation theory. *Mind & Language*, 7(1-2):104–119.

Goldman, A. I. (1993). Philosophical applications of cognitive science.

Goldman, A. I. (2012). A moderate approach to embodied cognitive science. *Review of Philosophy and Psychology*, 3(1):71–88.

Goldman, A. I. (2013). The bodily formats approach to embodied cognition. *Current Controversies in Philosophy of Mind*, page 91.

Goldman, A. I. and Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition*, 94(3):193–213.

Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological Bulletin*, 141(6):1266.

Grossmann, T. and Vaish, A. (2009). Reading faces in infancy: Developing a multi-level analysis of social stimulus. In Striano, T. and Reid, V., editors, *Social Cognition: Development, Neuroscience and Autism*. Blackwell Publishing, Oxford, UK.

Halliday, M. A. K. (1975). *Learning How to Mean–Explorations in the Development of Language*. Edward Arnold (Publishers) Ltd.

Hamlin, J. K., Wynn, K., and Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, 450(7169):557–559.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6):223–233.

Hess, U., Adams, R. B., and Kleck, R. E. (2009). The face is not an empty canvas: How facial expressions interact with facial appearance. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1535):3497–3504.

Iacoboni, M. (2009). Do adolescents simulate? Developmental studies of the human mirror neuron system. In Striano, T. and Reid, V., editors, *Social Cognition: Development, Neuroscience and Autism*. Blackwell Publishing, Oxford, UK.

Johnson, M. H., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1):1–19.

Johnson, M. H., Senju, A., and Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience & Biobehavioral Reviews*, 50:169–179.

Kennedy, D. P. and Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends in Cognitive Sciences*, 16(11):559–572.

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, 8(3):159–166.

Klinnert, M. D., Emde, R. N., Butterfield, P., and Campos, J. J. (1986). Social referencing: The infant's use of emotional signals from a friendly adult with mother present. *Developmental Psychology*, 22(4):427.

Knapp, M. L., Hall, J. A., and Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.

Lambie, J. A. and Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: A theoretical framework. *Psychological Review*, 109(2):219.

Langton, S. R., Watt, R. J., and Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, 4(2):50–59.

Leitan, N. D. and Chaffey, L. (2014). Embodied cognition and its applications: A brief review. *Sensoria: A Journal of Mind, Brain & Culture*, 10(1):3–10.

Leo, I., Angeli, V., Lunghi, M., Dalla Barba, B., and Simion, F. (2018). Newborns' face recognition: The role of facial movement. *Infancy*, 23(1):45–60.

Lewis, M. (2008). The emergence of human emotions. In *Handbook of Emotions*, pages 304–319. Citeseer.

Lewkowicz, D. J. and Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, 13(11):470–478.

Lipps, T. (1935). Empathy, inner imitation, and sense-feelings. *A Modern Book of Aesthetics*, pages 291–304.

Maranesi, M., Bonini, L., and Fogassi, L. (2014). Cortical processing of object affordances for self and others' action. *Frontiers in Psychology*, 5.

Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., and Frank, M. (2008). Facial expressions of emotion. *Handbook of Emotions*, 3:211–234.

Matsumoto, D. and Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of Personality and Social Psychology*, 96(1):1.

McKone, E. and Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review*, 16(5):778–797.

Meltzoff, A. N. (2007). 'Like me': A foundation for social cognition. *Developmental Science*, 10(1):126–134.

Meltzoff, A. N. and Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, pages 702–709.

Meltzoff, A. N. and Moore, M. K. (1992). Early imitation within a functional framework: The importance of person identity, movement, and development. *Infant Behavior and Development*, 15(4):479–505.

Meltzoff, A. N., Murray, L., Simpson, E., Heimann, M., Nagy, E., Nadel, J., Pedersen, E. J., Brooks, R., Messinger, D. S., Pascalis, L. D., et al. (2017). Re-examination of oostenbroek et al.(2016): evidence for neonatal imitation of tongue protrusion. *Developmental Science*.

Morton, J. and Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review*, 98(2):164.

Neisser, U. (1967). Cognitive psychology. *Memory*, 11(2).

Neisser, U. (2014). *Cognitive Psychology: Classic Edition*. Psychology Press.

Nelson, C. A. (2001). The development and neural bases of face recognition. In *Infant and Child Development*, volume 10, pages 3–18. Wiley Online Library.

Niedenthal, P., Wood, A., and Rychlowska, M. (2014). Embodied emotion concepts. *The Routledge Handbook of Embodied Cognition*, pages 240–249.

Nummenmaa, L., Hirvonen, J., Parkkola, R., and Hietanen, J. K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *Neuroimage*, 43(3):571–580.

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., Clark, S., and Slaughter, V. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology*, 26(10):1334–1338.

O'Toole, A. J., Roark, D. A., and Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, 6(6):261–266.

Oztop, E., Kawato, M., and Arbib, M. (2006). Mirror neurons and imitation: A computationally guided review. *Neural Networks*, 19(3):254–271.

O'Toole, A. and Roark, D. (2010). Memory for moving faces: The interplay of two recognition systems. *Dynamic faces: Insights from experiments and computation*, pages 15–29.

Palermo, R. and Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*, 45(1):75–92.

Pascalis, O., de Haan, M., and Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, 296(5571):1321–1323.

Pascalis, O., Scott, L. S., Kelly, D., Shannon, R., Nicholson, E., Coleman, M., and Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14):5297–5300.

Pell, P. J. and Richards, A. (2013). Overlapping facial expression representations are identity-dependent. *Vision Research*, 79:1–7.

Piepers, D. and Robbins, R. (2012). A review and clarification of the terms "holistic","configural", and "relational" in the face perception literature. *Frontiers in Psychology*, 3:559.

Prather, J. F., Peters, S., Nowicki, S., and Mooney, R. (2008). Precise auditory–vocal mirroring in neurons for learned vocal communication. *Nature*, 451(7176):305–310.

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(04):515–526.

Rauscher, F. H., Krauss, R. M., and Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7(4):226–231.

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neuroscience & Biobehavioral Reviews*, 32(1):123–142.

Reissland, N. (2013). *The Development of Emotional Intelligence: A Case Study*. Routledge.

Ricciardi, E., Bonino, D., Sani, L., Vecchi, T., Guazzelli, M., Haxby, J. V., Fadiga, L., and Pietrini, P. (2009). Do we really need vision? how blind people "see" the actions of others. *Journal of Neuroscience*, 29(31):9719–9724.

Rilling, J. K. and Young, L. J. (2014). The biology of mammalian parenting and its effect on offspring social development. *Science*, 345(6198):771–776.

Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., and Fazio, F. (1996). Localization of grasp representations in humans by PET:1. Observation versus execution. *Experimental Brain Research*, 111(2):246–252.

Robbins, R. and McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, 103(1):34–79.

Rochat, P., Passos-Ferreira, C., and Salem, P. (2009). Three levels of intersubjectivity in early development. In *Enacting Intersubjectivity. Paving the Way for a Dialogue Between Cognitive Science, Social Cognition and Neuroscience*, pages 173–190. Larioprint Como.

Salovey, P. and Mayer, J. D. (1989). Emotional intelligence. *Imagination, Cognition and Personality*, 9(3):185–211.

Sangrigoli, S. and De Schonen, S. (2004). Recognition of own-race and other-race faces by three-month-old infants. *Journal of Child Psychology and Psychiatry*, 45(7):1219–1227.

Sangrigoli, S., Pallier, C., Argenti, A.-M., Ventureyra, V., and De Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, 16(6):440–444.

Scherer, K. R. (1999). Appraisal theory. *Handbook of Cognition and Emotion*, pages 637–663.

Scherer, K. R. and Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, 7(1):113.

Scott, L. S., Pascalis, O., and Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, 16(4):197–201.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(03):417–424.

Shapiro, L. (2010). *Embodied Cognition*. Routledge.

Shouse, E. (2005). Feeling, emotion, affect. *M/c journal*, 8(6):26.

Simion, F., Cassia, V. M., Turati, C., and Valenza, E. (2003). Non-specific perceptual biases at the origins of face processing.

Simion, F. and Di Giorgio, E. (2015). Face perception and processing in early infancy: Inborn predispositions and developmental changes. *Frontiers in Psychology*, 6.

Simion, F., Macchi Cassia, V., Turati, C., and Valenza, E. (2001). The origins of face perception: specific versus non-specific mechanisms. *Infant and Child Development*, 10(1-2):59–65.

Simion, F., Turati, C., Valenza, E., and Leo, I. (2006). The emergence of cognitive specialization in infancy: The case of face preference. *Attention and Performance XXI, Processes of change in brain and cognitive development*, pages 189–208.

Simpson, E. A., Maylott, S. E., Heimann, M., Subiaul, F., Paukner, A., Suomi, S. J., and Ferrari, P. F. (2016). Commentary on "Animal studies help clarify misunderstandings about neonatal imitation" by Keven and Akins.

Singer, T. and Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156(1):81–96.

Trevarthen, C. (2006). The concept and foundations of infant intersubjectivity. In Bråten, S., editor, *Intersubjective Communication and Emotion in Early Ontogeny*, pages 15–46. Cambridge University Press.

Tsao, D. Y. and Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, 31:411.

Turati, C. (2004). Why faces are not special to newborns: An alternative account of the face preference. *Current Directions in Psychological Science*, 13(1):5–8.

Turati, C., Simion, F., Milani, I., and Umiltà, C. (2002). Newborns' preference for faces: what is crucial? *Developmental Psychology*, 38(6):875–881.

Tversky, B. (2009). Spatial cognition: Embodied and situated.

Varela, F., Rosch, E., and Thompson, E. (1991). The embodied mind. *Cognitive Science and Human Experience*.

Viola Macchi, C., Turati, C., and Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science*, 15(6):379–383.

Vivanti, G. and Hamilton, A. (2014). Imitation in autism spectrum disorders. In Volkmar, F., Rogers, S., Paul, R., and Pelphrey, K. A., editors, *Handbook of Autism and Pervasive Developmental Disorders*. John Wiley & Sons Inc, 4 edition.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., and van Leeuwen, C. (2012). A century of gestalt psychology in visual perception: Ii. conceptual and theoretical foundations. *Psychological Bulletin*, 138(6):1218.

Wilkinson, N., Paikan, A., Gredebäck, G., Rea, F., and Metta, G. (2014). Staring us in the face? An embodied theory of innate face preference. *Developmental Science*, 17(6):809–825.

Wilson, R. A. and Foglia, L. (2011). Embodied cognition. *Stanford Encyclopedia of Philosophy*.

Winkielman, P., Niedenthal, P., Wielgosz, J., Eelen, J., and Kavanagh, L. C. (2015). Embodiment of cognition and emotion. In Wagner, D. and Heatherton, T., editors, *APA Handbook of Personality and Social Psychology, Volume 1: Attitudes and Social Cognition*, volume 1, pages 151–175. American Psychological Association Washington, DC.

Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y.-H. P., and Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology*, 5.

Yankouskaya, A., Humphreys, G. W., and Rotshtein, P. (2014). The processing of facial identity and expression is interactive, but dependent on task and experience. *Frontiers in human neuroscience*, 8.

Zarbatany, L. and Lamb, M. E. (1985). Social referencing as a function of information source: Mothers versus strangers. *Infant Behavior and Development*, 8(1):25–33.

To study the abnormal is the best way of understanding the normal.

— William James —

3

Hypotheses of Face Processing Embodiment[1]

Chapter 2 provided the necessary background for the development of the computational tools and theories necessary to validate the thesis argument introduced in Chapter 1. As a secondary contribution, this dissertation aims to provide preliminary computational evidence in support of embodied cognition theories. Although this secondary contribution is not a necessary pre-condition for the validation of the thesis argument proposed in this dissertation, its validation would be of significant value for advancing embodied cognition research, thus considerably enriching the present work.

For this reason, in Section 2.3, I provided the necessary background on embodied cognition research and, in Section 2.3.2, I offered a list of the methodologies available in embodied cognition research to assess the validity of embodied cognition theses. Among the available options, I identified the constitution hypothesis as the most favourable one to evaluate the plausibility of the embodied cognition hypotheses intended in this work. Recall that in order to validate a hypothesis

[1]Part of the contents of this chapter have been published in *"Vitale, J., Williams, M.-A., and Johnston, B. (2014). Socially impaired robots: Human social disorders and robots' socio-emotional intelligence. In 6th International Conference on Social Robotics , pages 350–359".*

under the constitution methodology it is necessary to demonstrate that two processes **X** and **Y** are not only causally related but that the embodied process **X** is a central constituent of the cognitive process **Y**, which would fail or be something else without (or with dysfunctional) process **X**.

In this dissertation, I choose face processing as topic to develop and test my computational theories, since it requires nontrivial cognitive capabilities that are vital for social cognition development. I also proposed to provide a computational understanding of mirroring mechanisms, which I suggested is at the core of social cognition embodiment. Therefore, in order to investigate social cognition embodiment under the constitution hypothesis methodology, in this work, I aim to provide computational evidence suggesting that without or with dysfunctional embodied mechanisms (*i.e.* **X**), face processing, and consequently social cognition (*i.e.* **Y**), would fail. The objective of this chapter is to provide the necessary hypotheses justifying the evaluations used in Chapter 6 able to produce the computational evidence favouring embodied cognition theories and validating the aimed secondary contributions of this dissertation.

In order to derive these hypotheses, I will offer a review of some of the most investigated clinical populations affected by social disorders, namely autism, schizophrenia and psychopathy. The target of this chapter is to offer a novel reading of the available clinical literature on the deficits traditionally associated with these social disorders. I will argue that social dysfunctions exhibited by these clinical populations can find an explanation in alterations of embodied processes. Deficient embodied processes can explain the face processing impairments observed in the considered clinical populations. Therefore, these insights will serve as foundations to design the experiments offered in Chapter 6 employed to demonstrate that embodied processes can indeed constitute a crucial capability of social cognition, such as face recognition.

I will start this chapter by providing the definitions of the considered social disorders and the traditional understanding of their exhibited social impairments (Sections 3.1, 3.2 and 3.3). In Section 3.4, I will suggest how the reported deficits can link to embodied cognition research. In Section 3.5, I will discuss the face processing deficits observed in these clinical subjects and, in Section 3.6, I will finally link this evidence to the proposed embodied understanding of the considered social disorders. By doing this, I will introduce the desired hypotheses driving the validation of the secondary contributions offered by this dissertation.

3.1 Autism Spectrum Disorders

Individuals affected by Autism Spectrum Disorders are characterised by deficits in three main areas: (i) communication, (ii) social interaction, and (iii) restrictive and repetitive behaviours and interests (American Psychiatric Association, 2000). Autism is commonly divided into two main classes: 'high-functioning' and 'low-functioning' autism. For the sake of simplicity, in this chapter, I will not consider this distinction while reviewing the literature[2], and instead, I will refer to autistic individuals in general by considering the deficits typically attributed to this disorder.

Social cognition is profoundly impaired in autistic population (Farrow and Woodruff, 2007). This population has reduced Theory of Mind capabilities (Baron-Cohen et al., 1985) and, therefore, compared with control subjects, autistic individuals are poorer at reasoning about what others think, know, or believe, recognising emotional expressions and gestures, and making social attributions and judgements (Bachevalier and Loveland, 2006).

The observed deficits in emotion recognition/responding lead often to an impoverished facial affect (Brothers, 2002). For this reason, autistic individuals are perceived by laypersons as unable to feel emotions (Yirmiya et al., 1989). Nevertheless, studies with electrodermal responses and self-report measures suggest that autistic individuals have appropriate emotional responsiveness to others (Dziobek et al., 2008). Hence, people affected by autism seem to be able to experience at least normal phenomenological states underlying emotional reactions.

Autistic population is also characterised by dysfunctions in imitative behaviour. Decety and Moriguchi (2007) suggest that this population exhibits deficient automatic mimicry, although it performs well during voluntary imitation. In addition, autistic children have deficits in imitating the use of objects, facial gestures and vocalisations (Gallese et al., 2009). Thus, Gallese et al. (2009) propose that these problems are due to their inability to establish a motor equivalence between demonstrator and imitator.

It is widely accepted in autism literature that one of the earliest signs of autism is a lack of sensitivity to social cues (Grossmann, 2015), and it has been suggested that autism deficits may be explained by a reduced visual interest to the available social information (Birmingham and Kingstone, 2009). Indeed, this clinical population exhibits poor eye contact, thus lacking engagement during

[2]However, most of the available literature in Autism Spectrum Disorders considers high-functioning autistic subjects.

face-to-face interactions and showing disinterest to other people (Brothers, 2002; Burack et al., 2012; Farrow and Woodruff, 2007). Sasson (2006) found that this population look less at the eyes relatively to control participants. Similar findings have been found in many other studies (Elsabbagh et al., 2012; Falck-Ytter et al., 2013; Guillon et al., 2014; Klin et al., 2002; Navab et al., 2012).

Therefore, traditional literature in autism suggests that these individuals may stumble at this first perceptual step and consequently limit their social interactions and prevent the correct development of social cognition (Burack et al., 2012; Grossmann, 2015; Vivanti and Hamilton, 2014).

Accordingly, one of the most influential hypotheses in autism research is the *'social orienting hypothesis'* proposed by Dawson et al. (1998). According to this hypothesis, autistic individuals present deficient social rewarding processes, namely mechanisms facilitating pro-social behaviours by activating internal positive states during perception of social stimuli (*e.g.* faces, eyes, *etc.*) or episodes of social exchanges (Bons et al., 2013). These deficits may prevent this clinical population to focus the attention on social information necessary to develop social cognition capabilities promoting mind-reading skills (Dawson et al., 1998; Izuma et al., 2011). Indeed, this population shows abnormal amygdala activation when fixating the eyes region on face stimuli (Birmingham et al., 2011), a brain region often associated with emotional processing and likely providing social rewards driving visual attention (Birmingham et al., 2011).

In summary, autistic individuals exhibits deficiencies in the ability to understand others' mental state, which is vital for being proficient in social cognition. These deficiencies includes impairments in emotion recognition and expression, imitative behaviour and visual attention to social stimuli. Indeed, one of the earliest signs of autism is a lack of sensitivity to social cues. The social orienting hypothesis suggests that these individuals may have impaired social rewarding mechanisms, thus not orienting their attention to crucial social information and consequently leading to impaired social cognition development.

3.2 Schizophrenia

Schizophrenia is a severe psychiatric spectrum of disorders altering emotional, cognitive, and social functions (Parasuraman, 1998). In particular, significant impairment in social functioning is considered one diagnostic characteristic of schizophrenia (American Psychiatric Association, 2000). Such impairment can

have serious impacts on social relationships (Kennedy and Adolphs, 2012). Furthermore, schizophrenia disorder is characterised by delusions and hallucinations (American Psychiatric Association, 2000).

Schizophrenic individuals exhibit dysfunctions on general abilities in social cognition (Savla et al., 2012), with particular deficiencies in emotional empathy (Farrow and Woodruff, 2007; Sparks et al., 2010). Hence, some current models of schizophrenia suggest that this disorder can be understood as a deficit in representing others' mental states (Brüne, 2005) and of 'resonating' to others' emotional states (Farrow and Woodruff, 2007). Accordingly, schizophrenic individuals exhibit blunted feeling and usually have inappropriate affective responses in social situations (Farrow and Woodruff, 2007). This population shows abnormalities in skin conductance responses, and it mostly respond with negative affect (*e.g.* depression) (Farrow and Woodruff, 2007).

Schizophrenia population performs poorly on nearly all tests of sensory and cognitive vigilance and some studies also demonstrate deficits in selective attention (Parasuraman, 1998). It has also been shown that there are abnormalities in eye movements during the scanning of emotional facial expressions (Streit et al., 1997). Similarly to autistic people, schizophrenic individuals look less at the eye region of the face (Burack et al., 2012; Farrow and Woodruff, 2007). Again, in a similar way as in autistic population, schizophrenic patients show partial gaze avoidance to human faces, whereas they do not avoid gaze when they look to non-human faces (Williams, 1974). Sasson et al. (2007) suggest that, although both autism and schizophrenia may share impaired perceptual processes (*e.g.* avoidance of eye region), schizophrenic individuals also show a *temporal delay* in orienting the gaze to informative social information, whereas autistic individuals fail more in spatially orienting the gaze to similar social cues.

However, a study by Tempesta et al. (2014) suggests that basic visual processing may be preserved in schizophrenia population and that sustained attention over emotional stimuli is impaired instead. The progression of emotional processing in the mind of schizophrenic patients have been compared to healthy subjects using a method known as event-related potential (Horan et al., 2010). Early processing (first 50ms after stimulus) of emotions is of similar intensity in both populations but in late processing (after 200ms) the intensity is reduced in schizophrenic patients. This suggests that, although schizophrenia individuals may have functioning emotional response mechanisms, they exhibit disruption of a later component of sustained attention over the observed emotional stimuli (Horan et al., 2010).

This alternative understanding can explain the temporal delays and perceptual deficits observed in this population during social stimuli processing tasks (Tempesta et al., 2014). In addition, schizophrenic subjects unable to sustain attention over the observed social information cannot maintain this information in the working memory over time and consequently affect later mechanisms of behaviour regulation (Horan et al., 2010).

In summary, schizophrenia patients show relevant social dysfunctions. Similarly to autistic people, they present perceptual deficits. However, contrary to autism population, these deficits can be explained by impaired sustained attention. Deficits in sustained attention can prevent the selection of appropriate behavioural responses during social interactions and the interpretation of others' actions, thus leading to impaired social cognition.

3.3 Psychopathy

The World Health Organization (1992) classifies psychopathy as a form of antisocial (or dissocial) personality disorder. Characteristics of such disorder are: (i) callous unconcern for the feelings of others; (ii) incapacity to maintain enduring relationships, though having no difficulty in establishing them; (iii) very low tolerance to frustration and a low threshold for discharge of aggression, including violence; (iv) incapacity to experience guilt or to profit from experience, particularly punishment; (v) marked proneness to blame others, or to offer plausible rationalisations, for the behaviour that has brought the patient into conflict with society (World Health Organization, 1992).

In contradistinction to what is commonly believed, psychopathic individuals do not always show violent and criminal behaviour. Although people affected by psychopathy exhibit impairments in perceiving guilt and learn from punishment, this lack of regulatory mechanisms can lead to a vast spectrum of behaviours, which depends on other factors such as sex, age, dominance and social role (Farrow and Woodruff, 2007).

Contrary to typical autistic and schizophrenic patients, people affected by psychopathy do not exhibit abnormal levels of intelligence (Ermer et al., 2012). Accordingly, psychopathic individuals successfully pass Theory of Mind tasks, and currently, there is no evidence of impaired cognitive empathy ability (Farrow and Woodruff, 2007). However, due to deficits in processing emotions, psychopath individuals may have difficulties in internally monitoring emotional states as-

sociated to the observed stimuli. Therefore, this population relies exclusively on cognitive appraisal mechanisms (see Definition 2.12 on page 32) in order to fulfil a mind-reading task, without integrating the necessary peripheral emotional information (Decety and Moriguchi, 2007).

Indeed, psychopathy population is suggested to be mainly characterised by a lack of 'emotional empathy' (Farrow and Woodruff, 2007) (see Definition 2.11 on page 32); individuals affected by psychopathy have a reduced ability to feel other people's emotional state, especially sadness and fear (Decety and Moriguchi, 2007). Psychopathic subjects have deficits in moral emotions such as remorse and guilt, and they are usually indifferent to shaming and embarrassing situations (Ermer et al., 2012). This dysfunction is at the heart of the disorder; in fact, individuals unable to 'feel' sadness or fear of others are individuals with problems to conforming to social norms (Farrow and Woodruff, 2007).

Crucially, the impairments in emotional processing in this population are frequently associated with abnormal amygdala activation and structure, widely documented in these individuals (Coplan and Goldie, 2011). Accordingly, the amygdala is a brain area suggested to be particularly vital for emotional processing capabilities, although there is still no agreement on which specific function the amygdala covers in emotional processing (Adolphs, 2002).

One suggestion is that the amygdala plays a role in guiding the gaze to the eye region of facial stimuli, thus facilitating the recognition of emotional expressions, in particular, fearful faces (Dawel et al., 2012). Although psychopaths show impairments in fixating over the eye region of the face, this does not explain other deficits in emotional processing involving other modalities, such as vocal, postural and symbolic (*i.e.* emotional words) stimuli (Dawel et al., 2012).

Therefore, at least two other explanations were suggested (Dawel et al., 2012). The first is that the amygdala may be involved in directing attention to socially relevant cues in general. Thus, the eyes are an example of a social relevant cue, but also speech tone and biological motion. The second potential explanation is that amygdala may provide the *experience* of emotion, and for this reason, its dysfunction may contribute to multimodal integration deficits between sensory-motor cues and visceral peripheral signals (Dawel et al., 2012). The last interpretation is likely to be more plausible since it can also explain the lack of empathy characterising this population.

In summary, psychopaths have no major deficiencies in attributing mental states to others. This clinical population can easily use cognitive appraisal mechanisms to infer others' mental states. However, it exhibits severe dysfunctions in emotional

processing. In particular, it has been suggested that this population may not be able to integrate peripheral emotional information with the rest of available sensory-motor information. This lack of empathy towards other individuals lead to difficulties in socialising and conforming to social norms.

3.4 Embodied Understanding of Social Disorders

In Sections 3.1, 3.2 and 3.3, I provided the definition of autism, schizophrenia and psychopathy. I provided a review of the available literature in clinical studies and offered the causes suggested to be at the origin of the observed social disorders.

In particular, an influential hypothesis in autism research, the social orienting hypothesis (Dawson et al., 1998), suggests that this disorder may derive by poor selective attention to social stimuli that in turn leads to insufficient social information during the developmental process of the subject, thus affecting social cognition development. On the contrary, individuals affected by schizophrenia exhibit dysfunctional sustained attention over emotional stimuli, thus resulting unable to sustain the gathered emotional information over time and consequently selecting inappropriate social behaviours (Horan et al., 2010; Tempesta et al., 2014). Finally, psychopaths are not affected by perceptual or cognitive impairments like autistic and schizophrenic subjects, but they rather present significant emotional processing dysfunctions (Decety and Moriguchi, 2007; Farrow and Woodruff, 2007). It has been suggested that individuals affected by psychopathy can hardly monitor peripheral emotional signals, thus being unable to integrate this emotional information with other available cues (Dawel et al., 2012). This in turn prevents emotional empathy in this population, leading to their well known antisocial behaviour (Farrow and Woodruff, 2007).

In the following sections, I will discuss how the deficits identified in these clinical populations can explain alterations of the embodiment mechanisms plausibly at the core of social cognition capabilities. This analysis will motivate the hypotheses introduced in the remainder of this chapter and necessary to design an appropriate methodology to validate embodied cognition theories.

3.4.1 Embodied Understanding of Autism's Deficits

One of the previously discussed dysfunctions in autistic individuals is an impaired imitative behaviour (Decety and Moriguchi, 2007). In Section 2.1.6 (page 34), I discussed that imitative behaviour is particularly crucial for the correct development

of social cognition abilities. In this section, I offer plausible connections between the observed dysfunctions in autism imitative behaviour and the discussed social orienting hypothesis, thus suggesting an embodied understanding of autism's deficits.

Imitative behaviour requires a complex set of skills in order to be correctly accomplished. For example, a neuropsychological model by Vivanti and Hamilton (2014) suggests imitation process to starts with a visual encoding of the observed action, which can be matched to a familiar action by accessing an action knowledge representation. This in turn can provide input to the motor system in order to enact unconscious or conscious imitative mechanisms of that action and consequently engage the associated motor plan, goals, underlying intentions and beliefs. Importantly, when the action is not familiar, there is no way to match it with motor schema stored in the action knowledge representations (Rizzolatti and Fabbri-Destro, 2010). Thus, in this case, the visual representation of the observed action must be mapped *directly* to the motor system (Vivanti and Hamilton, 2014).

Therefore, in both the scenarios, the subject needs an intact motor system in order to reach an interpretation of the observed action. The core mechanism that can make this motor matching possible is mirroring, as discussed in Section 2.1.4 (page 29). Gallese et al. (2009), Iacoboni and Dapretto (2006) investigated electrophysiological activations in autistic subjects. Their findings suggest that autistic people fail in imitative behaviour because of underlying impairments in mirror neurons functioning. Furthermore, studies employing Electroencephalography (EEG) and Transcranial Magnetic Stimulation (TMS) (Oberman et al., 2005; Théoret et al., 2005) show that autistic individuals suffer from deficits in action inner simulation.

However, as this population performs well at least on voluntary imitation (Decety and Moriguchi, 2007), the reduced activation of mirror neurons observed in these individuals during the observation of others' motor actions is plausibly due to other causes. Indeed, linking to the previously discussed social orienting hypothesis, reduced activation of mirror neurons in autistic individuals may not be necessarily due to mirror neurons functional impairments, but rather due to a lack of attention to relevant social cues, such as the eyes, emotional motor sequences or faces (Bons et al., 2013). By not paying attention to others' motor actions, people affected by autism cannot sufficiently activate motor representations of the observed behaviour via mirroring mechanisms. These poor activations throughout their development may prevent the acquisition of correct sensory-motor mappings,

thus impairing automatic and unconscious mirroring processes when interacting with people (Bons et al., 2013).

Accordingly, it has been proposed that the origins of mirror neurons function may derive from domain-general associative learning processes (Cook et al., 2014). Although it is conceivable that primates born equipped with functioning mirror neurons, this might be limited to mirror elementary actions. Through developmental experience, mirroring ability is refined to map more complex actions and to represent the interactions between sensory, motor and visceral inputs via associative learning mechanisms (Berlucchi and Aglioti, 1997). Therefore, if specific intentions, beliefs, goals or emotions are associated with motor acts not included in the repertoire of an autistic individual, this subject cannot mirror them and understand the associated mental states (De Vignemont, 2009; Rizzolatti and Fabbri-Destro, 2010). In other words, autistic individuals can still have access to the 'outside' representation of the motor act, which unfortunately does not correctly activate the 'inside' representation of the observed action (*i.e.* the mental states associated to specific motor act), because of their poorly developed mirror neurons, vital for providing information about the intention of the observed motor act (Rizzolatti and Sinigaglia, 2010). Therefore, autistic people can still perform well in imitating simple motor acts (like most of the ones used in autism behavioural studies), although presenting deficits when imitating opaque and more complex actions (Vivanti and Hamilton, 2014).

In summary, although selective attention dysfunctions in autism can plausibly explain the development of this disorder (Dawson et al., 1998), this explanation can be further enriched by an embodiment understanding. In fact, mirroring is crucial for the development of social cognition capabilities, and the poor attention to social cues in this population may prevent the correct development of sensory-motor associations in their mirror neuron system (Berlucchi and Aglioti, 1997; Cook et al., 2014). This in turn would lead to impairments in their motor organisation, including deficits in chaining motor acts into appropriate internal mental states necessary the understanding of the observed action (Rizzolatti and Sinigaglia, 2010).

3.4.2 Embodied Understanding of Schizophrenia's Deficits

Recent findings suggest that impairment of emotional resonance in patients with schizophrenia, as previously discussed in Section 3.2, are likely rooted to abnormalities in the mirror neurons mechanism (Sestito et al., 2015). As widely discussed

in Section 2.1.4 (page 29), mirroring is plausibly situated at the core of social cognition capabilities. Thus, in this section, I will offer the necessary connections between schizophrenia dysfunctions in sustained attention and mirroring mechanisms.

Self-monitoring of the own internal state is of particular importance to correctly process external stimuli (Matthias et al., 2009). As suggested in Section 3.2, schizophrenia subjects may fail in sustaining attention over internal states representing the perceived social stimuli (Tempesta et al., 2014), thus leading to their social dysfunctions. Gallese (2014) proposes the *bodily-self* as minimal notion of self available since early stages of life in order to facilitate social exchanges. This concept of self is built on top of the agent's motor system, thus being strictly related to functioning mirroring mechanisms (Gallese, 2014; Gallese et al., 2009). Therefore, in schizophrenia subjects this essential self-hood may be perturbed, unstable and oscillating, leading to alarming and alienating experiences potentially giving rise to their well-documented hallucinations and delusions (Gallese and Ferri, 2013; Sestito et al., 2015). Indeed, it has been suggested that self-awareness may have evolved for the specific purpose of allowing us to understand our own and others' behaviour (Decety and Sommerville, 2003). In order to enable such capability, one has to co-ordinate self and other mental representations, thus requiring specific executive function resources (Decety and Sommerville, 2003). Sustained attention and self-monitoring onto mirror neurons activations may contribute in maintaining such distinction (Gallese, 2014; Matthias et al., 2009).

The literature review by Billeke and Aboitiz (2013) shows that Superior Temporal Sulcus (STS) activity in schizophrenic population significantly differs from normal population. This brain structure is often implicated in mental state attribution, emotion, and *self-representation or agency* (Billeke and Aboitiz, 2013). Thus, people affected by schizophrenia and lacking a sense of agency may not be able to take beliefs as subjective (*i.e.* self) representations of the reality, and they instead equate them with reality itself. This may lead to difficulties in distinguishing between subjectivity and objectivity, thus using delusional convictions (*e.g.* control of the patient body by an alien) as a way to make sense of these false beliefs (Brüne, 2005).

Mirroring dysfunctions in schizophrenia, and the consequently disrupted bodily-self, may be due to impaired perceptual and attentive processes, leading to a fragmented monitoring of the bodily experience in these individuals (Sestito et al., 2015). Indeed, it has been suggested that an imbalance in monitoring sensory stimuli perception potentially leads to dysfunctions in automatic low-level ability

of multisensory integration (Sestito et al., 2015). Hence, the disruption of this integration process could result in a sense of discontinuity and inconsistency with the environment and the self (Sestito et al., 2015). This misalignment between mind and bodily-self induces disturbances of subjective experience, leading to symptoms such as depersonalisation, blurred boundaries, and a diminished sense of ownership and agency (Sestito et al., 2015). Accordingly, multisensory disintegration is argued to be implicated in the experiential emergence of self-disorders (Postmes et al., 2014), like schizophrenia (Gallese and Ferri, 2013). Self-disorders specifically refers to a disturbed sense of the basic self, leading to a variety of anomalous subjective experiences mostly affecting the sense of being a self-present, embodied subject immersed in the world (Sestito et al., 2015).

In summary, impaired sustained attention in schizophrenia subjects may lead to difficulties in focusing and maintaining attention over bodily representations of the processed social information, thus resulting in a disrupted sense of bodily-self. This deficit would eventually lead to a sense of discontinuity and inconsistency with the environment and the self, consequently leading to delusional convictions and the incapacity of attributing right mental states to others typically characterising this social disorder.

3.4.3 Embodiment Understanding of Psychopathy's Deficits

In Section 3.3, I suggested that this population, differently from autism and schizophrenia populations, does not suffer from perceptual or cognitive impairments, but it rather presents a significant impairment of emotional processing capabilities. In this section, I will provide findings connecting these emotional dysfunctions to embodied mechanisms, thus offering a novel embodiment understanding of psychopathy disorder.

The documented emotional dysfunctions in this clinical population could happen in two ways supporting the idea that embodiment is at the core of this disorder (Agnew et al., 2007). On the one hand, this population may be unable to correctly activate their mirror neuron system, thus preventing a correct interpretation of the observed social stimuli (*e.g.* facial expressions of emotion, biological motion, etc). On the other hand, the mirror neuron system may function sufficiently well to inform about the sensory-motor content of the processed social information, but not enough to integrate sensory-motor cues with visceral ones, thus preventing emotional empathy and leading to antisocial behaviour.

The first explanation is likely implausible since it does not take into consideration findings showing amygdala's dysfunctions in this clinical population (Coplan and Goldie, 2011; Dawel et al., 2012). Therefore, the second explanation results to be the most likely. In fact, the amygdala may play a crucial role in realising appropriate visceral cues associated with the perceived social stimulus, or it may have a more general role of regulating attention to the realised peripheral visceral cues (Dawel et al., 2012).

Accordingly, the *response modulation hypothesis* suggests that individuals with psychopathy are capable of normal emotional responses, but have difficulty in processing affective information when it is peripheral to their primary attentional focus, thus impairing its integration with other sensory-motor information (Newman and Lorenz, 2003).

Indeed, a study by Fecteau et al. (2008) showed that psychopaths are able to activate their mirror neuron system in order to create mental representations of the motor, affective and sensory state of the observed subject. More importantly, their mirror neuron system activation was shown to be even higher than non-psychopathic subjects. Therefore, Fecteau et al. (2008) suggested that this process is sufficient for psychopaths to gather the necessary information allowing them to manipulate and to exploit weaknesses in others. However, since this process includes maladaptive or absent emotional/affective responses in psychopathic population, this clinical population cannot employ emotional empathy mechanisms, thus having only a merely cognitive understanding of others (Farrow and Woodruff, 2007). Therefore, although people affected by psychopathy can activate appropriate covert viscero-motor representations Fecteau et al. (2008), psychopaths exhibit significant dysfunctions in realising corresponding physical reactions (Aniskiewicz, 1979; Blair, 1999; Blair et al., 1997). In particular, people affected by psychopathy show peculiar deficits in activating autonomic responses (*e.g.* electrodermal responses) when exposed to distress cues and threatening stimuli, such as fearful, sad and painful social signals (Blair, 1999; Blair et al., 1997).

In summary, psychopaths do not have cognitive impairments, but they suffer from severe dysfunctions in emotional processing. This clinical population does not show impairments of their mirroring mechanisms, thus being able to realise appropriate sensory-motor and affective mental representations of the observed social stimuli. Unfortunately, their limited attention to peripheral visceral reactions may prevent them to activate appropriate bodily responses assisting emotional understanding of others, especially when exposed to fearful, sad and painful

social stimuli. Therefore, individuals affected by psychopathy can only use purely cognitive mechanisms to interpret others' behaviour, and their lack of emotional empathy can potentially lead to their well known antisocial behaviour.

3.4.4 Summary

In the previous sections, I provided readings of the discussed social disorders from an embodiment standpoint. In particular, I reported the following main ideas:

- Autism is typically suggested to origin from dysfunctional perception of social stimuli. This dysfunction may reduce the amount of available social information necessary to correctly shape sensory-motor mapping capabilities exhibited by a functioning mirror neuron system and plausibly developed via associative learning episodes. Therefore, dysfunctional mirroring mechanisms would provide limited internal representations of the observed motor behaviours, thus leading to poor mind-reading capabilities;

- Sustained attention impairments play a crucial role in determining schizophrenia social disorders. This dysfunction may cause difficulties in correctly self-monitoring over time the bodily information provided by mirroring activity. This information provides a minimal concept of bodily-self necessary to distinguish between subjective and objective beliefs. Failing in correctly monitoring mirroring activation may lead to a disrupted bodily-self and result in delusional convictions and false beliefs about others;

- Psychopathy can be explained by generalised impairments of emotional processing. Although it has been suggested that psychopaths can correctly activate their mirror neuron system, they may not be able to attend peripheral visceral representations realised by this system. Therefore, this clinical population can access to the sensory-motor information associated with the processed social stimulus, but it cannot integrate this information with visceral cues, thus being unable to realise emotional empathy mechanisms necessary to prevent antisocial behaviours.

In the following section, I will review findings concerning face processing capabilities in these clinical populations. This further analysis is necessary to link the identified embodiment understandings to face processing impairments, and to motivate the hypotheses introduced in the remainder of this chapter.

3.5 Face Processing Impairments in Social Disorders

In this section, I will provide an overall review of findings concerning face processing impairments in the considered social disorders. This review will be necessary to determine differences among the considered clinical populations and to motivate the hypotheses introduced in the remainder of this chapter. I will first present findings concerning performance observed in the considered clinical populations during facial expression recognition tasks. Then, I will focus on the performance observed in clinical studies investigating facial identity recognition.

3.5.1 Deficits in Facial Expression Recognition Tasks

The literature provides a large amount of studies investigating facial expression recognition capabilities in autism, schizophrenia and psychopathy (Bauser et al., 2012; Dawel et al., 2012; Lozier et al., 2014; Marwick and Hall, 2008; McCleery et al., 2015; O'Brien et al., 2014; Savla et al., 2012). It is of particular importance to notice that most of these works present conflicting results, although inconsistencies found in the literature may be plausibly due to differences in the demographic factors of the experimental subjects and cognitive demands of the task (Harms et al., 2010; Pomarol-Clotet et al., 2010; Schönenberg et al., 2015). However, recent literature reviews and comprehensive meta-analyses help to identify the most salient impairments observed in facial expression recognition tasks in the considered clinical populations.

Autism

The literature in autism studies includes a recent review by Lozier et al. (2014) investigating autism's facial expression recognition deficits. This analysis shows that autistic individuals exhibit a strong and generalised deficit in facial expression recognition and that this effect increases throughout their development. Hence, the authors suggest that facial expression recognition ability may follow a distinct developmental trajectory in this clinical population compared to healthy subjects. In addition, Lozier et al. (2014) found that autistic individuals were less accurate in recognising facial expressions than control subjects for all six basic facial expressions of emotions. The proposed review is in agreement with another

recent meta-analysis including most previous works in autism facial expression recognition studies (Uljarevic and Hamilton, 2013).

Schizophrenia

Similarly to literature in autism, the available literature in schizophrenia presents two recent literature reviews on facial expression recognition deficits in this clinical population. A meta-analysis by Savla et al. (2012) investigated schizophrenia individuals' deficits on several measures of social cognition, included some thought to affect face processing strongly. The analysis suggests that schizophrenia population performs significantly worse than healthy subjects across all the considered measures of social cognition, with particularly significant effects on social perception, emotion perception, and emotional processing tasks. The review by Marwick and Hall (2008) specifically focused on face processing in schizophrenia population. This work shows an overall impairment in facial expression recognition in schizophrenia population compared to non-psychiatric controls.

Behavioural studies from facial expression recognition literature in schizophrenia are complemented with works observing event-related potentials amplitudes during recognition tasks. A recent meta-analysis by McCleery et al. (2015) reveals consistent and significant impairments in N170 and N250 event-related potentials components of schizophrenia population across the considered literature. These components are suggested to be associated with structural encoding of the face, with N170 component being associated mostly with encoding invariant configuration aspects of the face, and N250 component being associated mostly with changeable aspects of the face stimuli (McCleery et al., 2015). Although the relationship between N170 and N250 components is not well understood yet, schizophrenic individuals show clear impairments in event-related potentials components crucially assisting face processing capabilities.

Importantly, face processing deficits in schizophrenia cannot be explained by general impairments of their cognitive skills. In fact, Megreya (2016) recently investigated the accuracy of schizophrenic individuals in matching upright faces, inverted faces and non-face objects. Processing upright faces is thought to involve specific face processing mechanisms, whereas processing inverted faces or non-face objects rely on more general features matching mechanisms. Despite this work suggests impairments of schizophrenia population on all the three considered tasks compared to healthy controls, upright faces matching task produced a stronger effect than the other two tasks. Indeed, these results suggest that schizophrenia

subjects may exhibit a generalised impairment of their cognitive skills (Pomarol-Clotet et al., 2010), but their face processing capabilities are still significantly more accentuated.

Psychopathy

Differently from autistic and schizophrenic subjects, traditional literature in psychopathy suggests more significant impairments in fearful and sad facial expression processing (Dawel et al., 2012; Farrow and Woodruff, 2007). Accordingly, an influential meta-analysis by Marsh and Blair (2008) found evidence for specific deficits in recognising fearful emotions in psychopathic population.

However, a recent meta-analysis by Dawel et al. (2012) shows that facial expression recognition impairments in psychopaths are broader and generalised to all the basic expressions of emotion, although presenting more marked deficits in processing expressions of fear and sadness compared to other emotions. Thus, Dawel et al. (2012) suggest that psychopathy is associated with significant impairments for positive as well as negative emotions and that these deficits are present across both facial and vocal modalities.

One possible explanation for the contradictory results emerging from the literature in psychopathy may be due to the ease with which some expressions (*e.g.* happiness) are visually recognised (Farrow and Woodruff, 2007). However, Dolan and Fullam (2006) found that individuals with personality disorder compared with controls showed a deficit in sad and happy affect recognition even at 100% intensity, thus suggesting that the observed deficits cannot be likely attributed to task difficulty.

As an alternative explanation for the discrepancies observed in psychopathy literature, Contreras-Rodríguez et al. (2014) suggest the implicit (*i.e.* matching the same expression) as opposed to explicit (*i.e.* asking which facial expression is shown) emotional processing demands of the considered face matching task can potentially lead to different results. Importantly, despite a similar behavioural performance on the matching task, the authors still found significant differences in brain activation of the clinical population compared to control one. In particular, they found greater activation of areas involving visual and prefrontal cortices, whereas a decrease in activation of putative emotional brain regions and a general disruption of connections between emotional and cognitive components of the face-processing network (Contreras-Rodríguez et al., 2014).

Additional evidence for a more pervasive impairment of putative emotional brain areas in psychopaths comes from a study by Decety et al. (2014). This study shows that individuals scoring high on psychopathy exhibit consistently less activation than controls in brain regions relevant for emotional processing during the viewing of dynamic video clips of happy, sad, fearful, and pain expressions. Furthermore, Gordon et al. (2004) demonstrated that individuals scoring low on emotional-interpersonal features of psychopathy utilised areas of the brain typically associated with emotion interpretation and response when engaged in decoding facial expressions of affect, whereas high-scoring participants relied mostly on areas of the brain related to visual perception.

3.5.2 Deficits in Facial Identity Recognition Tasks

The available literature in clinical populations includes many works assessing facial identity recognition capabilities in people affected by social disorders, with the exception of psychopathy. Similarly to works in facial expression recognition, the literature in facial identity recognition studies includes conflicting results and inconsistencies (Bortolon et al., 2015; Weigelt et al., 2012). For this reason, in the following paragraphs, I will make use of recent literature reviews and meta-analyses able identify the most salient deficits observed in the considered clinical populations during facial identity recognition tasks.

Autism

A recent review by Weigelt et al. (2012) suggests that autistic population is typically impaired on standardised facial identity recognition tasks. In particular, face recognition deficits in this population can be found whenever sample and test stimuli are not present at the same time on the screen, and even if the test stimulus is presented immediately after the sample stimulus with no delay. However, no deficits are found when the face stimuli are presented simultaneously. These results may suggest memory dysfunctions in autistic people (Weigelt et al., 2012). However there is at least another plausible explanation.

Morin et al. (2015) found that autistic individuals exhibit deficits when recognising identities between different viewpoint conditions. They suggested that autistic subjects *"does not have a general impairment for facial identity discrimination per se, but are consistent with the hypothesis that facial identity discrimination is more difficult for participants with autism when (i) access to local cues is minimised,*

and/or (ii) an increased dependence on integrative analysis increases" (Morin et al., 2015, page 502).

Although Weigelt et al. (2012) did not find compelling differences between control and autistic subjects on how these population qualitatively process face stimuli, they found modest evidence for a possible impairment of three face markers in the autistic population. These markers are suggested to characterise intrinsic qualities of face processing mechanisms in healthy individuals, thus becoming a valid method to assess if autistic individuals' impairments in face recognition derive from qualitative differences in face processing mechanisms (Weigelt et al., 2012). Importantly, Weigelt et al. (2012) found modest evidence suggesting that autistic individuals may rely on representations of the face stimuli differing to the one traditionally suggested for healthy individuals, namely representations shaped as points of a multidimensional face-space, as proposed by Valentine et al. (2015). This framework suggests that face stimuli are represented as points of a multidimensional space able to encode invariant features of the face, thus facilitating the recognition of their identities (see Chapter 5 for more details).

As I will show later in Chapter 5 and Chapter 6, the face-space framework can be realised so to have a structure able to encode invariant (*e.g.* identity) and dynamic (*e.g.* facial expression, viewpoint, *etc.*) features at the same time in the same representation. This integral representation facilitates both facial identity and expression recognition, and it provides advantageously *compressed representations* of the observed stimuli (Vitale et al., 2016, but see Chapter 5 for more details). Thus, it may be possible that corrupted or absent face-space representations in autistic individuals lead to difficulties in realising compressed representations of the observed stimuli. These compressed representations may be more advantageous than using features based strategies since they can be more easily managed by the working memory (Brady et al., 2009). Thus, absent or deficient face-space representations may induce overall poor face recognition capabilities during memory demanding tasks. These representations differ in quality among humans due to their different experiences (Dennett et al., 2012). This diversity contributes to observed differences in face recognition performance among people (Dennett et al., 2012) and it can justify face recognition deficits observed in autistic populations. Indeed, Rhodes et al. (2014) suggest that faces are adaptively coded relative to visual norms that are updated by experience and that this coding is compromised in autistic individuals. Their results show that autistic population exhibits intact functional role for adaptive coding in

face recognition ability. Hence, the authors concluded that adaptive face-coding mechanisms are intact in autism, but less readily calibrated by experience.

Schizophrenia

Impairment in schizophrenia population has been shown in both identity matching/discrimination tasks and familiarity recognition tasks (Marwick and Hall, 2008). Accordingly, the recent literature review by Bortolon et al. (2015) suggests significant impairments of facial identity recognition in schizophrenia population; only four studies suggests similar performance for both controls and schizophrenia, in contrast to the majority of the available studies. Bortolon et al. (2015) indicate that these mixed results can be likely due to the clinical characteristics of the considered clinical samples, such as a shorter mean duration of illness.

In addition, the difficulty of the required task may be another factor determining the ability of clinical population in exhibiting normal or abnormal capabilities in facial identity recognition (Bortolon et al., 2015). In particular, tasks demanding high memory lead to major impairments in schizophrenia individuals' recognition accuracy (Bortolon et al., 2015).

However, brain activation in schizophrenia population during face identity recognition task has been observed to differ from healthy controls. In particular, their fusiform gyrus has been demonstrated to have a number of structural and functional abnormalities (Marwick and Hall, 2008). Moreover, it has been observed reduced neural activation relative to controls of the right fusiform gyrus while matching facial identity and emotion (Marwick and Hall, 2008). This brain area is strongly related with face processing tasks. For example, volume reduction of this brain area proportionally correlates to impairment at remembering face identities in this clinical population (Marwick and Hall, 2008).

Psychopathy

Surprisingly, from psychopathy literature it did not emerge any specific work on facial identity recognition[3]. Nevertheless, from the literature emerged at least two contributions in facial affect recognition in psychopathic and conduct disorders populations that can provide modest evidence for facial identity recognition performance in psychopaths.

[3]The search strategy involved querying PyscINFO and SCOPUS databases with terms related to psychopathy (*e.g.* psychopat*, conduct disorders, callous unemotional, antisocial, asocial) together with face recognition terms (*e.g.* face recognition, face identity, facial identity, identity recognition, face discrimination, face matching, face processing) and title/abstract screening.

Fairchild et al. (2010) investigated the ability of female adolescents with conduct disorder, including in the study individuals scoring high in psychopathic traits (Young Psychopathic traits Inventory ≥ 2.5) in recognising facial expressions of emotions. To assess subjects' basic visual processing abilities they submitted to clinical and healthy population the Benton's Facial Recognition Test (BFRT) (Benton, 1994), evaluating their capacity in matching identities from face stimuli. The authors did not find any group difference in facial identity recognition, similarly to a previous work investigating male adolescents with conduct early-onset conduct disorder (Fairchild et al., 2009). Sully et al. (2015) measured facial affect recognition performance in conduct disorders subjects and their unaffected relatives with respect to healthy population. Similarly to the previous work, they submitted the BFRT (Benton, 1994) to assess basic visual processing abilities for study inclusion. Accordingly, they did not find any significant differences among the considered groups; however, this study, differently from one of Fairchild et al. (2010), included only clinical subjects scoring low in psychopathic traits (Young Psychopathic traits Inventory < 2.5). However, Duchaine and Weidenfeld (2003) demonstrated how the BFRT is not an ideal tool for assessing deficits in face identity recognition. Therefore, the results of the discussed works should be interpreted with caution.

In conclusion, to my best knowledge there are no works specifically assessing facial identity recognition in psychopathic population. As previously discussed, there is only modest evidence in favour of spared facial identity recognition capability in people affected by conduct disorders, suggested being a predictor of adult psychopathy (Burke et al., 2007).

3.5.3 Summary

In the previous sections, I provided a review on face processing impairments in the considered clinical population. This literature review suggested that:

- Autistic people show a generalised impairment in facial expression recognition among all the six basic facial expressions of emotion. In addition, this clinical population shows impairments in facial identity recognition, in particular when the task does not present the target and test stimuli at the same time. This can be plausibly due to impaired face processing mechanisms unable to provide advantageously compressed representations of the observed face stimuli;

- Schizophrenia subjects exhibit a generalised impairment of facial expression recognition similarly to autistic individuals, although presenting additional significant cognitive impairments further compromising face processing capabilities. This population also shows significant impairments during facial identity recognition tasks compared to healthy subjects;

- Psychopaths are traditionally suggested to be impaired in processing fearful and sad emotional stimuli. In addition, neuroscience findings suggest that this population has impaired putative emotional brain areas processing emotional stimuli. However, a recent literature review by Dawel et al. (2012) shows that these individuals present a more generalised impairment over the recognition of all the six basic facial expressions of emotion. Differently from autism and schizophrenia populations, psychopaths seems to not be associated with facial identity recognition impairments, although this might be due to the lack of studies using appropriate tools.

In the following section, I will provide a general discussion concerning findings discussed in this chapter and introduce the hypotheses that will drive the design of the experiments in Chapter 6. These experiments will provide preliminary computational evidence in favour of embodied cognition theories, thus validating the secondary contribution of this dissertation and adding more value to the present work.

3.6 Hypotheses

In the previous sections of this chapter, I introduced some of the most investigated human social disorders, namely autism, schizophrenia, and psychopathy. Traditional understanding of these disorders suggests deficits in perceptual, cognitive and emotional capabilities. These disorders are classified as spectra since the affected individuals may exhibit mild to serious deficits. Determining the origins of these disorders becomes particularly challenging, since many perceptual, cognitive and emotional processes can interact in different ways and with varying levels of intensity, thus giving rise to a heterogeneous set of observed dysfunctions.

However, in this chapter, I provided a review of the available literature in clinical studies describing some of the deficits crucially characterising these disorders and suggested to be plausibly at their origins. For instance, it has been proposed that lack of attention to social cues may be at the source of autism development. Schizophrenia individuals seem to be particularly impaired in sustained attention

over emotional information. Finally, psychopaths have a lack of attention over peripheral emotional information, thus being unable to integrate the gathered social information with emotional cues.

In Section 3.4, I proposed that the identified dysfunctions likely impact on the correct functioning of embodied mechanisms, and this, in turn, affects social cognition capabilities. For example, the perceptual deficits of autism can provide poor social information during the subject's development, thus insufficiently shape desirable mirroring function at the basis of social cognition. In schizophrenia, the lack of sustained attention over mirroring activity can produce distorted representations of the bodily-self, thus contributing to their delusions and false beliefs about the mental state of others. Finally, psychopaths may not be able to attend over the visceral representation of the social stimulus realised by mirroring mechanisms, thus being incapable of 'feeling' like others and creating empathic connections promoting socially acceptable behaviours. I demonstrated the plausibility of these embodied standpoints by providing appropriate literature.

In summary, I suggest that:

Autism and schizophrenia dysfunctions prevent these populations from correctly retrieving sensory-motor information of observed actions provided by mirroring mechanisms. In fact, since the mirroring mechanisms in autistic individuals is not correctly developed, this clinical population may be unable to map the sensory information into appropriate motor representations. Similarly, schizophrenia patients may be able to process sensory information into motor representations, but their lack of sustained attention over mirroring activation may prevent them to integrate social information correctly over time, thus leading to altered sensory-motor representations;

Sensory-motor mapping capabilities in psychopaths are not impaired, but this information cannot be integrated with visceral information realised by embodied mechanisms. Therefore, psychopaths do not have access to visceral sensations associated with the observed social stimulus and they cannot emotionally understand others.

Given the present insights, I suggest the following hypothesis:

Hypothesis 3.1 *The information provided by embodied mechanisms, shaped as bodily formatted representations, lies on at least two dimensions: a sensory-motor*

dimension, and a visceral dimension. The sensory-motor dimension describes perceptual aspects of the perceived action, such as the motor potentials, the viewpoint and the pose. The visceral dimension specify emotional aspects of the perceived action, such as feelings and sensations associated with the perceived social stimuli.

As I explained in Section 2.1.6, face processing capabilities are vital to shaping social cognition capabilities. Therefore, after having suggested that dysfunctions in social cognition can be explained by dysfunctional embodied mechanism, I reviewed literature in clinical populations investigating face processing capabilities. This analysis suggested that whereas autism and schizophrenia populations are affected by dysfunctional facial expression and identity capabilities, psychopaths exhibit impairments in facial expression recognition only. Therefore, I suggest the following hypothesis:

Hypothesis 3.2 *Alterations of the sensory-motor embodied process of face stimuli significantly impair both facial identity and facial expression recognition capabilities. Alterations of the visceral embodied process of face stimuli significantly impair facial expression recognition capabilities only.*

Here it is important to note that, although a functioning sensory-motor embodiment alone may provide sufficient information to facilitate face identity recognition mechanisms, it is still necessary that the subject does not exhibit impairments in making use of such information. In other words, if for some reason the information cannot be used (or it is used wrongly) when learning new associations between the perceived stimulus and the corresponding identity or these associations cannot be correctly stored in the memory, the subject would still fail to recognise identities from faces. Hence, the present hypothesis does not neglect interactions from other impaired cognitive processes that may further affect facial identity recognition capabilities, even in the presence of unimpaired sensory-motor embodied representations. The hypotheses are summarised by Figure 3.1, whereas Figure 3.2 shows examples of how alterations of visceral and sensory-motor information may impair facial expression recognition capabilities.

I argue that both sensory-motor and visceral information lie on continua supporting smooth and gradual transactions among the available bodily representations. Concepts (*e.g.* facial expression of emotion, such as happy or sad or more abstract concepts like valence and arousal of the considered expressions) can be attributed to specific contiguous areas of the overall bodily formatted space. A facial expression discrimination task would require the subject to attribute the

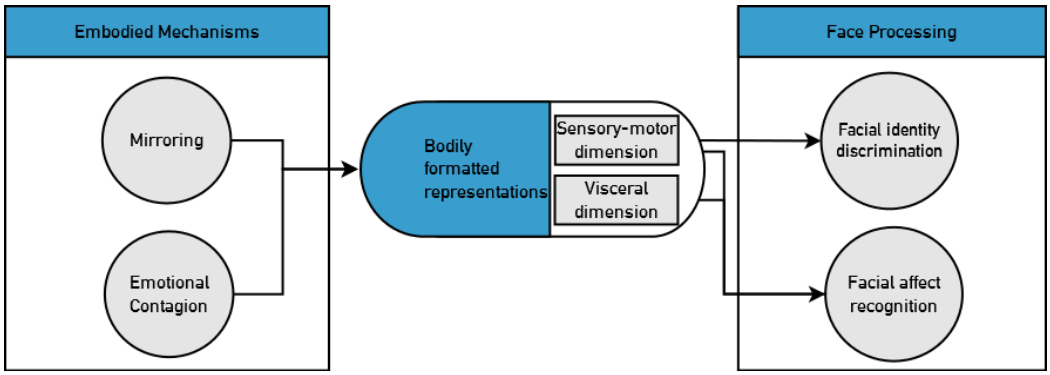


Figure 3.1 Embodied simulation mechanisms, via mirroring and emotional contagion, provides bodily formatted representations of the face stimulus (refer to Sections 2.1.3, 2.1.4 and 2.3.3 for definitions and supporting evidence). These representations exhibit two dimensions: a sensory-motor dimension and a visceral dimension. These dimensions provide necessary information to shape face processing mechanisms. In particular, the sensory-motor dimension alone interacts with face identity discrimination mechanisms, whereas the visceral dimension provides integral information that, together with the one available from sensory-motor dimension, interacts with facial affect recognition mechanisms.

facial expression associated with the the region closer to the currently available bodily formatted representation (see Figure 3.2a).

A lack of information from the visceral dimension would lead to *project* the concept's regions onto the sensory-motor dimension, consequently reducing the chances to discriminate among opaque stimuli (Figure 3.2b). This may happen, for example, in psychopathic individuals when discriminating among subtle facial expressions. On the other hand, missing information from corrupted sensory-motor embodiment would lead to multiple optimal solutions on the sensory-motor continua (Figure 3.2c). The information from the visceral dimension becomes useless in identifying the final optimal solution. Thus, the impaired subject may select a suboptimal solution leading to a shorter distance from the wrong concept (*i.e.* concept A in the visual example shown in Figure 3.2c), and attribute it to the observed face stimulus.

In Chapter 6, I will show how simulated alterations of sensory-motor information may impact on both facial identity and expression discrimination tasks and how simulated alterations of visceral information may impact on facial expression recognition only. These results will demonstrate that impaired embodied mechanisms significantly affect face processing capabilities, thus supporting the hypothesis that social cognition is crucially embodied.

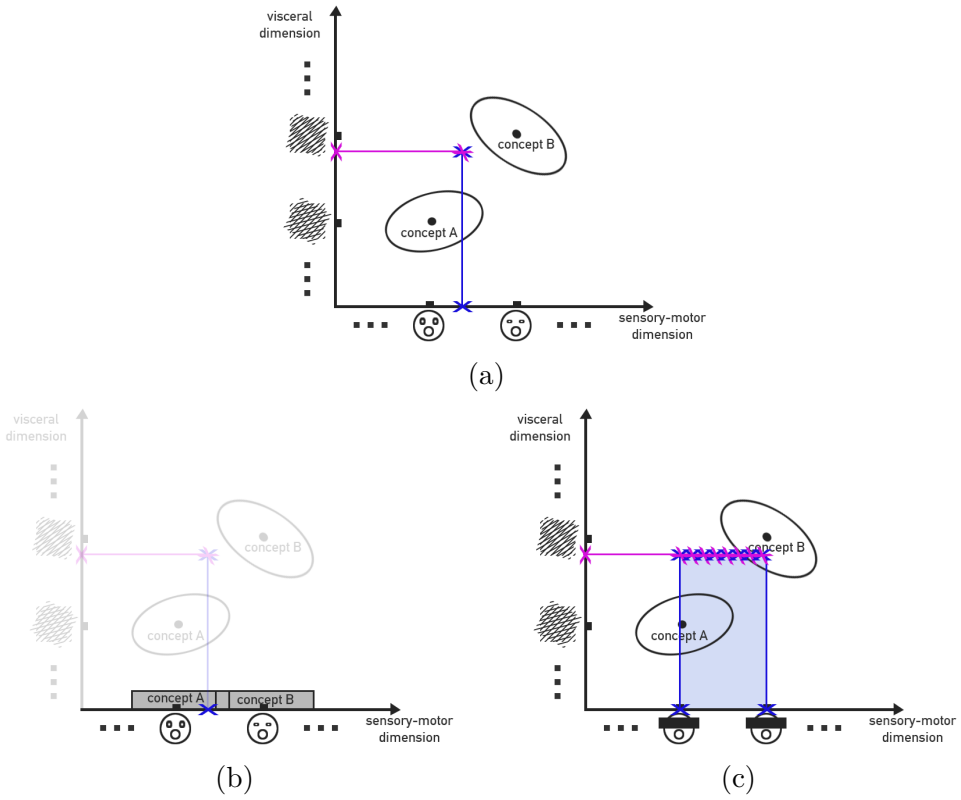


Figure 3.2 The pictures show examples of interactions between the sensory-motor (in blue, vertical axis) and visceral (in magenta, horizontal axis) dimensions during a facial affect recognition task. (a) A healthy subject can make use of both these dimensions to determine a point in the bodily formatted space closer to the region of the right concept to attribute to the face stimulus; This way it is possible to discriminate even between subtle expressions, by means of integral information coming from visceral sensations. (b) A subject without (or with limited) access to the visceral dimension (*e.g.* psychopaths) can still identify the correct point on the sensory-motor continuum but without the visceral dimension the individual projects the bodily formatted representation onto the sensory-motor dimension, and potentially attribute a wrong concept to the observed stimulus. (c) A subject unable to correctly detect sensory-motor features of the face (*e.g.* poor attention to the eye region) cannot identify a single optimal point on the sensory-motor continuum, but a set of possible suboptimal points (light-blue region). Thus, even with the presence of visceral dimension, the subject can still make wrong attributions (*e.g.* by choosing as the optimal solution for the sensory-motor continuum the one on the far left).

3.7 Conclusions

In this chapter, I reviewed works investigating three widely studied social disorders: autism, schizophrenia, and psychopathy. I compared the findings and proposed a novel reading, suggesting that the identified dysfunctions can plausibly impact on the correct functioning of embodied mechanisms. Furthermore, I reviewed the literature on their face processing capabilities compared to healthy individuals. Whereas autism and schizophrenia subjects exhibit deficits in both facial identity and affect recognition tasks, psychopathic individuals do not exhibit impairments in discriminating facial identities. In addition, contrary to autism and schizophrenia individuals, psychopathic subjects do not suffer from significant cognitive impairments, and they seem not to be affected by dysfunctional basic perceptual processes. Rather, psychopathic individuals are specifically affected by emotional processing dysfunctions, which may derive from a reduced attention to peripheral emotional signals.

I then introduced two hypotheses suggesting that:

- Embodied mechanisms provide information shaped in bodily formatted representations lying on two distinct dimensions: a sensory-motor dimension and a visceral dimension;

- Alterations of the sensory-motor embodiment process would impair facial expression and identity recognition capabilities, whereas alterations of the visceral embodiment process would impair facial expression recognition capability only.

Although these hypotheses are not necessary to validate the thesis argument proposed in this dissertation, they provide valuable insights to design an appropriate methodology able to advance the proposed secondary contribution, namely that social cognition is profoundly embodied.

In the following chapters, I aim to provide the computational tools necessary to validate my thesis argument and the aimed secondary contribution. In particular, I will provide:

1. A computational model of embodied simulation reflecting theories from Simulation Theory accounts. I will implement the model by limiting it to realise sensory-motor information of face stimuli (Chapter 4). This will demonstrate that embodied simulation mechanisms can plausibly realise

representations bodily in their format and able to facilitate face processing capabilities;

2. An extension of a framework widely used to explain phenomena underlying facial identity processing, the face-space framework (Valentine et al., 2015) (Chapter 5). I will provide and validate a novel hypothesis suggesting important new features of the offered spatial representation;

3. A set of experiments showing that the sensory-motor information plausibly provided by the proposed embodied simulation mechanisms is sufficient to implement a face-space representation of face stimuli facilitating identity discrimination (Chapter 6). This evidence will be enough to validate the thesis discussed in this dissertation from a computational perspective;

4. Additional computational simulations suggesting that simulated alterations of the sensory-motor embodiment process would lead to impaired facial expression and identity recognition capabilities, whereas simulated alterations of the visceral embodiment process would result in impaired facial expression recognition capabilities only (Chapter 6). These results will provide preliminary computational support of embodied cognition theories.

Chapter Bibliography

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12(2):169–177.

Agnew, Z. K., Bhakoo, K. K., and Puri, B. K. (2007). The human mirror system: A motor resonance theory of mind-reading. *Brain Research Reviews*, 54(2):286–293.

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR®*. American Psychiatric Publishing.

Aniskiewicz, A. S. (1979). Autonomic components of vicarious conditioning and psychopathy. *Journal of Clinical Psychology*.

Bachevalier, J. and Loveland, K. A. (2006). The orbitofrontal–amygdala circuit and self-regulation of social–emotional behavior in autism. *Neuroscience & Biobehavioral Reviews*, 30(1):97–117.

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1):37–46.

Bauser, D. S., Thoma, P., Aizenberg, V., Brüne, M., Juckel, G., and Daum, I. (2012). Face and body perception in schizophrenia: A configural processing deficit? *Psychiatry Research*, 195(1):9–17.

Benton, A. L. (1994). *Contributions to Neuropsychological Assessment: A Clinical Manual*. Oxford University Press, USA.

Berlucchi, G. and Aglioti, S. (1997). The body in the brain: Neural bases of corporeal awareness. *Trends in Neurosciences*, 20(12):560–564.

Billeke, P. and Aboitiz, F. (2013). Social cognition in schizophrenia: From social stimuli processing to social engagement. *Frontiers in Psychiatry*, 4:4.

Birmingham, E., Cerf, M., and Adolphs, R. (2011). Comparing social attention in autism and amygdala lesions: Effects of stimulus and task condition. *Social Neuroscience*, 6(5-6):420–435.

Birmingham, E. and Kingstone, A. (2009). Human social attention. *Annuals of the New York Academy of Sciences*, 1156(1):118–140.

Blair, R. J. R. (1999). Responsiveness to distress cues in the child with psychopathic tendencies. *Personality and Individual Differences*, 27(1):135–145.

Blair, R. J. R., Jones, L., Clark, F., and Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, 34(2):192–198.

Bons, D., Van Den Broek, E., Scheepers, F., Herpers, P., Rommelse, N., and Buitelaaar, J. K. (2013). Motor, emotional, and cognitive empathy in children and adolescents with autism spectrum disorder and conduct disorder. *Journal of Abnormal Child Psychology*, 41(3):425–443.

Bortolon, C., Capdevielle, D., and Raffard, S. (2015). Face recognition in schizophrenia disorder: A comprehensive review of behavioral, neuroimaging and neurophysiological studies. *Neuroscience & Biobehavioral Reviews*, 53:79–107.

Brady, T. F., Konkle, T., and Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, 138(4):487.

Brothers, L. (2002). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Foundations in Social Neuroscience*, pages 367–385.

Brüne, M. (2005). "Theory of mind" in schizophrenia: A review of the literature. *Schizophrenia Bulletin*, 31(1):21–42.

Burack, J. A., Enns, J. T., and Fox, N. A. (2012). *Cognitive Neuroscience, Development, and Psychopathology*. Oxford University Press.

Burke, J. D., Loeber, R., and Lahey, B. B. (2007). Adolescent conduct disorder and interpersonal callousness as predictors of psychopathy in young adults. *Journal of Clinical Child and Adolescent Psychology*, 36(3):334–346.

Contreras-Rodríguez, O., Pujol, J., Batalla, I., Harrison, B. J., Bosque, J., Ibern-Regàs, I., Hernández-Ribas, R., Soriano-Mas, C., Deus, J., and López-Solà, M. (2014). Disrupted neural processing of emotional faces in psychopathy. *Social Cognitive and Affective Neuroscience*, 9(4):505–512.

Cook, R., Bird, G., Catmur, C., Press, C., and Heyes, C. (2014). Mirror neurons: From origin to function. *Behavioral and Brain Sciences*, 37(02):177–192.

Coplan, A. and Goldie, P. (2011). *Empathy: Philosophical and Psychological Perspectives*. Oxford University Press.

Dawel, A., O'Kearney, R., McKone, E., and Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience & Biobehavioral Reviews*, 36(10):2288–2304.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., and Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders*, 28(6):479–485.

De Vignemont, F. (2009). Drawing the boundary between low-level and high-level mindreading. *Philosophical Studies*, 144(3):457–466.

Decety, J. and Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: Implications for intervention across different clinical conditions. *BioPsychoSocial Medicine*, 1(1):22.

Decety, J., Skelly, L., Yoder, K. J., and Kiehl, K. A. (2014). Neural processing of dynamic emotional facial expressions in psychopaths. *Social Neuroscience*, 9(1):36–49.

Decety, J. and Sommerville, J. A. (2003). Shared representations between self and other: A social cognitive neuroscience view. *Trends in cognitive sciences*, 7(12):527–533.

Dennett, H. W., McKone, E., Edwards, M., and Susilo, T. (2012). Face aftereffects predict individual differences in face recognition ability. *Psychological Science*, 23(11):1279–1287.

Dolan, M. and Fullam, R. (2006). Face affect recognition deficits in personality-disordered offenders: Association with psychopathy. *Psychological Medicine*, 36(11):1563–1569.

Duchaine, B. C. and Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia*, 41(6):713–720.

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., and Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with asperger syndrome using the multifaceted empathy test (MET). *Journal of Autism and Developmental Disorders*, 38(3):464–473.

Elsabbagh, M., Mercure, E., Hudry, K., Chandler, S., Pasco, G., Charman, T., Pickles, A., Baron-Cohen, S., Bolton, P., and Johnson, M. H. (2012). Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. *Current Biology*, 22(4):338–342.

Ermer, E., Kahn, R. E., Salovey, P., and Kiehl, K. A. (2012). Emotional intelligence in incarcerated men with psychopathic traits. *Journal of Personality and Social Psychology*, 103(1):194.

Fairchild, G., Stobbe, Y., Van Goozen, S. H., Calder, A. J., and Goodyer, I. M. (2010). Facial expression recognition, fear conditioning, and startle modulation in female subjects with conduct disorder. *Biological Psychiatry*, 68(3):272–279.

Fairchild, G., Van Goozen, S. H., Calder, A. J., Stollery, S. J., and Goodyer, I. M. (2009). Deficits in facial expression recognition in male adolescents with early-onset or adolescence-onset conduct disorder. *Journal of Child Psychology and Psychiatry*, 50(5):627–636.

Falck-Ytter, T., Bölte, S., and Gredebäck, G. (2013). Eye tracking in early autism research. *Journal of Neurodevelopmental Disorders*, 5(1):28.

Farrow, T. F. and Woodruff, P. W. (2007). *Empathy in Mental Illness*. Cambridge University Press Cambridge.

Fecteau, S., Pascual-Leone, A., and Théoret, H. (2008). Psychopathy and the mirror neuron system: Preliminary findings from a non-psychiatric sample. *Psychiatry Research*, 160(2):137–144.

Gallese, V. (2014). Bodily selves in relation: Embodied simulation as second-person perspective on intersubjectivity. *Philosophical Transactions of the Royal Society B*, 369(1644).

Gallese, V. and Ferri, F. (2013). Jaspers, the body, and schizophrenia: The bodily self. *Psychopathology*, 46(5):330–336.

Gallese, V., Rochat, M., Cossu, G., and Sinigaglia, C. (2009). Motor cognition and its role in the phylogeny and ontogeny of action understanding. *Developmental Psychology*, 45(1):103.

Gordon, H. L., Baird, A. A., and End, A. (2004). Functional differences among those high and low on a trait measure of psychopathy. *Biological Psychiatry*, 56(7):516–521.

Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological Bulletin*, 141(6):1266.

Guillon, Q., Hadjikhani, N., Baduel, S., and Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews*, 42:279–297.

Harms, M. B., Martin, A., and Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, 20(3):290–322.

Horan, W. P., Wynn, J. K., Kring, A. M., Simons, R. F., and Green, M. F. (2010). Electrophysiological correlates of emotional responding in schizophrenia. *Journal of Abnormal Psychology*, 119(1):18.

Iacoboni, M. and Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, 7(12):942–951.

Izuma, K., Matsumoto, K., Camerer, C. F., and Adolphs, R. (2011). Insensitivity to social reputation in autism. *Proceedings of the National Academy of Sciences*, 108(42):17302–17307.

Kennedy, D. P. and Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends in Cognitive Sciences*, 16(11):559–572.

Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, 59(9):809–816.

Lozier, L. M., Vanmeter, J. W., and Marsh, A. A. (2014). Impairments in facial affect recognition associated with autism spectrum disorders: A meta-analysis. *Development and Psychopathology*, 26(4pt1):933–945.

Marsh, A. A. and Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, 32(3):454–465.

Marwick, K. and Hall, J. (2008). Social cognition in schizophrenia: A review of face processing. *British Medical Bulletin*, 88(1):43–58.

Matthias, E., Schandry, R., Duschek, S., and Pollatos, O. (2009). On the relationship between interoceptive awareness and the attentional processing of visual stimuli. *International Journal of Psychophysiology*, 72(2):154–159.

McCleery, A., Lee, J., Joshi, A., Wynn, J. K., Hellemann, G. S., and Green, M. F. (2015). Meta-analysis of face processing event-related potentials in schizophrenia. *Biological Psychiatry*, 77(2):116–126.

Megreya, A. M. (2016). Face perception in schizophrenia: A specific deficit. *Cognitive Neuropsychiatry*, 21(1):60–72.

Morin, K., Guy, J., Habak, C., Wilson, H. R., Pagani, L., Mottron, L., and Bertone, A. (2015). Atypical face perception in autism: A point of view? *Autism Research*, 8(5):497–506.

Navab, A., Gillespie-Lynch, K., Johnson, S. P., Sigman, M., and Hutman, T. (2012). Eye-tracking as a measure of responsiveness to joint attention in infants at risk for autism. *Infancy*, 17(4):416–431.

Newman, J. P. and Lorenz, A. R. (2003). Response modulation and emotion processing: Implications for psychopathy and other dysregulatory psychopathology. In Davidson, R. J., Scherer, K. R., and Goldsmith, H. H., editors, *Handbook of Affective Sciences*, pages 904–929. Oxford University Press.

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., and Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, 24(2):190–198.

O'Brien, J., Spencer, J., Girges, C., Johnston, A., and Hill, H. (2014). Impaired perception of facial motion in autism spectrum disorder. *PloS One*, 9(7):e102173.

Parasuraman, R. (1998). *The attentive brain*. Mit Press Cambridge, MA.

Pomarol-Clotet, E., Hynes, F., Ashwin, C., Bullmore, E., McKenna, P., and Laws, K. (2010). Facial emotion processing in schizophrenia: A non-specific neuropsychological deficit? *Psychological Medicine*, 40(06):911–919.

Postmes, L., Sno, H., Goedhart, S., van der Stel, J., Heering, H., and de Haan, L. (2014). Schizophrenia as a self-disorder due to perceptual incoherence. *Schizophrenia Research*, 152(1):41–50.

Rhodes, G., Ewing, L., Jeffery, L., Avard, E., and Taylor, L. (2014). Reduced adaptability, but no fundamental disruption, of norm-based face-coding mechanisms in cognitively able children and adolescents with autism. *Neuropsychologia*, 62:262–268.

Rizzolatti, G. and Fabbri-Destro, M. (2010). Mirror neurons: From discovery to autism. *Experimental Brain Research*, 200(3-4):223–237.

Rizzolatti, G. and Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11(4):264–274.

Sasson, N., Tsuchiya, N., Hurley, R., Couture, S. M., Penn, D. L., Adolphs, R., and Piven, J. (2007). Orienting to social stimuli differentiates social cognitive impairment in autism and schizophrenia. *Neuropsychologia*, 45(11):2580–2588.

Sasson, N. J. (2006). The development of face processing in autism. *Journal of Autism and Developmental Disorders*, 36(3):381–394.

Savla, G. N., Vella, L., Armstrong, C. C., Penn, D. L., and Twamley, E. W. (2012). Deficits in domains of social cognition in schizophrenia: A meta-analysis of the empirical evidence. *Schizophrenia Bulletin*, page sbs080.

Schönenberg, M., Mayer, S. V., Christian, S., Louis, K., and Jusyte, A. (2015). Facial affect recognition in violent and nonviolent antisocial behavior subtypes. *Journal of Personality Disorders*, pages 1–12.

Sestito, M., Raballo, A., Umiltà, M. A., Leuci, E., Tonna, M., Fortunati, R., De Paola, G., Amore, M., Maggini, C., and Gallese, V. (2015). Mirroring the self: Testing neurophysiological correlates of disturbed self-experience in schizophrenia spectrum. *Psychopathology*, 48(3):184–191.

Sparks, A., McDonald, S., Lino, B., O'Donnell, M., and Green, M. J. (2010). Social cognition, empathy and functional outcome in schizophrenia. *Schizophrenia Research*, 122(1):172–178.

Streit, M., Wölwer, W., and Gaebel, W. (1997). Facial-affect recognition and visual scanning behaviour in the course of schizophrenia. *Schizophrenia Research*, 24(3):311–317.

Sully, K., Sonuga-Barke, E. J., and Fairchild, G. (2015). The familial basis of facial emotion recognition deficits in adolescents with conduct disorder and their unaffected relatives. *Psychological Medicine*, 45(09):1965–1975.

Tempesta, D., Stratta, P., Marrelli, A., Aloisi, P., Arnone, B., Gasbarri, A., and Rossi, A. (2014). Facial emotion recognition in schizophrenia: An event-related potentials study. *Rivista di Psichiatria*, 49(4):183–186.

Théoret, H., Halligan, E., Kobayashi, M., Fregni, F., Tager-Flusberg, H., and Pascual-Leone, A. (2005). Impaired motor facilitation during action observation in individuals with autism spectrum disorder. *Current Biology*, 15(3):R84–R85.

Uljarevic, M. and Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, 43(7):1517–1526.

Valentine, T., Lewis, M. B., and Hills, P. J. (2015). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, pages 1–24.

Vitale, J., Williams, M.-A., and Jonhston, B. (2016). The face-space duality hypothesis: A computational model. In *38th Annual Meeting of the Cognitive Science Society*, pages 514–519.

Vivanti, G. and Hamilton, A. (2014). Imitation in autism spectrum disorders. In Volkmar, F., Rogers, S., Paul, R., and Pelphrey, K. A., editors, *Handbook of Autism and Pervasive Developmental Disorders*. John Wiley & Sons Inc, 4 edition.

Weigelt, S., Koldewyn, K., and Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience & Biobehavioral Reviews*, 36(3):1060–1084.

Williams, E. (1974). An analysis of gaze in schizophrenics. *British Journal of Social and Clinical Psychology*, 13(1):1–8.

World Health Organization (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*, volume 1. World Health Organization.

Yirmiya, N., Kasari, C., Sigman, M., and Mundy, P. (1989). Facial expressions of affect in autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry*, 30(5):725–735.

Fortunately, most human behaviour is learnt observationally through modelling from others.

— Albert Bandura —

4

Embodiment of Sensory-Motor Facial Information: a Probabilistic Account[1]

In this chapter I propose a probabilistic computational theory of embodied simulation, limiting its implementation to the *embodiment* of face stimuli. To make this intention clearer, in the rest of this chapter I will use the term *embodiment* to describe *the process of deriving bodily formatted representations*, as per Definitions 2.23 and 2.25 provided on pages 56–57. This circumscribed task is considered crucial to promoting social cognition and, therefore, particularly necessary for interpreting others' minds (Gallese, 2016; Goldman and Sripada, 2005).

From a general standpoint, *mind-reading* is the process of inferring the mental state of other people based on their overt/observable behaviour (Goldman and Sripada, 2005), such as facial expressions (Jack and Schyns, 2015). This skill plays a major role in social interactions, empathy and effective communication (Brothers, 2002; De Vignemont and Singer, 2006). Embodied simulation mechanisms are suggested to be at the core of mind-reading processes. In fact, they plausibly

[1]This chapter is an adaptation of *"Vitale, J., Williams, M.-A., Johnston, B., and Boccignone, G. (2014). Affective facial expression processing via simulation: A probabilistic model. Biologically Inspired Cognitive Architectures , 10:30–41"*.

enable representations bodily in format able to promote mind-reading capabilities (Gallese and Sinigaglia, 2011).

As discussed in Section 2.3.4 (page 58), Gallese (2016) suggests that embodied simulation provides mechanisms necessary to achieve a simulation based mind-reading process, as supported by the Simulation Theory account from philosophy of mind (Goldman and Sripada, 2005). According to Simulation Theory, an observer arrives at a mental attribution by simulating, in his/her own mind and body, the same state as the target. This bodily simulation process can be realised through embodied simulation mechanisms, via mirroring and emotional contagion (see Figure 2.1).

Hence, having a computational model of simulation-based mind-reading describing embodied simulation mechanisms can significantly advance neuropsychological and theoretical understanding of embodiment phenomena (Gallese, 2016). In addition, the present contribution can also foster application-oriented areas such as social robotics and social signal processing (Boccignone et al., 2018; Pantic and Bartlett, 2007; Vitale et al., 2014a). Importantly, this contribution provides a computational explanation of embodied simulation mechanisms that does not only answer to whether embodied simulation shape the mind, but it also provides valuable insights to describe *how* embodied simulation may shape cognition, as I will argue in the remainder of this dissertation.

While the neuropsychological account of embodied simulation underlying a Simulation Theory process is modern and compelling (Gallese, 2016), a critical question remains poorly answered and largely unexplored (Gallese and Sinigaglia, 2011; Goldman and Sripada, 2005):

> *How is it possible to describe, from a computational level*[2]*, the embodied simulation mechanisms underlying a mind-reading process and reused to achieve other cognitive capabilities?*

Current literature includes some probabilistic models for motor action prediction and understanding inspired by simulation theories (Boccignone et al., 2018; Demiris and Johnson, 2003; Dindo et al., 2011; Watanabe et al., 2007; Wolpert and Flanagan, 2001). However, the model described in this chapter is novel in providing a computational account of simulation-style mind-reading via embodied simulation mechanisms, applied to the context of face-to-face interactions and further linking to face processing studies. In particular, this model aims to overcome some limitations not completely addressed by other previous works.

[2]The *"what"* level of explanation, in the sense of Marr (1982)

Among others, the generalisation of the model over several different observed identities and the implementation of both forward and inverse mechanisms able to not only embody but also to generate facial expressions from bodily formatted representations. In addition, this framework draws inspiration from neuroscience studies on mirror neuron system, thus making it consistent with biological human findings. Finally, this account links to face processing studies and it consequently offers a more pervasive role of embodied simulation for the development of other cognitive capabilities, as I will show in Chapter 6.

Therefore, the aim of this chapter is to answer the question introduced previously in this section and to take a step toward advancing the argument of this dissertation. In particular, I seek to apply Simulation Theory accounts to the problem of mapping an observed overt behaviour (in this dissertation limited to facial expressions) to a phenomenological internal latent space of the mind-reader[3]. I will evaluate the proposed model by showing that:

- The realised first-order phenomenological latent space provides representations bodily in format able to encode sensory-motor information of the observed face stimuli;

- The proposed embodied mechanisms can facilitate the classification of facial motor configurations.

I will use the offered computational account and the gathered experimental evidence to advance my thesis argument and validate the auxiliary hypotheses proposed in this dissertation.

In this study, I will mainly focus on describing plausible computational mechanisms of overt facial behaviour embodiment, employed during face-to-face social interactions. Thus, I will not investigate the attribution of the associated mental state given such internal representation (*i.e.* cognitive appraisal of the mind-reading process) since out of scope for the present dissertation.

4.1 Background

The attribution of a mental state (*e.g.* emotional state) to others can occur when a complex state of the organism is accompanied by variable degrees of awareness, variously indicated as *appraisal*.

[3]In this context proposed as the internal response of the subject given a particular stimulus

Two levels of appraisal can be distinguished (Lambie and Marcel, 2002): a first-order phenomenological state and a conscious second-order awareness. Both states can be either self-directed (first-person perspective) or world-directed (third person perspective).

The content of the first-order phenomenological state is physical and visceral, centred on one's body state and related neural underpinnings. In this dissertation I suggested that this state is available via embodied simulation mechanisms, in the form of bodily formatted representations. By contrast, the content of second-order conscious awareness can be either propositional or non-propositional. In this chapter, I will be concerned with modelling the first-order phenomenological experience, relying on bodily formatted representations.

In the following discussion, I will use a notation in order to describe more easily the simulation process as described by philosophy of mind literature (Goldman and Sripada, 2005). Therefore, I will generically refer to a behaviour as $\mathbf{X}$, to the first-order phenomenological state as $\mathbf{\Phi}$, and to a second-order conscious mental state as $\mathbf{\Psi}$. Furthermore, I will use the notation $\mathbf{A} \simeq \mathbf{B}$ meaning that $\mathbf{A}$ *is similar to* $\mathbf{B}$, and the notation $\mathbf{A} \mapsto \mathbf{B}$ meaning that $\mathbf{A}$ *corresponds to* $\mathbf{B}$. The suggested simulation-based mind-reading process can be summarised as follows (Goldman and Sripada, 2005):

(a) In a given situation, a subject (the *target*) experiences a phenomenological bodily state $\mathbf{\Phi}_{tg}$. The bodily state may be either triggered by an external event and/or mentally induced (*e.g.* through a particular memory or mental imagery). This bodily state elicits a corresponding behaviour $\mathbf{X}_{tg}$ (*e.g.* facial expression, gesture, heart beat, *etc.*) and it is associated with a specific mental state $\mathbf{\Psi}_{tg}$. Similarly, A second subject (*i.e.* the *mind-reader*), is experiencing a different bodily state $\mathbf{\Phi}_{mr}$ eliciting a corresponding behaviour $\mathbf{X}_{mr}$ and associated with a different mental state $\mathbf{\Psi}_{mr}$ (Figure 4.1a);

(b) While interacting with the target subject, the mind-reader perceives the observable behaviour $\mathbf{X}_{tg}$ of the interaction partner (Figure 4.1b);

(c) The mind-reader employs embodied simulation mechanisms to make sense of the observed behaviour. Therefore, the mind-reader embodies within himself the phenomenological bodily state $\mathbf{\Phi}_{mr}$ associated with a simulated behaviour $\check{\mathbf{X}}_{mr}$ similar to the one observed from the target subject (Figure 4.1c);

(d) Finally, the mind-reader selects a mental state $\mathbf{\Psi}_{mr}$ to attribute to the target subject by using as evidence the simulated phenomenological state $\mathbf{\Phi}_{mr}$ and the corresponding simulated behaviour $\check{\mathbf{X}}_{mr}$ (Figure 4.1d).

This account is illustrated in Figure 4.1.

4.1.1 Simulation Theory Accounts

There are several ways that the proposed account might be translated into a computational theory. Goldman and Sripada (2005) have devised four accounts of Simulation Theory-based mind-reading, having substantial plausibility and consistency with neuropsychological evidence:

1. Generate-and-test models;
2. Reverse simulation models;
3. Variants of the reverse simulation model that employ an *as-if* loop;
4. Unmediated resonance models.

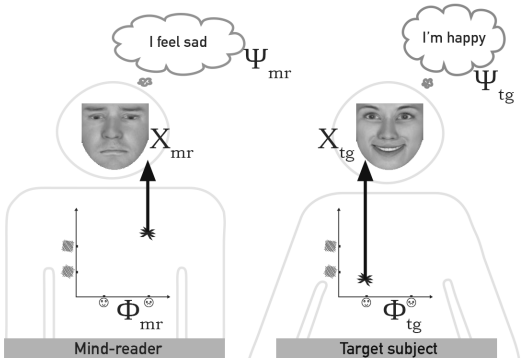
Generate-and-test models

Generate-and-test models assume that the mind-reader starts by hypothesising a certain phenomenological bodily state of the target subject $\mathbf{\Phi}_{tg}$ as the possible cause of the target's behaviour $\mathbf{X}_{tg}$. The mind-reader proceeds to mirror that very same state, namely producing a facsimile of it, $\mathbf{\Phi}_{mr}$, in his/her own system. If the behaviour $\mathbf{X}_{mr}$ resulting from such simulated process matches the behaviour observed in the target subject $\mathbf{X}_{tg}$, then the hypothesised phenomenological bodily state is classified with a specific interpretation available from the mind-reader conscious second-order awareness $\mathbf{\Psi}_{mr}$, which is then attributed to the target subject:

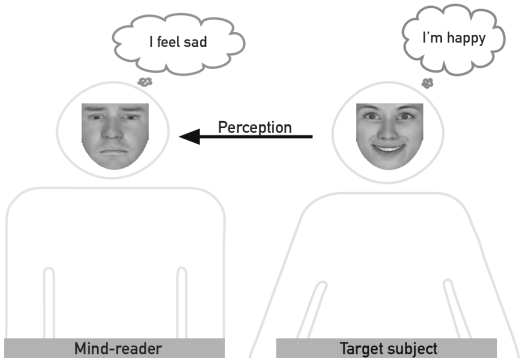
`Given a hypothesis:` $\mathbf{\Phi}_{tg} \simeq \mathbf{\Phi}_{mr}$
`IF:` $\mathbf{\Phi}_{mr} \mapsto \mathbf{X}_{mr} \simeq \mathbf{X}_{tg}$
`It follows:` $\mathbf{\Phi}_{mr} \mapsto \mathbf{\Psi}_{mr} \simeq \mathbf{\Psi}_{tg}$

Reverse simulation models

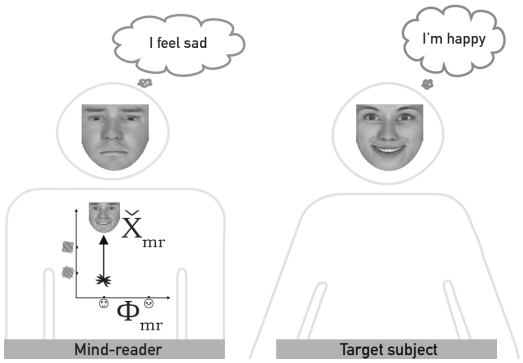
In reverse simulation, the mind-reader engages in the opposite direction. Namely, given the observed behaviour $\mathbf{X}_{tg}$, the mind-reader activates imitative mechanisms



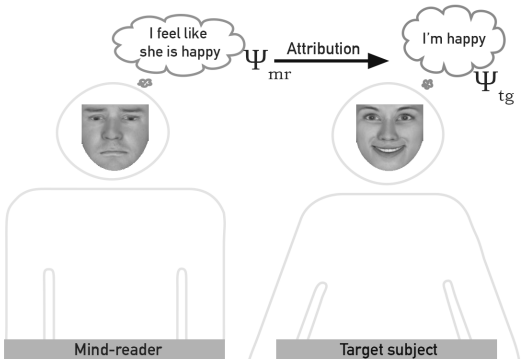
(a) The target subject experiences a phenomenological state $\mathbf{\Phi}_{tg}$ corresponding to a behaviour $\mathbf{X}_{tg}$ and a mental state $\mathbf{\Psi}_{tg}$, whereas the mind-reader experiences a different phenomenological state $\mathbf{\Phi}_{mr}$ corresponding to a different behaviour $\mathbf{X}_{mr}$ and a different mental state $\mathbf{\Psi}_{mr}$.



(b) The realised behaviour of the target $\mathbf{X}_{tg}$ is perceived by the mind-reader.



(c) The mind-reader embodies the observed behaviour $\mathbf{X}_{tg}$ experiencing a phenomenological state $\mathbf{\Phi}_{mr} \simeq \mathbf{\Phi}_{tg}$ and internally simulating the corresponding motor behaviour $\check{\mathbf{X}}_{mr} \simeq \mathbf{X}_{tg}$.



(d) The mind-reader selects a mental state $\mathbf{\Psi}_{mr} \simeq \mathbf{\Psi}_{tg}$ after using as evidence the experienced phenomenological state $\mathbf{\Phi}_{mr}$ following the embodiment process.

Figure 4.1 Embodiment of a face stimulus and attribution of a mental state during a face-to-face interaction.

to mimic the target's behaviour ($\mathbf{X}_{mr} \simeq \mathbf{X}_{tg}$). In order to overtly activate such behaviour (*e.g.* activating the appropriate facial muscles), the mind-reader activates the associated phenomenological bodily state $\mathbf{\Phi}_{mr}$. Under the assumptions that $\mathbf{\Phi}_{mr} \simeq \mathbf{\Phi}_{tg}$, this inner experience is classified as $\mathbf{\Psi}_{mr}$ and attributed to the actor:

Given: $\mathbf{X}_{mr} \simeq \mathbf{X}_{tg}$
It follows: $\mathbf{X}_{mr} \mapsto \mathbf{\Phi}_{mr} \mapsto \mathbf{\Psi}_{mr} \simeq \mathbf{\Psi}_{tg}$

Reverse simulation models employing an *as-if* loop

The *as-if* variants of reverse simulation assume that there may be direct links between the perceptual representation of the target's behaviour $\mathbf{X}_{tg}$ and the sensory-motor and visceral representation of *"what it would feel like"* were the mind-reader to exhibit that behaviour, *i.e.* $\mathbf{\Phi}_{mr}$. Thus, this model is similar to reverse simulation but avoids involving an explicit overt imitation to generate $\mathbf{X}_{mr}$, thus relying only on mental mirroring mechanisms:

Given: $\mathbf{X}_{tg}$
It follows: $\mathbf{X}_{tg} \mapsto \mathbf{\Phi}_{mr} \mapsto \mathbf{\Psi}_{mr} \simeq \mathbf{\Psi}_{tg}$

Unmediated resonance models

The rationale behind the unmediated resonance model is that observation of the target's behaviour directly triggers the activation of a same neural substrate in the mind-reader associated with the internal state of the target subject $\mathbf{\Phi}_{tg}$ in question. This is most similar to the *as-if* model, and there is no mediation of any kind.

Given: $\mathbf{X}_{tg}$
It follows: $\mathbf{\Phi}_{mr} \simeq \mathbf{\Phi}_{tg}$ and therefore $\mathbf{\Phi}_{mr} \mapsto \mathbf{\Psi}_{mr} \simeq \mathbf{\Psi}_{tg}$

At least two observations may be drawn regarding the suggested computational theories. The first is that some sort of projection or cross-modal matching should be introduced to map the perceptual representation $\mathbf{X}_{tg}$ of target's behaviour to an egocenric representation of the mind-reader's own behaviour realisation $\mathbf{X}_{mr}$. This egocentric representation is most likely a proprioceptive motor image, but see the work of Goldman and Sripada (2005) and Section 4.1.3 for a general discussion.

This is certainly evident for the generate-and-test and reverse simulation models, but are also relevant for the *as-if* and unmediated resonance models.

The second conclusion is that all models assume that the production of the relevant internal state in a mind-reader is eventually transmitted to some cognitive centre that *recognises* or *labels* the experienced phenomenological state (*i.e.* $\mathbf{\Psi}_{mr}$). This step relates to second-order cognitive appraisal, which I will not extensively consider in this dissertation.

In the remainder of this chapter, I propose a general probabilistic framework which generalises the most salient aspects of the four approaches discussed above, and I implement it to the specific domain of face-to-face interactions. Therefore, I formalise the embodied mechanisms at the core of any Simulation Theory approach by deriving probabilistic latent variable spaces able to:

1. Map a sensed facial expression to an egocentric motor representation of it (*i.e.* transcoding process): $\mathbf{X}_{tg} \mapsto \mathbf{X}_{mr}$;

2. Generate an egocentric motor representation of a facial motor configuration from a first-order phenomenological state (*i.e.* forward process): $\mathbf{\Phi}_{mr} \mapsto \mathbf{X}_{mr}$;

3. Map an egocentric motor representation to a phenomenological bodily representation (*i.e.* inverse process): $\mathbf{X}_{mr} \mapsto \mathbf{\Phi}_{mr}$.

4.1.2 The Two Dimensions of Bodily Representations

In order to model a Simulation Theory process, the mind-reader has to *experience* the very same internal state of the target subject (Goldman and Sripada, 2005). In this respect, all four approaches previously outlined satisfy this necessary condition. They do so by simulating the internal state using different embodied mechanisms, such as specific emotion production systems, the facial musculature and somatosensory centres.

From a neuropsychological perspective, some areas of the brain seem to be more involved in *emotional* representations of internal states (*e.g.* the insula or the amygdala) (Adolphs, 2002a), while others areas serve more as an internal action representations system (*e.g.* mirror neuron system), typically associated with producing actions and triggered during the observation of someone else's corresponding actions (Gallese, 2007; Gallese et al., 2004). More generally, a whole range of different mirror matching mechanisms instantiating simulation routines is likely to be present in our brain (Gallese, 2003; Gallese and Caruana,

2016). In Chapter 3 I argued that the bodily formatted representations, realised via by embodied mechanisms, exhibit two dimensions: a sensory-motor dimension, associated with perceptual and motor aspects of the face stimulus, and a visceral dimension, related to emotional and experiential aspects of the observed face stimulus.

In this study, I focus on implementing the 'sensory-motor' dimension of the considered internal representations, more in the vein of Gallese's *"shared manifold hypothesis"* (Gallese, 2001, 2003) and available studies on mirror neuron system (Gallese, 2007; Gallese et al., 2004). Hence, I will not provide a computational implementation of the 'visceral' dimension of the suggested bodily representations.

This choice is mainly due to the very limited amount of data available in literature to implement such visceral dimension in a computational model. In fact, whereas gathering images of face stimuli displaying motor configurations is nowadays quite simple, collecting visceral information associated with emotional responses is still a very challenging task, specifically due to limitations in sensors' hardware and the large varieties of the used methodological approaches (Greco et al., 2016)[4].

Nevertheless, this is not much of a restriction for the final aim of this dissertation. Firstly, as highlighted by Adolphs (2002a,b), the sensory-motor dimension appears to be critical for the recognition of emotions displayed by others, because sensory-motor systems support the reconstruction of what it would feel like to be in a particular emotion, by means of simulation of the related body state (Adolphs, 2002a). This was also stressed during the reviews in Chapter 2 and 3.

Secondly, to support my thesis I can limit the investigation to the sensory-motor dimension of the resulting bodily formatted representations. In fact, as I will show later in this chapter, the sensory-motor aspect of the resulting bodily formatted representations provides information about dynamic features of the face stimuli (*e.g.* facial configuration), facilitating facial affect recognition. Hence, in Chapter 6 I will be able to show that this information, available via embodied mechanisms, is sufficient to develop facial identity discrimination mechanisms.

4.1.3 Facial Mental Imagery *vs.* Facial Mimicry

Another critical issue about simulation theories is the main distinction between those where actual facial movements are put into work (generate-and-test and

[4]But see the dataset offered by Ringeval et al. (2013) including facial interactions and visceral responses successfully used by Boccignone et al. (2018) in a recent deep learning extension of the theory and model discussed in this chapter.

reverse simulation) and those supported by an *as-if* mechanism. This distinction is related to several controversies from a neuropsychological perspective.

"Wired" tendencies to micro-mimicking and imitation are well supported by early works of Meltzoff and Moore (1983) and Dimberg and Thunberg (1998). In these studies, the subjects spontaneously and overtly activate facial musculature corresponding to visually presented facial expressions; meanwhile, reverse simulation is consistent with the *"facial feedback hypothesis"* (Adelmann and Zajonc, 1989; Levenson et al., 1990).

As an alternative, facial mimicry may accompany but not actually facilitate recognition. A correlational, rather than causal, role for facial musculature in the recognition process is consistent with the results of Calder et al. (2000) and Keillor et al. (2002) where a patient with bilateral facial paralysis performed well on facial based emotion recognition tasks. In other words, the realisation of overt imitative mechanisms can enhance accuracy in attributing the correct mental state to others, but it is not necessary to promote facial expression recognition capability, which instead may make mainly use of covert mirroring mechanisms.

Even if this should be controversial from a neuropsychological perspective, it is not a critical conceptual issue from a strict computational modelling standpoint. Certainly, at the *"what"* level of explanation (Marr, 1982), it is mandatory to account for the mapping $\mathbf{X}_{tg} \mapsto \mathbf{X}_{mr}$ processing the external perceptual representation of the target's expression $\mathbf{X}_{tg}$ into the internally mind-reader-centred representation of the external stimulus $\mathbf{X}_{mr}$ (whether it is an internal image or a proprioceptive representation in somatosensory areas or a bare set of motor parameters).

Clearly, at the algorithmic level[5] this issue can be relevant, at least for practical purposes. However, this has been a largely studied problem, and elegant solutions are at hand. See for instance, in the field of robotics, the work of Lopes and Santos-Victor (2005) on how to compute and learn, through self-observation, a visuo-motor map suitable to transcode visual information to motor data for hand gesture imitation tasks.

To focus on the essential properties of the model, I will simply assume an internal representation $\mathbf{X}_{mr}$ of the mind-reader resembling the observed facial display of the target subject. This internal representation can be shaped in the form of mental image generated by the corresponding bodily formatted representation (Figure 4.1), without necessarily trigger overt facial mimicry mechanisms.

[5]The *"how"* level in Marr's terminology (Marr, 1982)

4.2 The Model

The proposed probabilistic model for simulation-based mental attribution via embodied simulation mechanisms is defined as follows. Assume two interacting individuals, the target subject and the mind-reader (Figure 4.1), and consider state variables $\mathbf{X}_{tg}$, $\mathbf{X}_{mr}$, $\mathbf{\Phi}_{tg}$, $\mathbf{\Phi}_{mr}$,$\mathbf{\Psi}_{tg}$, $\mathbf{\Psi}_{mr}$ as random variables. In this chapter, I will use lowercase letters to indicate samples of the corresponding uppercase random variables and the accent $\check{\cdot}$ to denote that the considered sample resulted from a forward simulation process.

Based on the experimental findings of Gallese (2001), these subjects are then assumed to share a latent manifold of first-order phenomenological states. This manifold is shared among individuals because of common representations bodily in format (Goldman, 2013), and corresponding neural realisations (Gallese, 2001). In this manifold, the random variable $\mathbf{\Phi}$ takes values and has forward and inverse mapping mechanisms $\mathbf{\Phi} \mapsto \mathbf{X}$ and $\mathbf{X} \mapsto \mathbf{\Phi}$, respectively.

I then define the following:

- $P(\mathbf{X}_{mr} \mid \mathbf{X}_{tg})$, the conditional probability density function (pdf) representing for the mind-reader the probability of realising an egocentric display $\mathbf{X}_{mr}$ when the target subject displays $\mathbf{X}_{tg}$;
- $P(\mathbf{\Phi}_{mr} \mid \mathbf{X}_{mr})$, the conditional pdf representing the probability for the mind-reader of being in an internal phenomenological bodily state $\mathbf{\Phi}_{mr}$ given the covert or overt egocentric facial display $\mathbf{X}_{mr}$ (inverse probability);
- $P(\mathbf{X}_{mr} \mid \mathbf{\Phi}_{mr})$, the conditional pdf that the mind-reader generates a facial display $\mathbf{X}_{mr}$ (covertly or overtly) given the phenomenological internal state $\mathbf{\Phi}_{mr}$ (forward probability);
- $\mathcal{M}(\mathbf{x}_{mr},\check{\mathbf{x}}_{mr})$, a decision or matching function comparing the similarity between the sampled mind-reader-centred expression of the target's display $\mathbf{x}_{mr}$ with a forward-simulated display $\check{\mathbf{x}}_{mr}$. The function is operationalised by choosing an appropriate measure (*e.g.* perceptual similarity, joint angles distances, *etc.*) and it returns a positive real value suggesting how much the two motor configurations are similar to each other.

The simulation-style embodiment process can be realised by deriving the following pdfs (where the symbol $\sim$ stands for the sampling operator):

$$\mathbf{x}_{mr} \sim P(\mathbf{X}_{mr} \mid \mathbf{X}_{tg} = \mathbf{x}_{tg}), \tag{4.1}$$

$$\varphi_{mr} \sim P(\mathbf{\Phi}_{mr} \mid \mathbf{X}_{mr} = \mathbf{x}_{mr}), \tag{4.2}$$

$$\check{\mathbf{x}}_{mr} \sim P(\mathbf{X}_{mr} \mid \mathbf{\Phi}_{mr} = \varphi_{mr}), \tag{4.3}$$

In summary:

- Equation 4.1 defines process transforming an instance of the target subject's face expression $\mathbf{x}_{tg}$ into mind-reader's self-centred image $\mathbf{x}_{mr}$, here denoted with the term *'transcoding'*;
- Equation 4.2 defines the *inverse process* of experiencing/detecting state φ_{mr} under (internal) face expression $\mathbf{x}_{mr}$;
- Equation 4.3 accounts for the mind-reader's *forward process* of sampling his/her own simulated internal expression $\check{\mathbf{x}}_{mr}$ when he/she is in the latent internal state φ_{mr};
- The function $\mathcal{M}(\mathbf{x}_{mr},\check{\mathbf{x}}_{mr})$ compares the transcoded egocentric image $\mathbf{x}_{mr}$ (via Equation 4.1) against a set of internally simulated samples $\check{\mathbf{X}}$ (sampled via Equation 4.3), and is used (via Equation 4.2) to control when the matching process has converged to the most likely solution.

These three pdfs and the matching function $\mathcal{M}$ are sufficient to provide the core basis for the simulation model.

Clearly, a full mind-reading process will only be complete after attributing to the target subject a mental state $\mathbf{\Psi}_{tg} = \psi^{\star}_{mr}$, where $\psi^{\star}_{mr}$ is the most likely mental state instantiated by the mind-reader through a further inferential step, by relying on the pdf $P(\mathbf{\Psi}_{mr} \mid \mathbf{\Phi}_{mr} = \varphi_{mr}, \mathbf{C})$ where the random variable $\mathbf{C}$ summarises general contextual or cultural factors (Ojha et al., 2017). However, the attribution level is beyond the scope of this work.

To define the transcoding pdf (Equation 4.1) and the inverse/forward pdfs (Equations 4.2 and 4.3), I will make use of two distinct latent spaces: the *self-projected latent space* $\mathbf{Z}$ and the *first-order phenomenological latent space* $\mathbf{\Phi}$, which I will derive in the following sections.

Before delving into the derivation of the suggested pdfs, recall that a latent variable is a variable having specific features of interest not directly observable from the direct measure, but rather inferred from the observed stimuli through mathematical models. Therefore, a latent space is a space realised by the estimated

latent variables exhibiting the desired inferred features. For example, a direct measure could be the set of intensities of the pixels describing an image of a face. This measure does not provide direct information on how the exhibited facial expression is related to other facial configurations. However, it may be possible to infer this information and realise a latent space able to communicate such feature more directly.

4.2.1 Derivation of the Self-Projected Latent Space

The self-projected latent space $\mathbf{Z}$ can be conceived in terms of Bayesian latent factor regression (Murphy, 2012):

$$P(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I}_{\mathscr{L}}) \tag{4.4}$$

$$P(\mathbf{x}_{tg} \mid \mathbf{z}) = \mathcal{N}(\mathbf{\Theta}^z_{tg}\mathbf{z} + \mu_{tg}, \sigma^2_{tg}\mathbf{I}_{\mathscr{D}}) \tag{4.5}$$

$$P(\mathbf{x}_{mr} \mid \mathbf{z}) = \mathcal{N}(\mathbf{\Theta}^z_{mr}\mathbf{z} + \mu_{mr}, \sigma^2_{mr}\mathbf{I}_{\mathscr{D}}) \tag{4.6}$$

where $\mathcal{N}(\cdot)$ denotes the Gaussian distribution; $\mathbf{\Theta}^z_{tg}$ and $\mathbf{\Theta}^z_{mr}$ the mapping parameters for the target subject and the mind-reader, respectively; μ, σ, mean and variances; $\mathbf{I}_{\mathscr{L}}, \mathbf{I}_{\mathscr{D}}$ identity matrices of dimension $\mathscr{L}, \mathscr{D}$ (respectively the reduced dimension of the latent space $\mathbf{Z}$ and the dimension of the vector representing the facial behaviour).
I then denote:

$$\mathbf{\Theta}^z = \begin{pmatrix} \mathbf{\Theta}^z_{tg} \\ \mathbf{\Theta}^z_{mr} \end{pmatrix}, \mu = \begin{pmatrix} \mu_{tg} \\ \mu_{mr} \end{pmatrix}, \mathbf{\Omega} = \begin{pmatrix} \sigma^2_{act}\mathbf{I}_{\mathscr{D}} & \mathbf{0} \\ \mathbf{0} & \sigma^2_{mr}\mathbf{I}_{\mathscr{D}} \end{pmatrix}.$$

Since the model is jointly Gaussian, then $\mathbf{x}_{tg}$ and $\mathbf{x}_{mr}$ are jointly Gaussian distributed, *i.e.* :

$$P(\mathbf{x}_{tg}, \mathbf{x}_{mr}) = \mathcal{N}(\mu, \mathbf{\Sigma}) \text{ with } \mathbf{\Sigma} = \mathbf{\Omega} + \mathbf{\Theta}^z {\mathbf{\Theta}^z}^\top$$

In order to implement Equation 4.1 it is necessary to derive the conditional distribution $P(\mathbf{X}_{mr} = \mathbf{x}_{mr} \mid \mathbf{X}_{tg} = \mathbf{x}_{tg})$, which again has a Gaussian distribution. Thus, it follows that:

$$\mathbf{x}_{mr} \mid \mathbf{x}_{tg} \sim \mathcal{N}(\widehat{\mu}_{mr}, \widehat{\mathbf{\Sigma}}_{mr}) \tag{4.7}$$

where the mean $\widehat{\mu}_{mr}$ can be computed as:

$$\widehat{\mu}_{mr} = \mu_{mr} + \mathbf{\Sigma}_c^\top \mathbf{\Sigma}_a^{-1}(\mathbf{x}_{tg} - \mu_{tg})$$

and the covariance matrix $\widehat{\mathbf{\Sigma}}_{mr}$ can be computed as the Schur complement (Zhang, 2006) of matrix $\mathbf{\Sigma} = \mathbf{\Omega} + \mathbf{\Theta}^z \mathbf{\Theta}^{z^\top}$ rewritten in the form of block matrix:

$$\widehat{\mathbf{\Sigma}}_{mr} = \mathbf{\Sigma}_b - \mathbf{\Sigma}_c^\top \mathbf{\Sigma}_a^{-1} \mathbf{\Sigma}_c \text{ with } \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_a & \mathbf{\Sigma}_c \\ \mathbf{\Sigma}_c^T & \mathbf{\Sigma}_b \end{pmatrix}$$

In summary:

1. Equation 4.4 gives the prior of the points in latent space $\mathbf{Z}$ with dimension $\mathscr{L}$;

2. Equation 4.5 gives the conditional probability of sampling a target's facial expression $\mathbf{x}_{tg}$ with dimension $\mathscr{D} \gg \mathscr{L}$ given a point $\mathbf{z}$ of the latent space $\mathbf{Z}$;

3. Equation 4.6 gives the conditional probability of sampling the mind-reader-centred facial display $\mathbf{x}_{mr}$ with dimension $\mathscr{D} \gg \mathscr{L}$ given a point $\mathbf{z}$ of the latent space $\mathbf{Z}$;

4. Equation 4.7 returns the conditional probability of sampling a mind-reader-centred facial display similar to the one observed from the target subject. To do so, it makes use of both the parameters $\mathbf{\Theta}_{tg}^z$ and $\mathbf{\Theta}_{mr}^z$ in a multivariate normal distribution as previously suggested.

4.2.2 Derivation of the First-Order Phenomenological Latent Space

For what concerns the first-order phenomenological latent space $\mathbf{\Phi}_{mr}$, the main issue here is to conceive a probabilistic latent space model in which:

(i) Either forward/inverse mapping is allowed;

(ii) The forward step is a nonlinear and continuous mapping in order to smoothly generate the variety of facial motor behaviours of the observer.

The first property is necessary to provide both forward and inverse mechanisms as suggested in Section 4.2. The second feature is required to realise a motor map

of the considered facial motor configurations having a bodily format, as described by Definition 2.25 on page 57.

Therefore, it is necessary to provide a mapping $\mathbf{x}_{mr} = g(\varphi_{mr}; \mathbf{\Theta}^{\varphi}) + \epsilon$, where ϵ is a zero-mean, isotropic, white Gaussian noise model and $g(\cdot; \mathbf{\Theta}^{\varphi})$ is a continuous non-linear function. In order to handle non-linearity, the latter can be written as a linear combination of basis functions:

$$g(\varphi_{mr}) = \sum_{j} \theta_j^{\varphi} \theta_j^{\varphi}(\varphi_{mr})$$

The problem of deriving a form for the pdf $\mathrm{P}(\mathbf{X}_{mr} \mid \mathbf{\Phi}_{mr})$ in Equation 4.3 can be formalised as

$$\check{\mathbf{x}}_{mr} \sim P(\mathbf{X}_{mr} \mid \mathbf{g}(\mathbf{\Phi}_{mr} = \varphi_{mr}; \mathbf{\Theta}^{\varphi})) \tag{4.8}$$

In this perspective, the mapping parameters of the first-order phenomenological latent space $\mathbf{\Theta}^{\varphi}$ can be learnt via the marginalisation:

$$P(\mathbf{X}_{mr} \mid \mathbf{\Phi}_{mr}) = \int P(\mathbf{X}_{mr} \mid \mathbf{g}) P(\mathbf{g} \mid \mathbf{\Phi}_{mr}) \, \mathrm{d}\mathbf{g}$$

This problem has been solved by Lawrence (2004) in terms of the Gaussian Process Latent Variable model (GPLVM), which can be expressed as a Gaussian density over the observations $\mathbf{X}_{mr}$, namely a product of Gaussian Processes (one for each of the $\mathscr{D}$ data dimensions).

Formally,

$$P(\mathbf{X}_{mr} \mid \mathbf{\Phi}_{mr}) = \prod_{d}^{\mathscr{D}} \mathcal{N}(\mathbf{x}_d; \mathbf{0}, \mathbf{K})$$

where $\mathbf{K}$ is a covariance matrix (or kernel) which depends on the q-dimensional latent variables (*cf.* Lawrence (2004) for a derivation).

As a result, an efficient closed form for the marginal likelihood can be derived (Lawrence, 2004):

$$P(\mathbf{X}_{mr} \mid \mathbf{\Phi}_{mr}) = \frac{1}{\sqrt{(2\pi)^{N\mathscr{D}} |\mathbf{K}^{\mathscr{D}}|}} \exp(-\frac{1}{2} tr(\mathbf{K}^{-1} \mathbf{X}_{mr} \mathbf{X}_{mr}^{\top})), \tag{4.9}$$

where $\mathbf{X}_{mr} = [\mathbf{x}_{1,mr}, \cdots, \mathbf{x}_{N,mr}]$ is the set of N training observations and the elements of the kernel matrix $\mathbf{K}$ are defined by a kernel function $(\mathbf{K})_{i,j} = \mathcal{F}(\varphi_{i,mr}, \varphi_{j,mr})$ (implemented in the following Section 4.3 as Radial Basis Function kernel).

Once the latent variable model has been learnt it is straightforward to obtain the inverse pdf $P(\mathbf{\Phi}_{mr} \mid \mathbf{X}_{mr})$ in Equation 4.2 either through GPLVM standard inversion (Lawrence, 2004), or by Monte Carlo sampling approximation from Equation 4.9.

Eventually, the matching process summarised by the mapping function $\mathcal{M}$ can be conceived of as an optimised search in the mind-reader's first-order phenomenological latent space for determining the optimal state $\mathbf{\Phi}_{mr} = \varphi^{\star}_{mr}$ that maximises the similarity between the mind-reader facial expression and one currently generated by using Equation 4.9 (further details in Section 4.3).

4.3 Model Implementation

The overall simulation scheme is outlined in Figure 4.2. The perceptual input $\mathbf{x}_{tg}$, namely the observed facial expression of the target subject, is transcoded via the self-projected latent space $\mathbf{Z}$ in the mind-reader's egocentric representation $\mathbf{x}_{mr}$. This mapping is suitably learnt so that $\mathbf{x}_{mr}$ will exhibit the same dynamic features displayed in $\mathbf{x}_{tg}$ (*e.g.* facial muscles configuration), but discarding invariant features of the observed face (*i.e.* the identity appearance of the target subject).

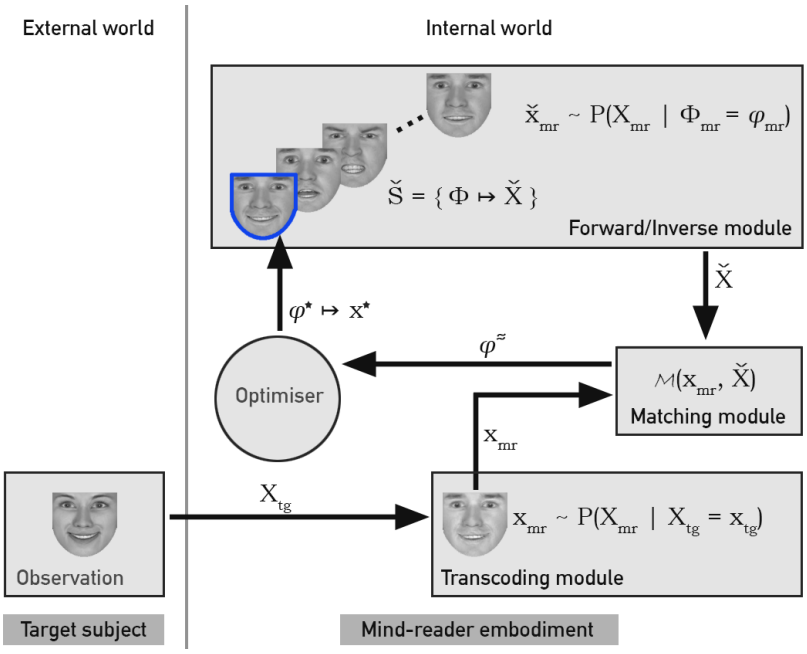


Figure 4.2 The schema of the proposed computational model

Denote $\check{\mathbf{S}} = \{\mathbf{\Phi}_{mr} \mapsto \check{\mathbf{X}}\} = \{\varphi^{1}_{mr} \mapsto \check{\mathbf{x}}^{1}_{mr}, \ldots, \varphi^{N}_{mr} \mapsto \check{\mathbf{x}}^{N}_{mr}\}$ a set of N samples from the mind-reader's first-order phenomenological space $\mathbf{\Phi}$ and associated with

the mind-reader's expressions generated through Equation 4.9. Along the matching process $\mathcal{M}$, a similarity measure is used in order to evaluate the likelihood between the samples in $\check{\mathbf{S}}$ and the egocentric observation $\mathbf{x}_{mr}$. Thus, given $\mathbf{x}_{mr}$ sampled from Equation 4.1, the initial state $\varphi^{\approx}$ is selected as:

$$\varphi^{\approx} \mapsto \check{\mathbf{x}}_{mr} \in \check{\mathbf{S}} \mid \check{\mathbf{x}}_{mr} = \max \mathcal{M}(\mathbf{x}_{mr}, \check{\mathbf{X}}) \tag{4.10}$$

Such choice is refined by resorting to a search optimisation process in the first-order phenomenological latent space leading to an optimal internal representation $\varphi^{\star}_{mr}$ of the observed target subject's expression $\mathbf{x}_{tg}$.

In the following I provide some implementation details and preliminary results of the proposed system.

4.3.1 The Transcoding Module

The aim of this mapping is to generate a self-centred expression $\mathbf{x}_{mr}$ exhibiting the same facial expression displayed in $\mathbf{x}_{tg}$, but replacing the identity of the target subject with the one of the mind-reader, as previously proposed with Equation 4.7.

Equation 4.6 creates subspaces of $\mathbf{Z}$ for each considered facial expression of the mind-reader and that Equation 4.5 does the same for the facial expression of the target subject (Mohammadzade and Hatzinakos, 2013).

Since I start from the assumption that both Equation 4.5 and Equation 4.6 share the same latent space $\mathbf{Z}$ and the same set of facial expressions, it is likely that similar facial expression of the mind-reader and the target subject would be clustered on close regions of the latent space $\mathbf{Z}$ (Calder et al., 2001; Turk and Pentland, 1991).

Thus, the process of parameter learning can be simplified by employing a Principal Component Analysis (PCA) over a training set of the mind-reader face displaying different expressions and using the estimated mapping parameters in order to project the facial expression of the target subject onto the latent space and back to the original space, thus obtaining a new observation exhibiting the mind-reader identity, but maintaining the target subject's facial expression as illustrated by Figure 4.3.

This procedure is dual to that recently proposed by Mohammadzade and Hatzinakos (2013) (to which I refer for more technical details), where images of different subjects' faces with the same facial expression are located in a common subspace; here, instead the same facial expression is maintained, changing the identity of the subject into a desired one.

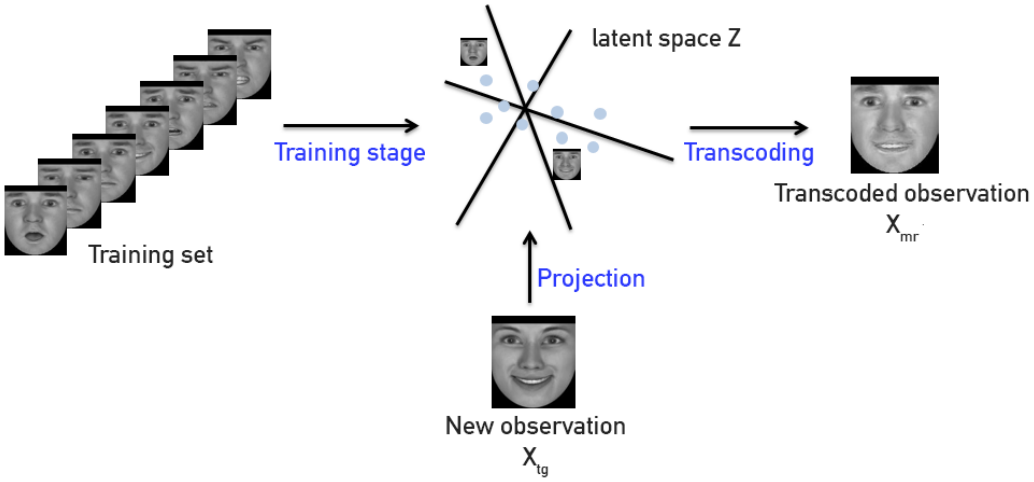


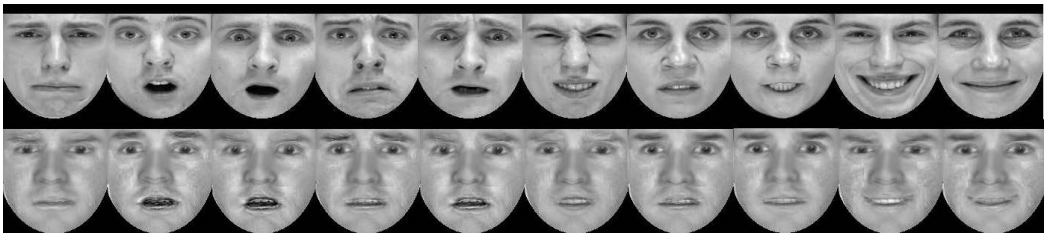
Figure 4.3 Diagram of the transcoding process. Observations of the mind-reader motor configurations are used to infer the self-projected latent space. A novel face stimulus can be projected onto this latent space, thus obtaining the corresponding egocentric face stimulus exhibiting a similar facial motor configuration, although centred on the mind-reader body.

Importantly, for this process, it is only necessary a *training set* of different facial expressions exhibited by the same subject, which can be either a specific identity (*e.g.* mind-reader identity) or a prototypical average face. However, since the training process may lead to a synthesis error, this error can be reduced by using an additional validation set of face stimuli from other identities exhibiting facial configurations very similar to the one included in the training set (for more details see Mohammadzade and Hatzinakos (2013) and Vitale et al. (2014b)).

In order to further improve the transcoding performance and reducing the number of the required training images, the face is split in semantic/spatial parts (*i.e.* the eyebrows, the eyes, the nose, the mouth and the cheeks), thus estimating several self-projected spaces, one for each of these face parts.

Figure 4.4a shows examples of projected synthetic images, whereas Figure 4.4b shows examples of projected real images (from the MMI-Facial Expression Database collected by Pantic et al. (2005); Valstar and Pantic (2010)). To generate such images I used a training set and validation set including only 10 synthetic facial expressions. The resulted self-centred images exhibit facial expressions resembling the ones displayed by the input stimuli, although having the identity of the training subject. This is an important feature since it suggests that the realised latent space can captures abstract aspects of the exhibited facial expressions that are independent of identity.

(a) Some examples of synthetic faces transcoded using the self-projected latent space. In the top row the observations, in the middle row the transcoded faces with the identity removed and the facial configuration preserved, whereas in the bottom row the ground truth images.



(b) Some examples of real faces transcoded using the self-projected latent space. In the top row the observations, whereas in the bottom row the projected image with the identity removed and the facial configuration preserved.

Figure 4.4 Examples of synthetic and real faces transcoded in self-centred stimuli.

4.3.2 The Forward/Inverse Module

The GPLVM introduced in Section 4.2 has the capability of learning with few samples and of generating smooth dynamics between points of the latent space (if an appropriate kernel is used, in this case a Radial Basis Function). These are two desired characteristics in order to obtain bodily formatted representations for the considered face stimuli. In fact, as I will show in the remainder of this section, it is possible to implement the desired latent space by using a training set including only the motor samples of the training subject (*i.e.* the cogniser's own body as suggested by Definition 2.23 on page 56) and, given the properties of the kernel function, the computational tool can realise a latent space constrained by aspects of the body (*i.e.* the possible facial configurations) as required by Definition 2.25 on page 57.

The GPLVM model was implemented in the form of a Hierarchical Gaussian Process Latent Variable Model (HGP-LVM) (Lawrence and Moore, 2007). The HGP-LVM is an extension of the original GPLVM (Lawrence, 2004). I use this

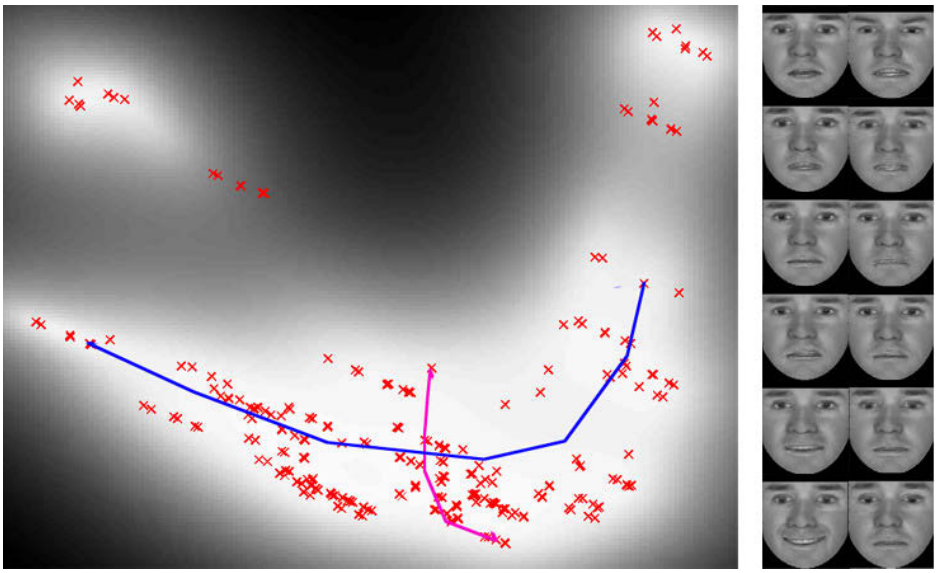


Figure 4.5 Example of a first-order phenomenological space realised by using HGP-LVM. On the right examples of 12 sampled generated images from the trajectories highlighted on the shown latent space: the samples in the left column comes from the 'top-to-bottom' trajectory, whereas the samples in the right column comes from the 'right-to-left' trajectory.

tool to introduce hierarchical constraints based on the different parts composing the face (*i.e.* eyebrows, eyes, nose, mouth and cheeks).

Using this model, it is possible to sample a vocabulary $\mathbf{S} : \mathbf{\Phi} \mapsto \{\mathbf{X}^{\mathbf{j}}_{\mathbf{mr}}\}^{\infty}_{\mathbf{j=1}}$ generating an infinite set of synthetically simulated facial expressions of a specific desired identity from low-dimensional representations. A subsample $\check{\mathbf{S}}$ of such vocabulary is used by the matching module to determine the starting condition via Equation 4.10 (Figure 4.2).

Hence, each point of the latent space represents an internal phenomenological state φ in bodily format. Such representation has dimension q (in the present work 2-dimensional). The likelihood between the projected observation $\mathbf{x}_{mr}$ and the observations generated from sampled latent points $\check{\mathbf{x}}$ provides an approximation of the conditional distribution $\mathbf{P}(\mathbf{X} \mid \mathbf{\Phi_{mr}} = \mathbf{x_{mr}})$, that represents the activation levels of the topology of the motor map.

In Figure 4.5 is shown a latent space generated from a high number of different facial expressions (238) and examples of generated facial expressions from points of such latent space.

4.3.3 Matching Module and Optimisation

Since in this implementation the observed stimuli are in the form of images, to operationalise the matching function $\mathcal{M}$ I use as similarity index the Structural Similarity (SSIM) measure (Wang et al., 2004). On the basis of the statistical model behind the mapping, one can conceive the transcoded images as noisy representations of the ground truth images; in this perspective, the SSIM was shown to be a consistent measure (Wang et al., 2004). Additional tests with different similarity measures (*e.g.* Pearson's correlation measure) produced poorer results, further motivating this choice. By taking into account future real time applications of the suggested model, measures with a high computational cost were not considered.

The SSIM metric is processed among different windows of an image, and then the average among them is used as a final measure. The measure between two windows x and y of size $N \times N$ (in this model a Gaussian window of 80×80 pixels with sigma 3) is given by:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4.11}$$

Where μ_x and μ_y are respectively the means of x and y, σ_x^2, σ_y^2 and σ_{xy} are respectively the variances of x and y and the covariance of x and y, c_1 and c_2 are two constants to stabilise the division with weak denominator (Wang et al., 2004).

Since only a finite number of points of the first-order latent space is available (and consequently the corresponding synthetically simulated images), it is possible to use the criteria in Equation 4.10 to select a point of the first-order latent space as the initial condition of an interior–point method optimisation (Nesterov et al., 1994). This optimisation method can refine the selected position $\varphi^{\approx}$ within the latent space in order to find the closer position $\varphi^{\star}$ that generates the image with the maximum value of SSIM respect to $\mathbf{x}_{mr}$.

4.4 Model Evaluation

The aim of this chapter is to demonstrate that the realised first-order phenomenological latent space, together with the self-projected latent space, is a plausible computational account of embodied simulation mechanisms, as suggested by Gallese and Caruana (2016). Namely, I will demonstrate that the present imple-

Figure 4.6 The training images used for the tests of our architecture. From left to right: anger, annoyance, delight, fake smile, fear, neutral, sadness, smile, surprise, and wonder.

mentation of the model is able to map perceptual face stimuli into sensory-motor representations of it, shaped in bodily formats (Goldman, 2013).

Hence, in this section I evaluate the model considering two aspects:

1. **Quantitative assessment of the self-centred mapping.** I will show the ability of the self-projected latent space to preserve the dynamic features of the observed face stimuli (*i.e.* facial expression), while at the same time replacing the observed identity with the desired one. This evaluates the computational plausibility of the proposed transcoding process (Equation 4.1) resembling embodied mechanisms;

2. **Quantitative assessment of classification performance.** I will evaluate the performance of the present model (self-projected and first-order latent spaces) to promote the classification of facial expressions. This evaluates the computational plausibility of the present model and theory in enhancing the performance of a face expression recognition task, in comparison with basic features matching strategies.

4.4.1 Dataset

The dataset has been generated by using the FaceGen software[6]. Synthetically generated images were used to facilitate the evaluation of the results, as with the currently available datasets of face stimuli it was not possible to have a sound validation set; as a matter of fact, the subjects in the available datasets do not display the very same facial configuration or viewpoint, even when the facial expression is attributed to the same concept. This makes impossible to find conclusive correspondences between observations. These correspondences are critical for assessing the present model.

The dimension of the images used was 140×154 pixels. In order to get a range of distinctive expressions, I considered only 10 facial configurations resembling

[6]http://www.facegen.com/

10 corresponding concepts, namely anger, annoyance, delight, fake smile, fear, neutral, sadness, smile, surprise and wonder (Figure 4.6). These labels were subjectively attributed to specific motor configurations of the face stimuli, realised choosing specific parameters of FaceGen software.

These facial expressions were selected to be the most representative among the ones obtainable by using FaceGen. The synthesis process required to select the whole list of emotional expressions available in the software (*i.e.* specific motor configuration modifying a set of facial muscles suggesting facial expressions of basic emotions) plus further combinations between them. Some additional features were added to the basic emotional expressions when necessary; for example, facial expressions of sadness are decoded by focusing on the eye region of the subject (Eisenbarth and Alpers, 2011), hence sadness expression included a 50% looking down tilt of the eyes. Importantly, this list of expressions is not limited to expressions of basic emotions. In fact, here it is important to address the problem of modelling a more general discrimination of facial movements. For this reason, the considered list included expressions exhibiting similar facial movements difficult to discriminate, such as the fake smile (Ekman et al., 1990) and the smile, neutral and sadness, and surprise and wonder.

The dataset includes:

- A training set of 10 images, namely a subject exhibiting all the selected facial expressions (Figure 4.6);
- A validation/test set of 28 subjects, each one exhibiting the 10 selected facial expressions (some examples in the top row of Figure 4.4a);

A 4-fold cross validation approach was used. Thus for each test I split the validation/test set in 2 sets: a validation set of 21 subjects ($21 \times 10 = 210$ validation images) and a test set of 7 subjects ($7 \times 10 = 70$ test images).

The training set was used by the transcoding module and by the forward/inverse process to estimate mapping parameters. The validation sets were exploited to reduce the synthesis error of the transcoding module. The test sets were used to obtain the results provided in the next sections.

Figure 4.7 illustrates an example of the realised first-order latent space's topology, obtained by using the 10 facial expressions of the mind-reader. Here it is important to note that in this space the facial expressions are located based on their visual similarities. Therefore, features like arousal and valence characterising such observations are not clearly distinguishable in this model. However, Boccignone

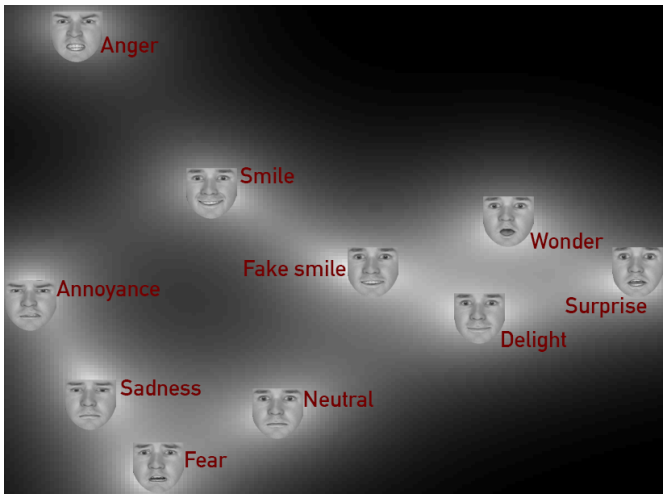


Figure 4.7 The topology of the first-order phenomenological latent space resulting from the used training set. Perceptually similar facial configurations are clustered in nearby regions of the space, although maintaining a certain separation facilitating their classification.

et al. (2018) recently demonstrated that by extending the model discussed in this chapter to accommodate the visceral information accompanying the observed face stimuli it is possible to generate a core affect state-space representing more precise valence and arousal information.

4.4.2 Quality of Self-Centred Mapping

Using the measure introduced in Section 4.3.3, the evaluation process requires to determine the structural similarity between the egocentric images $\mathbf{x_{mr}} \sim P(\mathbf{X}_{mr} \mid \mathbf{X}_{tg})$ and the associated ground truth images $g \in G$ by using the measure introduced in Section 4.3.3. As a baseline for comparisons, I evaluated the structural similarity between the raw observations $\mathbf{x}_{tg}$ and the associated ground truth images $g \in G$.

The 4-fold cross validation test produced the results summarised in Table 4.1 and in Figure 4.8.

The maximum similarity for the selected measure (SSIM) can be 1 if and only if the two images under evaluation are identical. This means that a similarity measure of 1 between the realised self-centred image and the corresponding ground truth image would imply a perfect embodiment of the observed face stimulus.

Hence, I define the *embodiment deficit* to be the difference between 1 and the gathered similarity measure. This measure conveys the deficit underlying embodiment of sensory-motor features of the face stimuli in an egocentric representation.

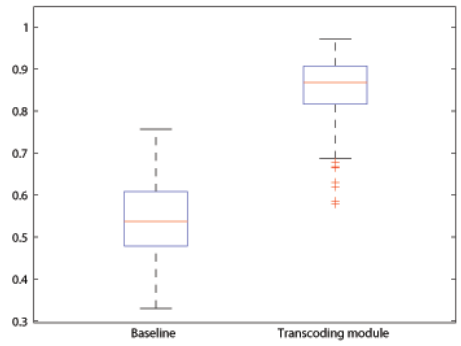


Figure 4.8 Quality of self-centred mapping (boxplot). A higher result underlies desirable better quality, with 1 denoting optimal quality.

Case	Mean	Median	Std
Baseline	0.5425	0.5370	0.0871
Model	0.8525	0.8674	0.0713

Table 4.1 Quality of self-centred mapping evaluated as similarity likelihood to ground truth images. The highest similarity can be 1 for identical images and 0 for completely dissimilar images.

By comparing the embodiment deficits between the baseline strategy and the proposed model after transcoding module processing, the average embodiment deficit was reduced from 0.46 to 0.15. This means that the proposed model promotes embodiment of face stimuli by approximately 68%, compared to a basic features matching strategy (*i.e.* baseline).

4.4.3 Classification Performance

In this chapter I did not extensively investigate the attribution process, that highly depends on contextual and cultural factors of the mind-reader. Furthermore, the implementation of this model is limited to provide the sensory-motor dimension of bodily formatted representations; as I discussed in Chapter 3 (Section 3.6 on page 98), this dimension may not be sufficient to promote fully functioning facial expression recognition mechanisms and emotional empathy. Indeed, integrating a visceral dimension may significantly improve accuracy, especially for opaque and subtle facial expressions. However, in this section I can still demonstrate that using this model (and so its representations and underlying processes), instead of an alternative basic features matching strategy, will significantly enhance the accuracy of a facial expression recognition task.

I measured the classification performance with respect to the 10 possible facial expressions included in the considered dataset.

Given a set $G = \{g_j\}_{j=1}^{10}$ of ground truth images for each class j, these have a corresponding set of latent positions $\varphi_{\mathbf{G}}^{\mathbf{j}}$ in the first-order latent space.

Given the optimal latent point $\varphi^\star$ estimated from a new observation $\mathbf{x}_{tg}$, via Equations 4.1, 4.2, and 4.3, the concept c underlying the representation $\varphi^\star$ (and so of the observation $\mathbf{x}_{tg}$) is j of $\varphi^{\mathbf{j}\star}_{\mathbf{G}}$ spatially closer to $\varphi^\star$. In other words, each ground truth image corresponding to a specific facial expression acts as the centroid of a specific region in the first-order latent space (see Figure 4.7). If the experienced phenomenological representation happens to be in this region or close to it, the attributed concept would be the one associated with the centroid of such region.

For each test of the 4-fold cross validation process, I provide a confusion matrix of the 70 test images classified. Thus, the four confusion matrices were summed, obtaining a new overall confusion matrix of the 280 images used to verify the implemented model during the 4-fold cross validation process.

For the baseline approach, the concept attributed to the observation $\mathbf{x}_{tg}$ is the concept associated with the ground truth image in $\mathbf{G}$ that is more structurally similar to $\mathbf{x}_{tg}$. This means that the baseline approach is based on a features matching strategy, without the mediation of embodied mechanisms. Thus, to estimate the confusion matrix of the baseline approach, each observation $\mathbf{x}_{tg}$ of the validation/test set was classified using the structural similarity measure with respect to the ground truth images in $\mathbf{G}$.

Table 4.2 and Table 4.3 show the confusion matrices respectively of the baseline strategy and the embodiment strategy proposed in this chapter. An optimal classification would result in a confusion matrix with all zeros except for the diagonal (where in this case the optimal score is 28).

Furthermore, Table 4.4 provides the sensitivity, specificity, accuracy, precision and negative predictive value (NPV) of, for both the baseline method and the proposed model. Precision, accuracy and sensitivity of the two approaches in each considered facial configuration are illustrated in Figure 4.9.

The proposed embodied mechanisms of the present model overcome the features matching approach of the baseline strategy with respect to all the considered measures. In particular, the proposed model shows a critical gain in accuracy for the expression of delight. Table 4.2 suggests that the delight expression was misunderstood as neutral or sadness by the baseline approach only relying on visual features. On the contrary, the topology realised by the proposed model was able to facilitate the discrimination of the delight expression, since in a region of the space separated from the one of neutral and sadness expressions.

In order to verify a statistically significant enhancement of the performance, I tested the considered measures with a paired Wilcoxon signed-rank test (Wilcoxon, 1945), as the assumption of normality required by other statistical tests was not

Expression	Anger	Annoyance	Delight	fake Smile	Fear	Neutral	Sadness	Smile	Surprise	Wonder
Anger	20	1	3	4	0	0	0	0	0	0
Annoyance	0	11	7	2	1	2	0	0	3	2
Delight	0	0	25	1	0	0	0	0	2	0
fake Smile	1	0	3	18	0	0	0	6	0	0
Fear	0	0	4	0	16	3	0	0	4	1
Neutral	0	0	15	0	0	9	2	0	2	0
Sadness	0	0	14	0	0	11	2	0	1	0
Smile	1	0	6	9	0	0	0	12	0	0
Surprise	0	0	5	0	1	5	0	0	17	0
Wonder	0	0	0	0	1	6	0	0	4	17

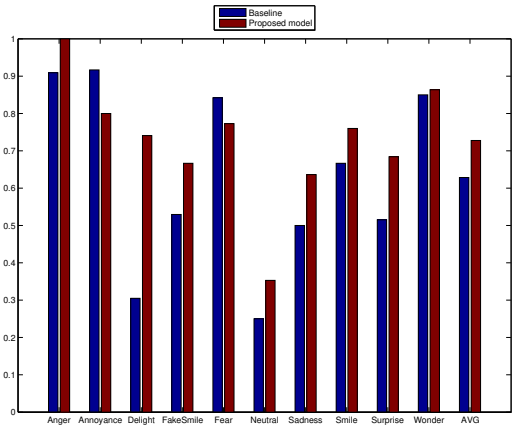
Table 4.2 Confusion matrix using the baseline strategy.

Expression	Anger	Annoyance	Delight	fake Smile	Fear	Neutral	Sadness	Smile	Surprise	Wonder
Anger	23	1	1	2	0	0	1	0	0	0
Annoyance	0	24	0	0	0	2	1	0	0	1
Delight	0	0	20	1	0	5	0	0	2	0
fake Smile	0	0	0	22	0	0	0	6	0	0
Fear	0	1	1	0	17	5	0	0	3	1
Neutral	0	0	1	0	1	24	2	0	0	0
Sadness	0	2	2	0	0	17	7	0	0	0
Smile	0	0	1	8	0	0	0	19	0	0
Surprise	0	1	1	0	2	10	0	0	13	1
Wonder	0	1	0	0	2	5	0	0	1	19

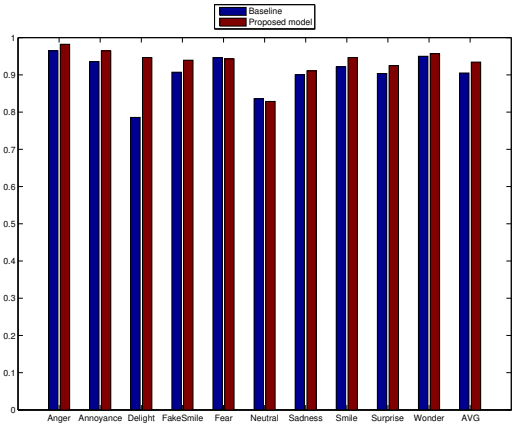
Table 4.3 Confusion matrix using the embodied mechanisms suggested in this chapter.

Case	Sensitivity	Specificity	Accuracy	Precision	NPV
Baseline	0.5250	0.9472	0.9050	0.6283	0.9483
Proposed model	0.6714	0.9635	0.9343	0.7277	0.9641
p-values	.0537	.7646	**.0049**	**.0322**	.0674

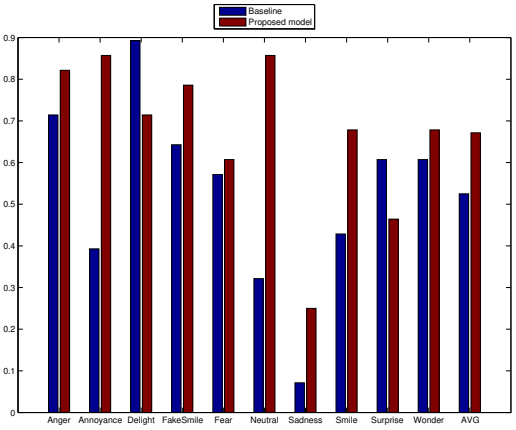
Table 4.4 Results of classification performance between a basic features matching strategy (baseline) and an embodiment strategy (proposed model). The results are the averages of the 10 considered concepts.



(a) Precision



(b) Accuracy



(c) Sensitivity

Figure 4.9 Classification performance of the proposed model.

satisfied. The increment in accuracy and precision resulted statistically significant (respectively p-values = .0049 and .0322). There was not enough evidence to suggest significant differences for sensitivity (p-value = .0537), NPV (p-value = .0674) and specificity (p-value = .7646). These results demonstrate that the proposed computational model and underlying theory of embodiment can better promote the accuracy and precision of facial expression recognition, as compared to basic features matching strategies.

4.5 Conclusions

In this chapter, I investigated notable accounts of Simulation Theories underlying mechanisms of embodiment. I provided a theoretical and computational model of the presented accounts. The current investigation was limited to the sole process of mapping external face stimuli of a target to its corresponding phenomenological representation, bodily in format, of the mind-reader. The development of such computational model advances the argument of this thesis, by providing evidence for computationally plausible mechanisms of embodied simulation during face-to-face interactions.

In this work, I suggested that a simulation-style mind-reading process can plausibly underlie embodied simulation mechanisms. I proposed that these embodied mechanisms can be operationalised by two modules: one mapping the perceived stimuli into self-centred images of the stimuli, and one mapping the realised egocentric representations onto a first-order phenomenological latent space having bodily format.

Significantly, the proposed model is general enough to be extended to other behaviours, as long as suitable data is provided for its implementation. Gestures, posture and vocal cues can be implemented in the model with a similar procedure and mapped to other self-projected and first-order phenomenological latent spaces, or aggregated in a single multimodal latent space (Boccignone et al., 2018).

In fact, these signals can be treated as vectors, similarly to the used facial expression images. Then, for example, it is likely that a self-projected latent space of a speech signal would map the pitch of the voice of the target subject to that of the mind-reader, maintaining the same principal frequencies. A similar outcome can be realised for gesture and posture signals, where physical constraints of the target subject's body would be mapped to those of the mind-reader's body.

More importantly, the visual stimulus sensed by the retina passes through a complex neural system of the visual cortex, which extracts primitive features representing the observed stimulus. In this work, I used raw pixels of the face stimulus as input to my model. However, the face images can be processed with filters resembling human visual cortex capabilities, such as Gabor filters (Jones and Palmer, 1987), thus using the new extracted features as input to the present model.

Each considered modality can generate a self-projected latent space and the associated first-order phenomenological latent space. Such internal representations in bodily formats, shared among different modalities and easily treatable from a probabilistic point of view, might solve the challenging problem of multimodality integration in computational affect systems (Boccignone et al., 2018; Zeng et al., 2009), or provide insights to understand self-disorders (Gallese and Ferri, 2013), suggested to arise from impairments in multisensory integration (Sestito et al., 2015) (see Section 3.4.2 for a discussion). In addition, the proposed model can be easily extended to consider temporal dynamics. Such extension might make use of tools like Markov Chains or Dynamic Bayesian Networks, or more generally it is possible to use temporal sequences of stimuli as inputs to the model, instead of single instances of them.

Finally, it is important to acknowledge some limitation of the present study. First, the considered facial expressions are well aligned frontal faces. However, this is not a major problem to advance the general argument of this dissertation, since, as I discussed previously, the model and theory are general enough to employ different inputs and implementation strategies, potentially making use of more sophisticated and precise computational tools (Boccignone et al., 2018). Secondly, in this work, I used synthetic face stimuli. However, in Figure 4.4b I provided preliminary qualitative results showing that it is possible to also transcode real face stimuli and reach an acceptable quality. It is likely to expect that for real face stimuli it would be necessary to implement the transcoding process in a different way, for instance by using non-linear algorithms or a different set of features (*i.e.* not the raw pixels).

Chapter Bibliography

Adelmann, P. K. and Zajonc, R. B. (1989). Facial efference and the experience of emotion. *Annual Review of Psychology*, 40(1):249–280.

Adolphs, R. (2002a). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12(2):169–177.

Adolphs, R. (2002b). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, 1(1):21–62.

Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., and Lanzarotti, R. (2018). Deep construction of an affective latent space via multimodal enactment. *IEEE Transactions on Cognitive and Developmental Systems*.

Brothers, L. (2002). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Foundations in Social Neuroscience*, pages 367–385.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., and Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision research*, 41(9):1179–1208.

Calder, A. J., Keane, J., Cole, J., Campbell, R., and Young, A. W. (2000). Facial expression recognition by people with möbius syndrome. *Cognitive Neuropsychology*, 17(1-3):73–87.

De Vignemont, F. and Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, 10(10):435–441.

Demiris, Y. and Johnson, M. (2003). Distributed, predictive perception of actions: A biologically inspired robotics architecture for imitation and learning. *Connection Science*, 15(4):231–243.

Dimberg, U. and Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology*, 39(1):39–45.

Dindo, H., Zambuto, D., and Pezzulo, G. (2011). Motor simulation via coupled internal models using sequential monte carlo. In *Proceedings of the International Joint Conference on Artificial Intelligence*, volume 22, page 2113.

Eisenbarth, H. and Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion*, 11(4):860.

Ekman, P., Davidson, R. J., and Friesen, W. V. (1990). The duchenne smile: Emotional expression and brain physiology: Ii. *Journal of personality and social psychology*, 58(2):342.

Gallese, V. (2001). The 'shared manifold' hypothesis. From mirror neurons to empathy. *Journal of Consciousness Studies*, 8(5-6):33–50.

Gallese, V. (2003). The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431):517–528.

Gallese, V. (2007). Before and below 'theory of mind': Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):659–669.

Gallese, V. (2016). Finding the body in the brain. From simulation theory to embodied simulation. In McLaughlin, B. P. and Kornblith, H., editors, *Goldman and His Critics*, pages 299–314. John Wiley & Sons.

Gallese, V. and Caruana, F. (2016). Embodied simulation: Beyond the expression/experience dualism of emotions. *Trends in Cognitive Sciences*.

Gallese, V. and Ferri, F. (2013). Jaspers, the body, and schizophrenia: The bodily self. *Psychopathology*, 46(5):330–336.

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8(9):396–403.

Gallese, V. and Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, 15(11):512–519.

Goldman, A. I. (2013). The bodily formats approach to embodied cognition. *Current Controversies in Philosophy of Mind*, page 91.

Goldman, A. I. and Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition*, 94(3):193–213.

Greco, A., Valenza, G., Citi, L., and Scilingo, E. P. (2016). Arousal and valence recognition of affective sounds based on electrodermal activity. *IEEE Sensors Journal*.

Jack, R. E. and Schyns, P. G. (2015). The human face as a dynamic tool for social communication. *Current Biology*, 25(14):R621–R634.

Jones, J. P. and Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258.

Keillor, J. M., Barrett, A. M., Crucian, G. P., Kortenkamp, S., and Heilman, K. M. (2002). Emotional experience and perception in the absence of facial feedback. *Journal of the International Neuropsychological Society*, 8(1):130–135.

Lambie, J. A. and Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: A theoretical framework. *Psychological Review*, 109(2):219.

Lawrence, N. D. (2004). Gaussian process latent variable models for visualisation of high dimensional data. *Advances in Neural Information Processing Systems*, 16(329-336):3.

Lawrence, N. D. and Moore, A. J. (2007). Hierarchical gaussian process latent variable models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 481–488. ACM.

Levenson, R. W., Ekman, P., and Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, 27(4):363–384.

Lopes, M. and Santos-Victor, J. (2005). Visual learning by imitation with motor representations. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(3):438–449.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, New York.

Meltzoff, A. N. and Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, pages 702–709.

Mohammadzade, H. and Hatzinakos, D. (2013). Projection into expression subspaces for face recognition from single sample per person. *IEEE Transactions on Affective Computing*, 4(1):69–82.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

Nesterov, Y., Nemirovskii, A., and Ye, Y. (1994). *Interior-Point Polynomial Algorithms in Convex Programming*, volume 13. SIAM.

Ojha, S., Vitale, J., and Williams, M.-A. (2017). A domain-independent approach of cognitive appraisal augmented by higher cognitive layer of ethical reasoning. In *39th Annual Meeting of the Cognitive Science Society*, pages 2833–2838.

Pantic, M. and Bartlett, M. S. (2007). Machine analysis of facial expressions. In Delac, K. and Grgic, M., editors, *Face Recognition*, pages 377–416. I-Tech Education and Publishing.

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. (2005). Web-based database for facial expression analysis. In *IEEE International Conference on Multimedia and Expo*. IEEE.

Ringeval, F., Sonderegger, A., Sauer, J., and Lalanne, D. (2013). Introducing the recola multimodal corpus of remote collaborative and affective interactions. In *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pages 1–8. IEEE.

Sestito, M., Raballo, A., Umiltà, M. A., Leuci, E., Tonna, M., Fortunati, R., De Paola, G., Amore, M., Maggini, C., and Gallese, V. (2015). Mirroring the self: Testing neurophysiological correlates of disturbed self-experience in schizophrenia spectrum. *Psychopathology*, 48(3):184–191.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.

Valstar, M. and Pantic, M. (2010). Induced disgust, happiness and surprise: An addition to the MMI facial expression database. In *Proceeding of the International Conference on Language Resources and Evaluation, Workshop on Emotion*, pages 65–70.

Vitale, J., Williams, M.-A., and Johnston, B. (2014a). Socially impaired robots: Human social disorders and robots' socio-emotional intelligence. In *6th International Conference on Social Robotics*, pages 350–359.

Vitale, J., Williams, M.-A., Johnston, B., and Boccignone, G. (2014b). Affective facial expression processing via simulation: A probabilistic model. *Biologically Inspired Cognitive Architectures*, 10:30–41.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Watanabe, A., Ogino, M., and Asada, M. (2007). Mapping facial expression to internal states based on intuitive parenting. *Journal of Robotics and Mechatronics*, 19(3):315.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, 1(6):80–83.

Wolpert, D. M. and Flanagan, J. R. (2001). Motor prediction. *Current Biology*, 11(18):R729–R732.

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58.

Zhang, F. (2006). *The Schur complement and its applications*, volume 4. Springer Science & Business Media.

The perception of identity is so intimately bound up with the perception of the human form.

— David Hanson —

5

Face Identity Discrimination via a Dual Face-Space[1]

In Chapter 3 I suggested that embodiment is at the core of social cognition and vital for promoting face processing mechanisms. embodied mechanisms provide representations of face stimuli bodily in their format. I suggested that these representations exhibit two dimensions: a sensory-motor dimension, providing information about dynamic features of the face, and a visceral dimension, providing information about the emotional reactions associated with the observed face stimuli.

In Chapter 4 I focused on providing a computational model of embodied simulation mechanisms able to demonstrate a plausible realisation of the suggested sensory-motor dimension for face stimuli. I showed that this model provides embodied representations shaped in bodily formats. Furthermore, I showed that this dimension is beneficial in enhancing facial expression recognition accuracy, compared to a basic feature matching strategy. In addition, in Hypotheses 3.1 and 3.2, I suggested that the sensory-motor dimension of the realised bodily representations

[1]This chapter is an adaptation of *"Vitale, J., Williams, M.-A., and Jonhston, B. (2016). The face-space duality hypothesis: A computational model. In 38th Annual Meeting of the Cognitive Science Society , pages 514–519"*.

is not only necessary to facilitate facial expression recognition, but it interacts also with facial identity discrimination mechanisms.

The '*face-space*' framework (Valentine, 1991) is a widely used tool in face perception and processing research able to explain many of the phenomena underlying facial identity discrimination in both human experimental settings (Lee et al., 2000; Rhodes et al., 2011) and computational simulations (Calder et al., 2001; Vitale et al., 2016, 2017). This framework is so important in face studies that it is *"virtually impossible to explain the interactions between the computational and cognitive approaches to understanding face recognition without reference to this model. It serves as the glue that binds the theoretical and computational aspects of the problem together."* (Calder, 2011, page 17).

Valentine (1991) used formal models of concept representations to propose that faces are represented in a psychologically plausible multidimensional space, *i.e.* the *face-space*. Faces are points of this space based on their perceived properties. This structure can plausibly account for coding *identity-related* features, such as sex, distinctiveness, age and attractiveness (Calder et al., 2001; Valentine et al., 2015). For example, the feature 'eyebrows' can vary from marked to delicate, thus possibly being one of the perceivable features crucial for coding the sex of a face. Hence, the face-space framework is a particularly valuable tool to investigate facial identity discrimination mechanisms in humans.

Unfortunately, dynamic aspects of faces, such as the ones modelled in Chapter 4, were neglected in the traditional face-space account. This significant limitation prevents the analysis of the interactions potentially happening between facial expression and facial identity recognition mechanisms. In addition, this gap precludes from providing a link between the model proposed in Chapter 4 and facial identity discrimination mechanisms of the face-space framework, necessary to validate the thesis argument of the present dissertation.

Thus, in this chapter I will answer the following questions necessary to advance the proposed thesis argument:

> *Can Valentine's framework be extended to support codings integrating both dynamic and invariant facial features and facilitating facial expression and identity recognition?*
> *If so, how can the structure of this space be possibly shaped to accommodate this interdependent nature of dynamic and invariant facial features?*

How can the hypothesised structure be plausibly implemented in a computational model?

Answering these questions is significantly important to link face processing mechanisms to embodied mechanisms. As I will argue in Chapter 7, this outcome will provide computational evidence able to advance an understanding unifying traditional and modern studies in face processing and promoting an integration of cognitive science and embodied cognition research. Therefore, in this chapter, I will show that a *single* face-space can exhibit a structure that supports both identity and facial expression recognition. I refer to this twofold structure as a *dual face-space*. The structure of this dual face-space can be realised in a parsimonious way by integrating both invariant and dynamic features of the face stimuli in a single multidimensional representation. I will demonstrate the computational validity of the suggested *duality hypothesis* through a rigorous mathematical presentation and related experiments.

5.1 Background

Valentine's face-space framework (Valentine, 1991) is a notable cognitive model for face representation. According to this framework, facial representations are encoded in a multidimensional psychological space. The dimensions of this space are assumed to encode properties of the facial signals that better discriminate one face from another. The distance between two representations underlies their dissimilarity from a psychological perspective.

Identity and expression are two forms of facial information crucial for many social skills (Grossmann, 2015). Identity recognition is suggested to arise from coding invariant features of face stimuli, whereas facial expression is proposed to result from coding dynamic facial features (Fitousi and Wenger, 2013). Early brain lesion and neuroimaging studies suggested that face identity and expression are independent dimensions (Tranel et al., 1988). Studies suggested a complete separation of identity and expression systems after the completion of a structural encoding stage and that identity and expression are represented and processed by separate systems that process faces in parallel (Bruce and Young, 1986; Bruyer et al., 1983). Accordingly, Haxby et al. (2000) argue that invariant features, such as identity, and dynamic features, such as facial expression, are computed by separate regions of the brain.

However, new evidence from recent findings indicates that these systems operate interdependently (Pell and Richards, 2013), thus suggesting that identity and expression are more closely connected than previously thought. Indeed, it has been argued that common-codings of face stimuli can respond to both identity and expression related features, suggesting that the underlying processes can indeed interact (Ganel et al., 2005; Kadosh et al., 2016; Rhodes et al., 2015). Furthermore, Ganel and Goshen-Gottstein (2004) found that familiarity of faces increases the perceptual inter-dependence of identity and expression recognition. They suggested that differences between the facial configurations of individuals should lead to systematic differences in the way emotions are expressed by people. For this reason, every individual can express each facial expression in a unique way. Knowledge of the identity of the observed subject can therefore facilitate the process of his or her facial expression.

From a computational perspective, Calder et al. (2001) supported the plausibility of common codings for both identity and expression. They demonstrated that a multidimensional space derived from a Principal Component Analysis (PCA) (Turk and Pentland, 1991) can provide a set of components being either identity-independent, expression-independent or identity-expression-interdependent (Calder and Young, 2005). Their results demonstrated that this common representation can support both identity and expression recognition and that the representations of identity and expression partially overlap. However, in order to perform identity and expression recognition, the authors used two distinct latent discriminant analysis (LDA) modules, one choosing the best components to support identity recognition, whereas the other choosing the best components to support expression classification. In this chapter, I will show that it is possible to realise similar inter-dependencies within a single face-space representation.

5.2 The Face-Space Duality Hypothesis

In the previous sections I suggested:

(i) A multidimensional spatial representation of faces, the face-space, to be a plausible model for explaining many face perception phenomena;

(ii) Traditional research in face perception proposing neural areas able to promote separation between invariant and dynamic facial features, which facilitates identity and expression recognition through distinct processes happening in parallel;

(iii) A modern understanding of face perception, proposing interdependencies between codings of invariant and dynamic facial features;

Points (ii) and (iii) appear to suggest different perspective, which may be difficult to unify in a single hypothesis. However, in this chapter I introduce the *'face-space duality hypothesis'*, aiming to reduce this conflict by suggesting that:

Hypothesis 5.1 *Face stimuli can be encoded according to their perceived properties in a multidimensional dual face-space (i). This multidimensional representation has a twofold structure enhancing the separation of dynamic and invariant features of the face (ii), although underlying common codings for both dynamic and invariant facial features (iii). This dual face-space is able to facilitate facial expression and identity classification of novel face stimuli.*

5.2.1 Modelling the Hypothesis

Consider a $\mathscr{D}$-dimensional face stimulus x_i shaped as a column vector of a matrix X containing a set of N observed face stimuli. Consider also a corresponding d-dimensional point y_i of a multidimensional psychological face-space shaped as a column vector of a matrix Y containing the d-dimensional representations of the face stimuli in X. The spatial representations in Y encode most of the original information of the input face stimuli in X and can be realised through a mapping function $\mathcal{S}(X) \mapsto Y$.

I introduce the functions $\mathcal{C}^{\mathscr{E}}_{\mathscr{D}} : \mathbb{R}^{\mathscr{D}\times N} \to \mathbb{N}$ and $\mathcal{C}^{\mathscr{I}}_{\mathscr{D}} : \mathbb{R}^{\mathscr{D}\times N} \to \mathbb{N}$, respectively providing the number of correctly classified facial expressions and identities from a set of N novel face stimuli x_i having dimension $\mathscr{D}$ and shaped as column vectors of the matrix X. I also introduced the functions $\mathcal{C}^{\mathscr{E}}_{d} : \mathbb{R}^{d\times N} \to \mathbb{N}$ and $\mathcal{C}^{\mathscr{I}}_{d} : \mathbb{R}^{d\times N} \to \mathbb{N}$, respectively providing the number of correctly classified facial expressions and identities from a set of N novel face stimuli y_i represented as points of a d-dimensional face-space and shaped as column vectors of the matrix Y. Finally, I introduce a permutation function $\sigma : \mathbb{R}^{a\times b} \to \mathbb{R}^{a\times b}$ sorting the column vectors $\left[m^1, \ldots, m^b\right]$ of a matrix M in the inverse order:

$$\sigma(M) = \tilde{M} = \begin{pmatrix} m^1 & m^2 & m^3 & \ldots & m^b \\ m^b & m^{b-1} & m^{b-2} & \ldots & m^1 \end{pmatrix} \tag{5.1}$$

I make use, here and for the rest of the chapter, of the superscript $\tilde{}$ to denote a matrix to which is applied the permutation in Equation 5.1. Then, given the set

of perceived face stimuli X and the associated face-space codings Y, the mapping function $\mathcal{S}$ is defined such that:

1. $\mathcal{S}(X) \mapsto Y$;
2. $\mathcal{C}_d^{\mathscr{E}}(Y) \gg \mathcal{C}_{\mathscr{D}}^{\mathscr{E}}(X)$;
3. $\mathcal{C}_d^{\mathscr{I}}(\tilde{Y}) \gg \mathcal{C}_{\mathscr{D}}^{\mathscr{I}}(X)$;

In other words, the face-space duality hypothesis assumes a function $\mathcal{S}$ able to map the perceived properties of face stimuli onto a psychological multidimensional space having a twofold structure. This new representation results in codings $Y = \Big[y_1, y_2, \ldots, y_n\Big]$ and associated permutations $\tilde{Y} = \Big[\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n\Big]$ able to support significantly higher recognition rates, compared to basic features matching strategies implemented on the original stimuli X.

The rationale behind this idea is that this twofold face-space, in order to maximise the separation between dynamic and invariant features of the face in a single multidimensional representation, will order the components of the resulting space in such a way that the first ones will mostly encode dynamic features of the face, whereas the latter will mostly encode invariant features of the face. Therefore, the resulting face-space would provide a single multidimensional representation (as per point (i) of Hypothesis 5.1) able to facilitate expression and identity classification (as per point (ii) of Hypothesis 5.1), although under a common coding where invariant and dynamic features of the face are inter-dependent (as per point (iii) of Hypothesis 5.1). Figure 5.1 shows an example of the rationale behind the proposed hypothesis.

As I will show in the following sections, the hypothesis is plausible from a computational perspective.

5.3 Dimensionality Reduction Models

The previously introduced mapping function $\mathcal{S}$ can be modelled as a dimensionality reduction mapping function. A dimensionality reduction function maps a high-dimensional signal onto a point of a low-dimensional space. For example, consider an image of a face having resolution 100×100 pixels. This observed signal is represented by a set of pixels and can be posed as a column vector x_i of dimension $\mathscr{D} = 10000$. Dimensionality reduction models provide a mapping function $\mathcal{S} : \mathbf{R}^{\mathscr{D}\times 1} \rightarrow \mathbf{R}^{d\times 1}$, with $d \ll \mathscr{D}$, such that the low-dimensional representation $y_i = S(x_i)$ is able to explain the observed data x_i (Yan et al., 2007).

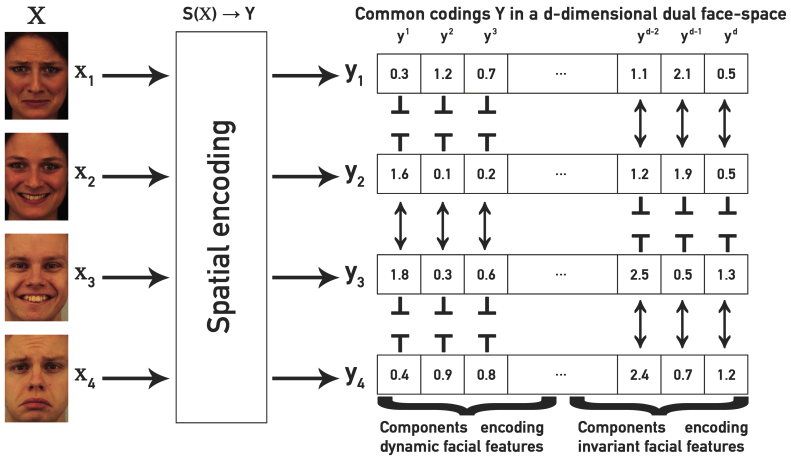


Figure 5.1 The dual face-space presents a twofold structure: on one side the components of the space allow observations with similar facial configurations to lie within close spatial locations (↕), whereas at the same time "repulsing" observations of similar identities away (⊣ ⊢); on the other side, the components of the proposed space allow observations of similar identities to lie in close locations, whereas at the same time "repulsing" observations of similar facial configurations away. This facilitates respectively facial expression and identity recognition, under common multidimensional codings.

Linear dimensionality reduction techniques make use of a linear projection matrix $V \in \mathbf{R}^{\mathscr{D} \times d}$ in order to map the high dimensional observed sample onto the low-dimensional target space. So the projection y_i of an observation x_i can be computed as $y_i = V^\top x_i$. When V is an orthogonal matrix, an approximation of the original observation can be reconstructed from its projection: $x_i \approx V y_i$. The projection matrix V can be estimated by solving an *objective function*. The objective function models desired constraints that the structure of the target low-dimensional space is required to satisfy.

For the purposes of this dissertation, I limit the investigation in providing an implementation of the proposed hypothesis through *linear* dimensionality reduction techniques. As I will show in the remainder of this chapter, this choice is sufficient to validate my hypothesis. However, it is not my intention to suggest that face processing capabilities in humans are shaped as linear dimensionality reduction algorithms. Rather, in this chapter, I propose that, from a computational level of analysis, face-space representations can plausibly exhibit a twofold structure, as suggested by Hypothesis 5.1.

5.3.1 Graph-based Dimensionality Reduction Framework

In this work I adopt the framework of Yan et al. (2007) used to unify many dimensionality reduction models available in the literature, which briefly introduce below. Let $X = [x_i, \ldots, x_n]$ be a matrix of the N observations represented as column vectors with dimension $\mathscr{D}$. The structure of the target low-dimensional space can be constrained by a similarity matrix W and a penalty matrix $W^{(p)}$. For each pair of samples (x_i, x_j), the similarity matrix W_{ij} encodes the associated non-negative similarity measure, whilst the penalty matrix $W^{(p)}_{ij}$ encodes the related penalty measure. This penalty measure can be used as a repulsive force between pairs of samples to prevent samples with high similarity but belonging to different classes from being placed in close proximity in the low-dimensional space (Kokiopoulou and Saad, 2009). The similarity and penalty matrices define two graph structures over the input data, thus the name *'graph-based'*.

Let D and L respectively denote the objective diagonal matrix and the objective Laplacian matrix. These matrices can be defined as:

$$D_{ii} = \sum_j W_{ij}, \quad L = D - W \tag{5.2}$$

and similarly the penalty diagonal matrix $D^{(p)}$ and the penalty Laplacian matrix $L^{(p)}$ as:

$$D^{(p)}_{ii} = \sum_j W^{(p)}_{ij}, \quad L^{(p)} = D^{(p)} - W^{(p)} \tag{5.3}$$

The main property of a Laplacian matrix is that for any matrix X of N column vectors x_i we have that:

$$Tr(XLX^\top) = \frac{1}{2}\sum_i^N \sum_j^N W_{ij}\|x_i - x_j\|_2^2 \tag{5.4}$$

with $Tr(\cdot)$ denoting the matrix trace operator and $\|\cdot\|_2^2$ denoting the Frobenius norm. Then, when an objective function is specified by means of similarity and penalty measures over each pair of samples $(x_i,\ x_j)$ to be reflected in the corresponding encodings $(y_i,\ y_j)$, it is possible to set the previously defined similarity and penalty matrices $W, W^{(p)}$, and determine the optimal mapping matrix $V^\star$ by solving the following objective function:

$$V^\star = \underset{V \in \mathbf{R}^{\mathscr{D}\times d}}{\arg\min} \frac{Tr(V^\top XLX^\top V)}{Tr(V^\top XL^{(p)}X^\top V)} \tag{5.5}$$

Unfortunately, there is no closed-form solution to this optimisation problem (Ngo et al., 2012). However, the problem can be solved numerically with iterative algorithms whenever the matrix $XL^{(p)}X^\top$ is positive definite. The resulting optimal solution $V^\star$ is unique up to unitary transforms of the columns (Ngo et al., 2012).

To ensure that the matrix $XL^{(p)}X^\top$ is positive definite, the process is usually split into two phases:

1. given a dimension $\mathscr{D}' < N$, the observations X are provided in input to a PCA, and a first mapping matrix $V_{pca} \in \mathbf{R}^{\mathcal{D}\times\mathcal{D}'}$ is estimated;

2. the samples $\bar{X} = V_{pca}^\top X$ are provided as input to the objective function in (5.5) and the optimal mapping matrix $V^\star \in \mathbf{R}^{\mathcal{D}'\times d}$ is estimated.

Hence, the overall mapping matrix able to reduce the dimensionality from dimension $\mathscr{D}$ to dimension d and to perform the constraints specified in the objective function (5.5) is given by:

$$V_{overall} = V_{pca}V^\star \tag{5.6}$$

5.4 Model Implementation

Consider a set X of N observations of frontal faces. Each observation consists of $\mathscr{D}$ pixel values, shaped as a $\mathscr{D} \times 1$ vector. In this chapter, I limit the implementation and evaluation of the model for observations varying in identity and facial expression only, thus not considering additional dynamic features of the face (*e.g.* illumination and viewpoint). However, the model can be easily extended to accommodate other dynamic features of the face stimuli, as I will briefly explain in the following Section 5.4.1.

I set a dimension $d < N \ll \mathscr{D}$ and I estimate a mapping matrix $V_{pca} \in \mathbf{R}^{\mathscr{D}\times d}$ by submitting the samples X to a PCA, thus obtaining the corresponding PCA-encodings $\bar{X} = V_{pca}^\top X$. I aim to estimate another mapping matrix $V^\star \in \mathbf{R}^{d\times d}$, such that the final overall matrix $V_{overall} = V_{pca}V^\star$ validates the suggested duality hypothesis.

Denote the identity class of the sample $\bar{x}_i$ with $\mathscr{I}(\bar{x}_i)$ and the facial expression class of the sample $\bar{x}_i$ with $\mathscr{E}(\bar{x}_i)$.

Before proceeding with designing the appropriate similarity and penalty matrices able to implement the proposed dual face-space via Equation 5.5, it is

Figure 5.2 Some examples of prototypes. On the left are two prototypical identities (F05 and M07) in which expression-related features are reduced, whereas on the right are two examples of prototypical facial expressions (happiness and surprise).

important to note that invariant features of the face extend to more regions of the face than dynamic ones, thus explaining most of the variance in the considered dataset of observations (Turk and Pentland, 1991). Hence, most of the variance in face stimuli relates to identity (Turk and Pentland, 1991), and it is likely to expect that, at least on average, faces with same identity and different facial expression would vary less (*i.e.* they are more similar) than faces with the same facial expression but different identity. This means that during facial expression classification the identity can potentially introduce a bias on the samples (the *identity-bias*), thus increasing their similarity and reducing their distance in the face-space even when they belong to a different class of facial expression (Sariyanidi et al., 2015).

Therefore, as a first step of the proposed implementation, I design the similarity matrix to encourage pairs of samples associated with the same facial expression to be in close proximity in the resulting space, and the penalty matrix to provide a repulsive force between pairs of samples belonging to the same identity. This would result in maximising the separation between dynamic and invariant components of the face, thus facilitating the classification of facial expression.

Accordingly, I define the similarity matrix $W^{\mathscr{E}}$ as:

$$W_{ij}^{\mathscr{E}} = \begin{cases} \frac{1}{n_{\mathscr{E}_i}}, & \text{if } \mathscr{E}(\bar{x}_i) = \mathscr{E}(\bar{x}_j) \\ 0, & \text{otherwise.} \end{cases} \tag{5.7}$$

where $n_{\mathscr{E}_i}$ is the number of samples in $\bar{X}$ belonging to facial expression class $\mathscr{E}(\bar{x}_i)$.

Similarly, I define the penalty matrix $W^{\mathscr{I}}$ as:

$$W_{ij}^{\mathscr{I}} = \begin{cases} \frac{1}{n_{\mathscr{I}_i}}, & \text{if } \mathscr{I}(\bar{x}_i) = \mathscr{I}(\bar{x}_j) \\ 0, & \text{otherwise.} \end{cases} \tag{5.8}$$

where $n_{\mathscr{I}_i}$ is the number of samples in $\bar{X}$ belonging to identity class $\mathscr{I}(\bar{x}_i)$.

Then, the objective function in Equation 5.5 becomes:

$$\underset{V \in \mathbf{R}^{d \times d}}{\arg\min} \frac{Tr(V^\top \bar{X}(I_N - W^{\mathscr{E}})\bar{X}^\top V)}{Tr(V^\top \bar{X}(I_N - W^{\mathscr{I}})\bar{X}^\top V)} \tag{5.9}$$

By using the similarity and penalty matrices in Equations 5.7 and 5.8, the resulting Laplacians becomes $L = I_N - W^{\mathscr{E}}$ and $L^{(p)} = I_N - W^{\mathscr{I}}$, with I_N a $N \times N$ identity matrix. Hence, these Laplacians behave as block centering matrices. These matrices remove respectively the corresponding prototypical facial expression (*i.e.* an average identity showing the averaged facial expression) and the corresponding prototypical identity (*i.e.* the considered identity showing a neutral facial expression) from samples $\bar{X}$ (see Figure 5.2 for examples of prototypical expressions and identities obtained by averaging samples from the dataset presented in Section 5.5).

In fact, since a centring matrix is symmetric idempotent and $Tr(AA^\top) = \|A\|_2^2$, the objective function in Equation 5.9 is equivalent to:

$$\underset{V \in \mathbf{R}^{d \times d}}{\arg\min} \frac{\|V^\top \bar{X} - V^\top \bar{X} W^{\mathscr{E}}\|_2^2}{\|V^\top \bar{X} - V^\top \bar{X} W^{\mathscr{I}}\|_2^2} \tag{5.10}$$

Thus, the objective function in Equation 5.10 (and so equivalently the one in Equation 5.9) attempts in minimising the distances between the encodings $Y = V^\top \bar{X}$ and their corresponding prototypical facial expressions $Y^{\mathscr{E}}_{proto} = V^\top \bar{X}^{\mathscr{E}}_{proto} = V^\top \bar{X} W^{\mathscr{E}}$, while maximising their distances with respect to the prototypical identity $Y^{\mathscr{I}}_{proto} = V^\top \bar{X}^{\mathscr{I}}_{proto} = V^\top \bar{X} W^{\mathscr{I}}$, overall facilitating expression recognition.

This means that the suggested weight matrices promote a *norm-based space*, namely a space described in terms of distances between points coding the input observations, and respectively centroids coding their prototypical facial expression and identity.

The objective function in Equation 5.9 can be solved by the iterative algorithm proposed by Ngo et al. (2012). Given the matrices $M^{\mathscr{E}} = \bar{X}(I_N - W^{\mathscr{E}})\bar{X}^\top$ and $M^{\mathscr{I}} = \bar{X}(I_N - W^{\mathscr{I}})\bar{X}^\top$ the optimal mapping matrix $V^\star$ can be found through the Algorithm 5.1. Because this algorithm returns an orthogonal matrix, the resulting mapping can be reversed by simple matrix transposition.

From Algorithm 5.1, it is possible to note that the optimal mapping matrix $V^\star$ results to be the set of eigenvectors related to the smallest eigenvalues of $\mathcal{G}(\rho^\star)$, with $\rho^\star$ being the result of the trace ratio in Equation 5.9 when posing the optimal solution $V^\star$. If, instead of taking the eigenvectors associated with the smallest

Data: Matrices $M^{\mathscr{E}}$, $M^{\mathscr{I}}$, a maximum number of iterations K and a tolerance ϵ.
Result: A mapping matrix V of dimension $\mathscr{D} \times d$.
$V \leftarrow I_{\mathscr{D} \times d}$;
for $i \leftarrow 1$ **to** K **do**
  $\rho \leftarrow \frac{Tr(V^\top M^{\mathscr{E}} V)}{Tr(V^\top M^{\mathscr{I}} V)}$;
  $\mathcal{G}(\rho) \leftarrow M^{\mathscr{E}} - \rho M^{\mathscr{I}}$;
  Compute the smallest (for minimisation) or largest (for maximisation) d eigenvalues $[\lambda_1, \ldots, \lambda_d] \equiv \Lambda$ of $\mathcal{G}(\rho)$ and associated eigenvectors $[v_1, \ldots, v_d] \equiv V$;
  if $|\sum_{j=1}^{d} \Lambda| < \epsilon$ **then**
    break;
  end
end

Algorithm 5.1: Newton-Lanczos algorithm for optimisation of objective function in Equation 5.9.

eigenvalues, we take the eigenvectors associated with the largest eigenvalues, it is possible to get the optimal solution for the following objective function:

$$\arg\max_{V \in \mathbf{R}^{d \times d}} \frac{Tr(V^\top \bar{X}(I_N - W^{\mathscr{E}})\bar{X}^\top V)}{Tr(V^\top \bar{X}(I_N - W^{\mathscr{I}})\bar{X}^\top V)} \tag{5.11}$$

By using simple properties of trace and eigenvalues, it follows that the objective function in Equation 5.11 can be equivalently posed as:

$$\arg\min_{V \in \mathbf{R}^{d \times d}} \frac{Tr(V^\top \bar{X}(I_N - W^{\mathscr{I}})\bar{X}^\top V)}{Tr(V^\top \bar{X}(I_N - W^{\mathscr{E}})\bar{X}^\top V)} \tag{5.12}$$

Similarly to Equation 5.9, the objective function in Equation 5.12 is equivalent to:

$$\arg\min_{V \in \mathbf{R}^{d \times d}} \frac{\|V^\top \bar{X} - V^\top \bar{X} W^{\mathscr{I}}\|_2^2}{\|V^\top \bar{X} - V^\top \bar{X} W^{\mathscr{E}}\|_2^2} \tag{5.13}$$

Thus, dual to the previous scenario, the objective function in Equation 5.13 (and equivalently the ones in Equations 5.11 and 5.12) attempts in minimising the distances between the encodings $Y = V^\top \bar{X}$ and their corresponding prototypical identities $Y^{\mathscr{I}}_{proto} = V^\top \bar{X}^{\mathscr{I}}_{proto} = V^\top \bar{X} W^{\mathscr{I}}$, while maximising their distances with respect to the prototypical facial expression $Y^{\mathscr{E}}_{proto} = V^\top \bar{X}^{\mathscr{E}}_{proto} = V^\top \bar{X} W^{\mathscr{E}}$, overall facilitating identity recognition.

Since the eigenvectors of $V^\star$ are estimated from the same matrix $\mathcal{G}(\rho^\star) = M^{\mathscr{E}} - \rho^\star M^{\mathscr{I}}$ in both the objective functions in Equation 5.9 and Equation 5.11, the components of the two spaces are the same, but differing by order.

In other words, given $V^\star$ as the optimal matrix resulting from objective function in Equation 5.9 it is easy to get the optimal matrix $\tilde{V}^\star$ of the objective function in Equation 5.11, defined as a matrix with the same components of $V^\star$, but arranged in an inverse order (as per permutation in Equation 5.1).

Finally, given the matrix $V_{pca} \in \mathbf{R}^{\mathscr{D} \times d}$ and the matrix $V^\star \in \mathbf{R}^{d \times d}$ it is possible to estimate the final mapping matrix $V_{overall}$ of the face-space through Equation 5.6, which leads respectively to the mapping $Y = V_{overall}^\top X$ and the associated permutated mappings $\tilde{Y} = \tilde{V}_{overall}^\top X$ as suggested by the present hypothesis. This was achieved by modelling a *single* mapping function, in this chapter implemented via the objective function in Equation 5.9, which integrates identity and facial expression features together in a norm-based dual face-space.

It is important to note that this may not be the only way to implement the desired properties of this space promoted by the mapping function $\mathcal{S}$ (for example it can be generalised to non-linear models). Furthermore, the neural implementation of this hypothesis may not resemble the present computational implementation. However, the aim of this chapter is to validate, from a computational level, a more general hypothesis suggesting a twofold structure of the face-space promoting both facial expression and identity recognition and not its neural realisation.

5.4.1 Generalising to Other Dynamic Facial Features

The proposed implementation can be easily generalised to accommodate other dynamic features of the face, such as variation in illumination and in viewpoint.

Denote $P^{(d)} = \{\mathcal{P}_1^{(d)}, \ldots, \mathcal{P}_N^{(d)}\}$ a set of N dynamic properties displayed by the considered face stimuli $\bar{X}$. Then, each property $\mathcal{P}_i^{(d)}$ can be described as a function $\mathcal{P}_i^{(d)}(\bar{x}) \mapsto C_i$, attributing to a face stimulus $\bar{x}$ a class in $C_i = \{c_i^1, \ldots, c_i^q\}$, defined as the set of the possible classes realisable by the property $\mathcal{P}_i^{(d)}$.

Then, the function $\mathscr{E}(\bar{x})$ introduced in the previous section (and so the similarity matrix in Equation 5.7) can be generalised to attribute to a face stimulus $\bar{x}$ a set of classes $\{c_1^a, c_2^b, \ldots, c_N^c\}$ for each dynamic property in $P^{(d)}$, which will be denoted with $C_{(d)}^{a,b,\ldots,c}$. Thus, if the sets of classes of two stimuli $\bar{x}_i$ and $\bar{x}_j$ would differ of at least one element, the corresponding edge $W_{ij}^{\mathscr{E}}$ would be set to 0.

For example, suppose to consider the dynamic properties *facial expression* $\mathcal{P}_{exp}^{(d)}$ and *viewpoint* $\mathcal{P}_{view}^{(d)}$. The facial expression is classified into the six basic

emotions, thus possibly realising the classes c_{exp}^{anger}, $c_{exp}^{disgust}$, c_{exp}^{fear}, $c_{exp}^{happiness}$, $c_{exp}^{neutral}$, $c_{exp}^{sadness}$, and $c_{exp}^{surprise}$. The viewpoint is classified as frontal, side left and side right, thus possibly realising the classes $c_{view}^{frontal}$, c_{view}^{left}, and c_{view}^{right}. Thus, a face stimulus of a frontal happy face will be allocated to the class $C_{(d)}^{frontal,happy}$, whereas a face stimulus of a side left happy face will be attributed to the class $C_{(d)}^{left,happy}$, which will result in two differing sets of classes and consequently leading to a missing edge in the similarity matrix $W^{\mathscr{E}}$.

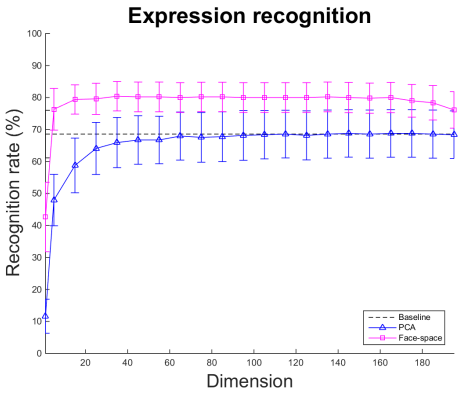
5.5 Hypothesis Validation

I further validate the present hypothesis by using images from the Karolinska Directed Emotional Faces (KDEF) dataset (Lundqvist et al., 1998). The dataset contains static images of 70 subjects—35 female and 35 male—exhibiting 7 different prototypical facial expressions of basic emotions (anger, disgust, fear, happiness, neutral, sadness and surprise). The pictures are taken in different face orientations and in two different sessions (A and B). I used this dataset for my study since commonly used in face perception and processing studies. In addition, the face stimuli from this dataset are already aligned and they do not present extreme illumination variation. These features are important considering I assumed face detection and normalisation mechanisms to be given and not investigated in my dissertation.

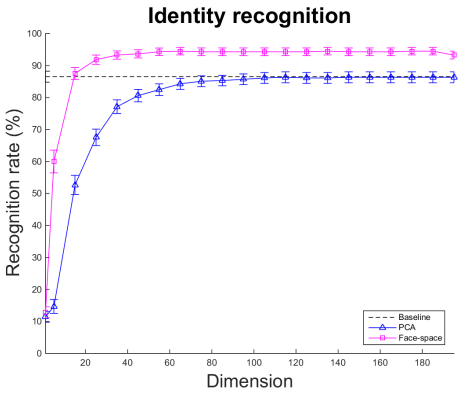
I used the frontal pictures taken in session A. The facial region was extracted from the images and its resolution reduced to 80×80 pixels. Eyes and mouth were at approximately the same position. Illumination variations were reduced by applying a simple equalisation process to the images (the `histeq` function in Matlab software[2]). The data was first pre-processed by submitting the pixels of the images in input to a PCA, as explained previously. In the first experiment were retained the components able to explain 95% of the variance of the original data resulting in 200 components, while in the second experiment were retained the data explaining the 85% (so reducing the number of components and facilitate the visual inspection of the resulted plot) resulting in 100 components.

I performed two experiments: the first to test the ability of the proposed face-space in promoting both identity and expression recognition, and the second to demonstrate the twofold nature of the resulting face-space.

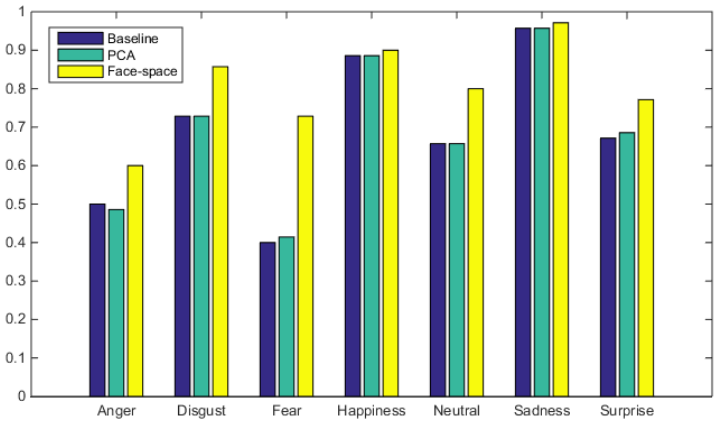
[2]https://au.mathworks.com/products/matlab.html



(a) Expression recognition task



(b) Identity recognition task



(c) Best recognition rates for each expression

Figure 5.3 Performance of the proposed dual face-space model. For each dimension and considered model, in each of the 10 cross-fold iterations each face sample was assigned to one of the 7 expressions (a) or one of the 70 identities (b) depending on the recognition task. In (a) and (b) each point of the plots shows the recognition accuracy for a specific dimension averaged among the 7 expressions (a) or 70 identities (b). The error bars depicts the standard error of the mean (SEM). In (c) are collected the recognition rates of each facial expression for each model during the expression recognition rate. In the case of the PCA and face-space model the dimensions having the best recognition performance led to the selection of the recognition rates depicted in the bar plot.

5.5.1 Identity and Facial Expression Recognition

The first experiment tests the ability of the present implementation to support subsequent processes of identity and facial expression recognition. I used a 10-fold cross validation approach. For each iteration, I divided the data by taking one fold as the test set and the rest for model training.

	Accuracy mean (%)	Accuracy SME (%)	Best dimension
Baseline	68.57	7.44	N/A
PCA	68.78	7.4	145
Dual face-space	80.41	4.59	35

Table 5.1 Expression recognition rates for the best dimensions among the compared models.

With each training data, I estimated the mapping matrix $V_{overall}$ as per Equation 5.9 and Equation 5.6. Then I mapped each test data onto the corresponding face-space, thus obtaining the encodings $Y^{\mathscr{E}} = V_{overall}^{\top} X$ and $Y^{\mathscr{I}} = \tilde{Y}^{\mathscr{E}} = \tilde{V}_{overall}^{\top} X$ respectively used during the expression and identity recognition tasks. Recall that $\tilde{V}$ is a matrix V with its columns selected in an inverse order, and that $\tilde{Y}^{\mathscr{E}}$ is a set of codings with components permutated as per permutation in Equation 5.1.

The classification was performed using the nearest neighbour algorithm on three multi-dimensional spaces derived from the three compared approaches: a baseline approach, a PCA model and the suggested face-space model. Specifically, the baseline approach considers each face sample x_i as a point in the $\mathscr{D}$-dimensional observation space, with $\mathscr{D}$ being the number of pixels of the image sample and each pixel value being one coordinate of such space; the PCA model compresses each face sample x_i in a d-dimensional vector $\bar{x}_i$ preserving as much information as possible (*i.e.* 95% of the original information leading to a space having 200 components); the dual face-space transforms the d-dimensional PCA compressed vector $\bar{x}_i$ in a new d-dimensional vector y_i reflecting the desired constraints encoded in the two weight matrices $W^{\mathscr{E}}$ and $W^{\mathscr{I}}$. For each sample x_i, $\bar{x}_i$ and y_i I computed the Euclidean distances with respect to the centroids of each class (*i.e.* the prototypical identities in the case of identity recognition or the prototypical expressions in the case of facial expression recognition) in the corresponding space (*i.e.* respectively the observation space, the PCA space, and the dual face-space) and selected the label associated with the centroid having the lower distance to the sample. For the encodings $\bar{x}_i$ and y_i I repeated this process for each dimension $k = [1, \ldots, 200]$ by taking only the first k components of the encodings. I computed the recognition rate for each considered expression or identity, depending on the recognition task, among each cross-fold iteration and averaged their classification accuracy.

The results of facial expression and identity recognition are shown, respectively, in Figure 5.3a and Figure 5.3b. This face-space realised through a single process integrating invariant and dynamic facial features facilitates both identity and expression recognition capabilities as compared to basic feature matching strategies,

	Accuracy mean (%)	Accuracy SME (%)	Best dimension
Baseline	86.53	1.72	N/A
PCA	86.33	1.71	115
Dual face-space	94.49	1.17	65

Table 5.2 Identity recognition rates for the best dimensions among the compared models.

such as a simple PCA of pixels similarities as in the baseline approach. Tables 5.1 and 5.2 summarises the recognition rates for the best dimensions of the compared models. We can also notice that the PCA approach does not overcome the baseline approach. This means that the PCA is only capable of reducing the dimensionality of the face stimuli, but it is not capable of providing an optimised representation facilitating the classification of new face stimuli. Another fact we can observe is the drop in performance of the dual face-space when considering almost all the spatial components. This effect is likely due to additional noisy information introduced by the last components (*i.e.* approximately from component #160). In fact, the structure of the dual face-space leads to a separation between the components related to invariant aspects of the face (*i.e.* identity recognition oriented) and the components related to dynamic aspects of the face (*i.e.* expression recognition oriented). The components more relevant to the specific classification task are posed as the very first components of the resulted face-space. Thus, using early components of the obtained space would lead to a rapid gain in classification accuracy, followed by a steady accuracy rate due to the lack of new significant information and ending with a drop in the accuracy rate due to noise introduced by components not relevant to the current classification task. Figure 5.3c compares the best recognition rates for each facial expression among the considered models. The proposed face-space model overcomes the baseline approach and PCA for each considered facial expression. These results demonstrate the versatility of the dual face-space in addressing the classification of different facial expressions.

5.5.2 Face-Space Twofold Structure

I was able to confirm the present hypothesis on the twofold structure of the face-space using the dataset presented previously in Section 5.5. I estimated the mapping matrix $V_{overall}$ as per Equation 5.9 and Equation 5.6 using the full dataset as training data. In order to map an observation x_i onto the face-space, it is possible to select $k \leq d$ components of the mapping matrix $V_{overall}$, thus realising a k-dimensional representation y_i. This new representation can converge to the

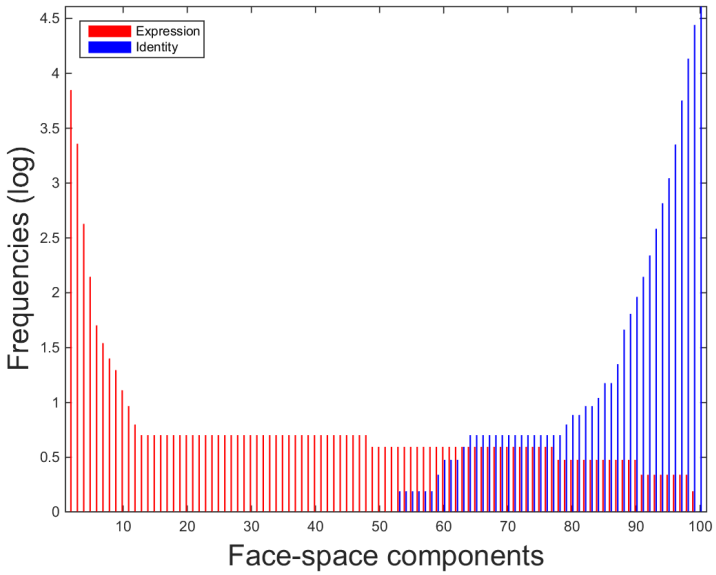


Figure 5.4 Components used in recognition tasks

right class or misclassify it. In this experiment, I want to count how many times a set of k components of the mapping matrix $V_{overall}$ are sufficient to correctly classify a facial expression or a facial identity of each considered sample.

Given the matrix $V_{overall} = [v_1, \ldots, v_d]$ the minimum set of expression components for a sample x_i is the smallest set $V_{i_{min}} = [v_1, \ldots, v_k]$ such that $y_i = V_{i_{min}}^\top x_i$ is classified with the correct expression label $\mathscr{E}(x_i)$ through nearest neighbour algorithm, as in the previous experiment.

Similarly, given the matrix $\tilde{V}_{overall} = [v_d, \ldots, v_1]$ the minimum set of identity components for a sample x_i is the smallest set $\tilde{V}_{i_{min}} = [v_k, \ldots, v_1]$ such that $y_i = \tilde{V}_{i_{min}}^\top x_i$ is classified with the correct identity label $\mathscr{I}(x_i)$ through nearest neighbour algorithm, as in the previous experiment.

For each sample x_i I computed its minimum set of expression and identity components. Then, I set $n_k^{\mathscr{E}}$ the number of times the component k was included in the minimum sets of expression components and $n_k^{\mathscr{I}}$ the number of times the component k was included in the minimum sets of identity components. I computed the final results for expression and identity and for each component k as $f_k^{\mathscr{E}} = \log(\frac{n_k^{\mathscr{E}}}{N} \times 100 + 1)$ and $f_k^{\mathscr{I}} = \log(\frac{n_k^{\mathscr{I}}}{N} \times 100 + 1)$, with N the number of samples in the dataset (*i.e.* 490). The logarithm is used for better readability of the results. The resulting log-frequencies are illustrated in Figure 5.4.

From the results, it is possible to clearly see two peaks placed in the extremes of the face-space components. There are expression components clearly independent

from identity components (components #1 to #52), components shared among expression and identity classification tasks (components #53 to #99) and just one identity component independent from expression ones (component #100).

Crucially, these results are in agreement with the study of Ganel and Goshen-Gottstein (2004), which suggest that expression is perceptually separable from identity, but identity is not perceptually separable from expression. This experiment further supports the proposed twofold structure of face-space as suggested by the face-space duality hypothesis.

5.6 Conclusions

The final aim of this dissertation is to demonstrate that sensory-motor information, possibly provided in forms of bodily formatted representations, does not limit in facilitating facial affect recognition, but it can be used (and I suggest it is sufficient) to promote facial identity discrimination mechanisms. This means that my thesis argument requires a tool able to explain facial identity discrimination mechanisms providing a link to the embodied mechanisms modelled in Chapter 4.

Hence, in this chapter, I extended a conceptual tool, the face-space, widely used in the literature to explain face perception phenomena. This framework is specifically designed to represent invariant features of the face in a multidimensional psychological space. Thus, it is of particular importance to explain facial identity discrimination capabilities.

Unfortunately, the traditional formulation of the face-space framework neglects the integration of dynamic facial features. This is a significant limitation preventing:

1. the explanation of facial affect recognition mechanisms within a single model integrating both invariant and dynamic features of the face, as suggested by modern findings in face processing;

2. a link between the sensory-motor information provided by embodied mechanism and a psychological space for facial identity discrimination, necessary to validate the argument of the present dissertation.

Thus, here I provided a hypothesis, the duality hypothesis, demonstrating that the face-space framework can be extended to accommodate both invariant and dynamic features of the face under common codings of the same multidimensional face-space. In particular, the proposed representation results in a twofold

structured multidimensional space, which I call the dual face-space. Reading the components of this space from one direction facilitates facial affect recognition capability, whereas reading the components of this space from the opposite direction promotes facial identity discrimination capability. I validated the face-space duality hypothesis, from a computational perspective, through a formal mathematical presentation, by considering a general dimensionality reduction framework. The hypothesis was further supported by experimental data. Specifically, the provided face-space model is capable of enhance facial expression and identity recognition accuracy as compared to other basic feature matching approaches.

However, the model was tested on a limited set of face samples with low or no variations in illumination, viewpoint, obstructing features and other dynamic features not related to the facial expression. In addition, the dataset included only one sample for each facial expression of each individual. Adding more samples of the same facial expression for the same individual under different conditions or presenting temporal relationships may further facilitate individuation (Xiao et al., 2014; Yankouskaya et al., 2014). Finally, this model cannot provide answers to differences in face processing between gender, since the present model does not take into account such variable. These acknowledged limitations can be addressed by future studies by extending the proposed model and testing it on more extended datasets.

In Chapter 4, I provided a plausible computational account modelling embodied mechanisms able to promote the classification of facial sensory-motor dynamics. These mechanisms are suggested to be at least partially innate (Meltzoff and Decety, 2003) and refined through sensory-motor learning (Catmur et al., 2007). Unfortunately, information about the displayed facial configuration, available via the suggested embodied mechanisms, is not sufficient to implement the proposed twofold face-space. In fact, in order to computationally implement the dual face-space it is necessary to provide a training set of labelled data with information about the facial expression and the identity exhibited by the face stimuli. For this reason, in the following chapter, I will show that the proposed dual face-space can be further generalised to necessitate only information about the facial dynamics, without the need of knowing the identities exhibited by the training face stimuli during the training process. This will eventually validate the thesis argument proposed in this dissertation.

Chapter Bibliography

Bruce, V. and Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77:305–327.

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., and Rectem, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. *Brain and Cognition*, 2(3):257–284.

Calder, A. J. (2011). *Oxford handbook of face perception*. Oxford University Press.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., and Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, 41(9):1179–1208.

Calder, A. J. and Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6(8):641–651.

Catmur, C., Walsh, V., and Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Current Biology*, 17(17):1527–1531.

Fitousi, D. and Wenger, M. J. (2013). Variants of independence in the perception of facial identity and expression. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1):133.

Ganel, T. and Goshen-Gottstein, Y. (2004). Effects of familiarity on the perceptual integrality of the identity and expression of faces: The parallel-route hypothesis revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3):583.

Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., and Goodale, M. A. (2005). The involvement of the "fusiform face area" in processing facial expression. *Neuropsychologia*, 43(11):1645–1654.

Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological Bulletin*, 141(6):1266.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6):223–233.

Kadosh, K. C., Luo, Q., de Burca, C., Sokunbi, M. O., Feng, J., Linden, D. E., and Lau, J. Y. (2016). Using real-time fMRI to influence effective connectivity in the developing emotion regulation network. *NeuroImage*, 125:616–626.

Kokiopoulou, E. and Saad, Y. (2009). Enhanced graph-based dimensionality reduction with repulsion laplaceans. *Pattern Recognition*, 42(11):2392–2402.

Lee, K., Byatt, G., and Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science*, 11(5):379–385.

Lundqvist, D., Flykt, A., and Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, pages 91–630.

Meltzoff, A. N. and Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431):491–500.

Ngo, T. T., Bellalij, M., and Saad, Y. (2012). The trace ratio optimization problem. *SIAM Review*, 54(3):545–569.

Pell, P. J. and Richards, A. (2013). Overlapping facial expression representations are identity-dependent. *Vision Research*, 79:1–7.

Rhodes, G., Jaquet, E., Jeffery, L., Evangelista, E., Keane, J., and Calder, A. J. (2011). Sex-specific norms code face identity. *Journal of Vision*, 11(1):1.

Rhodes, G., Pond, S., Burton, N., Kloth, N., Jeffery, L., Bell, J., Ewing, L., Calder, A. J., and Palermo, R. (2015). How distinct is the coding of face identity and expression? Evidence for some common dimensions in face space. *Cognition*, 142:123–137.

Sariyanidi, E., Gunes, H., and Cavallaro, A. (2015). Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6):1113–1133.

Tranel, D., Damasio, A. R., and Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology*, 38(5):690–690.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43(2):161–204.

Valentine, T., Lewis, M. B., and Hills, P. J. (2015). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, (ahead-of-print):1–24.

Vitale, J., Williams, M.-A., and Jonhston, B. (2016). The face-space duality hypothesis: A computational model. In *38th Annual Meeting of the Cognitive Science Society*, pages 514–519.

Vitale, J., Williams, M.-A., and Jonhston, B. (2017). Facial motor information is sufficient for identity recognition. In *39th Annual Meeting of the Cognitive Science Society*, pages 3447–3452.

Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y.-H. P., and Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology*, 5.

Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., and Lin, S. (2007). Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51.

Yankouskaya, A., Humphreys, G. W., and Rotshtein, P. (2014). The processing of facial identity and expression is interactive, but dependent on task and experience. *Frontiers in human neuroscience*, 8.

You can get along very well in this world by simply coming up with a quantity of reasonably valid statements.

— B.F. Skinner —

6

Thesis Validation[1]

In Chapter 2, I discussed why face processing capabilities are of paramount importance for the development of social skills. These capabilities are so critical that infants show innate mechanisms preferentially orienting their attention to face-like stimuli (Grossmann, 2015), thus enabling social exchanges with their caregivers (Trevarthen, 2006). However, detecting a face is just one of the steps necessary to promote social cognition skills. Indeed, humans need to *attribute an identity* to the observed face stimulus (Palermo and Rhodes, 2007; Simion and Di Giorgio, 2015) and to attribute mental states by *interpreting the exhibited facial motor configuration* (Niedenthal et al., 2014).

In Sections 2.1.6 and 2.2, I reported studies showing newborns' ability to match observed facial motor configurations via imitative behaviour, even well before the development of early cognitive capabilities (Meltzoff and Moore, 1983, 1992) and an experiment demonstrating that dynamic facial information facilitates face recognition from early stages of life (Leo et al., 2018). In Chapter 2, I reported literature proposing that facial expression recognition can be mediated by rudimentary imitative mechanisms available since the very beginning of life and

[1]Part of this chapter is an adaptation of *"Vitale, J., Williams, M.-A., and Jonhston, B. (2017). Facial motor information is sufficient for identity recognition. In 39th Annual Meeting of the Cognitive Science Society , pages 3447–3452"*.

refined through reciprocal social exchanges between the caregiver and the baby (Iacoboni, 2009). Infants' imitative behaviour can be plausibly realised by neural mechanisms mapping sensory information of the observed facial configuration into a proprioceptive motor format (Gallese and Caruana, 2016; Iacoboni, 2009), as I demonstrated in Chapter 4.

In Section 2.2.2, I discussed that face identity processing capabilities are not entirely innate, but they follow a developmental process (Grossmann and Vaish, 2009). Currently, facial identity processing development is not yet well understood. For example, we do not know yet how and where in the face processing hierarchy representations of invariant (*e.g.* identity) and dynamic (*e.g.* motor configuration) features interact (Simion and Di Giorgio, 2015; Yankouskaya et al., 2014). To propose how facial identity and expression representations interact, I provided the dual face-space model in Chapter 5. I demonstrated that this face-space having a twofold structure could accommodate both dynamic and invariant facial features into a shared representation. This representation can facilitate both facial expression and identity recognition capabilities.

In this chapter, I offer a new understanding of the face-space proposed in Chapter 5 able to validate my thesis argument:

> *Sensory-motor information of face stimuli is sufficient to facilitate the acquisition of face recognition capabilities because this information is available early in life via embodiment mechanisms able to map sensory information of novel face stimuli, encountered throughout social exchanges, onto corresponding motor representations.*

I suggest that to acquire face recognition skills it is *sufficient* to interpret the motor configuration of observed face stimuli. This interpretation is likely to be mediated by embodied simulation mechanisms, as described in Chapter 4. Hence, as secondary contribution of this dissertation, I propose that face-space representations acquired by using *altered* sensory-motor information would lead to dysfunctional face processing capabilities. These results will support Hypotheses 3.1 and 3.2 presented in Chapter 3, thus adding evidence advancing embodied cognition theories.

In summary, in this chapter, I will show that:

1. It is possible to provide a computational explanation of a face-space representation able to facilitate both facial expression and identity recognition capabilities by only interpreting the motor information displayed by the observed face stimuli;

2. The interpretation of the facial motor information is plausibly shaped by embodied simulation mechanisms. Thus, a dysfunctional embodiment would realise altered motor information that would consequently impair face processing skills at the core of social cognition development.

The two contributions of this chapter will promote a better understanding of face processing mechanisms and advance embodied cognition research.

6.1 Summary of Previous Findings

In this section, I will provide a set of propositions summarising the key findings illustrated so far. It is possible to validate each proposition by referencing to previous results offered in this work. This process is necessary to connect the evidence provided in the previous chapters in a single section, easily guiding the reader throughout the validation of the proposed thesis and the offered auxiliary hypotheses.

Denote with X_{self} a set of self-centred face stimuli describing the motor potentials of the subject's face (*i.e.* the potential facial configurations and poses the subject can realise).

Proposition 6.1 *The set X_{self} is sufficient to implement forward and inverse mechanisms embodying observed face stimuli.*
Proof: *See the equations in Section 4.2 on page 125, their derivation in Section 4.2.1 on page 127, and the implementation in Section 4.3 on page 130.*

Denote with Φ a set of representations bodily in their format and with X a set of observed face stimuli.

Proposition 6.2 *The inverse embodied mechanisms map observed face stimuli X into bodily formatted representations Φ.*
Proof: *See the Definition 2.23 of embodiment on page 56, the Definition 2.25 of bodily formatted representations on page 57, and the implementation of the inverse process in Section 4.3.2 on page 133 reflecting the provided definitions.*

Proposition 6.3 *The forward embodied mechanisms map bodily formatted representations Φ of facial motor potentials into self-centred facial motor configurations X_{self}.*
Proof: *See the implementation of the forward process in Section 4.3.2 on page 133, and the results in Section 4.4.2 on page 138.*

Denote with X_{dev} a set of novel face stimuli encountered during development, with Φ_{dev} the corresponding bodily formatted representations, and with $\ell_{X_{dev}}$ the set of motor configurations displayed by the face stimuli X_{dev}.

Proposition 6.4 *It is possible to use the bodily formatted representations Φ_{dev}, realised by embodying face stimuli X_{dev} via inverse mechanisms, to realise a set $\ell_{X_{dev}}$ able to provide accurate interpretations of the motor configurations displayed by the observed face stimuli X_{dev}.*
Proof: *See the classification performance in Section 4.4.3 on page 139.*

Denote with $\mathscr{E}_{X_{dev}}$ and with $\mathscr{I}_{X_{dev}}$ respectively the set of facial expression classes and the set of identity classes exhibited by the set of face stimuli X_{dev}.

Proposition 6.5 *The face stimuli X_{dev} and the corresponding facial expression and identity classes $\mathscr{E}_{X_{dev}}$ and $\mathscr{I}_{X_{dev}}$ are sufficient to implement a dual face-space representation able to encode dynamic and invariant facial features under shared codings.*
Proof: *See the mathematical presentation in Section 5.4 on page 159, and the results in Section 5.5.2 on page 167.*

Proposition 6.6 *The realised face-space representation exhibits a twofold structure able to facilitate both facial expression and identity recognition capabilities.*
Proof: *See the mathematical presentation in Section 5.4 on page 159, and the results in Section 5.5.1 on page 165.*

This set of propositions summarises the key findings discussed in the previous chapters of this dissertation.

6.2 Face Processing Development via Motor Information

Given Proposition 6.5, in order to implement the proposed dual face-space, it is necessary to know the set of facial expression classes $\mathscr{E}_{X_{dev}}$, and the set of facial identity classes $\mathscr{I}_{X_{dev}}$ exhibited by the training set X_{dev}. In Chapter 4, I demonstrated that the set $\ell_{X_{dev}}$ is accurate enough to classify facial expressions of novel stimuli, namely its classification of facial expressions is above chance level and having significantly better accuracy as compared to a feature matching baseline approach. Hence, $\ell_{X_{dev}}$ is a plausible approximation of the set $\mathscr{E}_{X_{dev}}$.

Proposition 6.7 *The set $\ell_{X_{dev}}$ interpreting the motor configuration of observed face stimuli X_{dev} can plausibly approximate the facial expression classes set $\mathscr{E}_{X_{dev}}$ corresponding to the face stimuli X_{dev}.*
Proof: *See the results in Section 4.4.3 on page 139 and how all the considered expressions similarly present in $\mathscr{E}_{X_{dev}}$ were classified by the model presented in Chapter 4.*

Since it is likely that the facial motor potentials X_{self} are hard-wired and available from birth (Gallese, 2001; Meltzoff and Moore, 1983, but see also Section 2.2 and Section 2.3.4), the interpretations in the set $\ell_{X_{dev}}$ can be available early in life by employing the embodied mechanisms (Boccignone et al., 2018) offered in this dissertation (Proposition 6.4). This set can sufficiently approximate the desired set of facial expression classes $\mathscr{E}_{X_{dev}}$ (Proposition 6.7). Therefore, Propositions 6.4 and 6.7 connect Chapter 4 to Chapter 5. This link significantly advances the argument of this dissertation by proposing that face processing capabilities can be plausibly shaped by embodied mechanisms.

Unfortunately, the present evidence is not sufficient to validate my thesis argument. In fact, I showed that sensory-motor information, plausibly available from the embodied mechanisms proposed in Chapter 4, can realise the set $\mathscr{E}_{X_{dev}}$ necessary to implement the face-space promoting facial identity recognition proposed in Chapter 5 but not the set $\mathscr{I}_{X_{dev}}$. Hence, in order to validate my thesis there are two options available:

1. The sensory-motor dimension of embodied representations can also approximate the set $\mathscr{I}_{X_{dev}}$. Thus, the set $\mathscr{I}_{X_{dev}}$ is available throughout development and it supports the implementation of the proposed dual face-space facilitating facial identity recognition;

2. The set $\mathscr{I}_{X_{dev}}$ is not necessary to implement the dual face-space.

The first option is highly implausible. In fact, if the sensory-motor dimension of embodied representations can also approximate a set of the exhibited facial identities, there is no reason to implement a face-space, since facial expression and facial identity classification can be approximated by embodied mechanisms available early in life. Thus, the face-space would result in an unnecessary duplication.

The implausibility of the first option is further supported by modern literature in face processing, which provides evidence in favour of a face-space representation in humans able to explain phenomena underlying both facial expression and facial

identity classification (Valentine et al., 2015). In addition, it has recently been suggested that perceptual features of facial expression and identity in face stimuli are processed inter-dependently (Ganel et al., 2005; Kadosh et al., 2016; Rhodes et al., 2015). Finally, it is widely supported that facial identity recognition is not an entirely innate capability, but it develops with experience, and it is biased by the social environment in which the individual is situated (De Heering et al., 2010; Goodman et al., 2007; Meissner and Brigham, 2001; Pascalis et al., 2002).

Therefore, the second option is the only plausible one, and I suggest to replace Proposition 6.5 with the following proposition:

Proposition 6.8 *The face stimuli X_{dev} and the corresponding facial expression classes $\mathscr{E}_{X_{dev}}$ are sufficient to implement a dual face-space representation able to encode dynamic and invariant facial features under shared codings.*

I also aim to demonstrate that Proposition 6.6 is still true under this new standpoint.

Therefore, in the remainder of this section, I will provide a new framework, the Δ face-space, and its implementation. I will demonstrate that this new face-space representation can be implemented without knowing the set of identity classes $\mathscr{I}_{X_{dev}}$. Thus, this tool will validate Proposition 6.8, while at the same time maintaining Proposition 6.6 true, as I will show in Section 6.2.2.

6.2.1 The Δ Face-Space

The objective of this section is to demonstrate that the dual face-space offered in Chapter 5 can be implemented without knowing the set of identity classes $\mathscr{I}_{X_{dev}}$.

In Chapter 5, I showed that it is possible to implement the dual face-space by solving the following objective function promoting the classification of facial expressions:

$$\arg\min_{V \in \mathbf{R}^{d \times d}} \frac{Tr(V^\top \bar{X}(I_N - W^{\mathscr{E}})\bar{X}^\top V)}{Tr(V^\top \bar{X}(I_N - W^{\mathscr{I}})\bar{X}^\top V)} \tag{5.9}$$

and by making use of the following permutation function sorting the column vectors $\left[m^1, \ldots, m^b\right]$ of a matrix M in the inverse order:

$$\sigma(M) = \tilde{M} = \begin{pmatrix} m^1 & m^2 & m^3 & \ldots & m^b \\ m^b & m^{b-1} & m^{b-2} & \ldots & m^1 \end{pmatrix} \tag{5.1}$$

In fact, given $V^\star$ as the optimal solution of the objective function in Equation 5.9, I demonstrated that the mapping matrix $\tilde{V}^\star = \sigma(V^\star)$ is the optimal

solution of another objective function promoting facial identity discrimination:

$$\underset{V \in \mathbf{R}^{d \times d}}{\arg\min} \frac{Tr(V^\top \bar{X}(I_N - W^{\mathscr{I}})\bar{X}^\top V)}{Tr(V^\top \bar{X}(I_N - W^{\mathscr{E}})\bar{X}^\top V)} \tag{5.12}$$

The objective function in Equation 5.12 is dual to the objective function in Equation 5.9. Thus, $V^\star$ and $\tilde{V}^\star$ share the same components but permutated in the opposite order, giving rise to common codings able to facilitate on the one hand facial expression classification ($V^\star$), and on the other hand facial identity discrimination ($\tilde{V}^\star$).

The objective function in Equation 5.9 requires two weight matrices $W^{\mathscr{E}}$ and $W^{\mathscr{I}}$. To operationalise the weight matrix $W^{\mathscr{E}}$ it is necessary the set $\mathscr{E}_{X_{dev}}$, whereas to operationalise the weight matrix $W^{\mathscr{I}}$ it is necessary the set $\mathscr{I}_{X_{dev}}$. Since in this section I want to show that the set $\mathscr{I}_{X_{dev}}$ is not necessary to implement the face-space, the objective of this section narrows down to the following:

Objective. Find a weight matrix W^{Δ} such that W^{Δ} approximates $W^{\mathscr{I}}$ and implementing W^{Δ} does not require the set $\mathscr{I}_{X_{dev}}$.

In this way the weight matrix $W^{\mathscr{I}}$ in Equation 5.9 can be replaced by W^{Δ} and realise the following objective function:

$$\underset{V \in \mathbf{R}^{d \times d}}{\arg\min} \frac{Tr(V^\top \bar{X}(I_N - W^{\mathscr{E}})\bar{X}^\top V)}{Tr(V^\top \bar{X}(I_N - W^{\Delta})\bar{X}^\top V)} \tag{6.1}$$

The optimal solution of the objective function in Equation 6.1 is the mapping matrix $V^{\Delta\star}$. Thus, given a mapping matrix V_{pca} realised by submitting the training data to a Principal Component Analysis (PCA) (Turk and Pentland, 1991), the final mapping matrix for the Δ face-space can be obtained as following:

$$V^{\Delta}_{overall} = V_{pca} V^{\Delta\star} \tag{6.2}$$

The mapping matrix $V^{\Delta}_{overall}$ is able to implement face-space representations facilitating both facial expression and facial identity classification, although without the need of knowing the identities exhibited by the training samples. In other words, the face stimuli given as training samples to the Δ face-space model need to be labelled with an expression class, but there is no need to label them with an identity class.

Defining the New Weight Matrix

The weight matrix $W^{\mathscr{I}}$ in Equation 5.9 is necessary to avoid that two face stimuli sharing the same identity, but exhibiting different facial expressions, would get placed to nearby locations of the face-space. This way it is possible to prevent their misclassification within the same facial expression class. This misclassification can easily happen since face stimuli of the same identity share most of their perceptual features, and, on average, they are close-by in the perceptual space (Sariyanidi et al., 2015; Turk and Pentland, 1991). This property possessed by face stimuli can be used to our advantage to realise the desired weight matrix W^{Δ}.

For each of the N training face stimuli x_i, shaped as column vectors $i \in \{1, \ldots, N\}$ of the matrix X_{dev}, I denote with Δ_{x_i} the set containing the perceptual distances $\delta(x_i, x_j) = \|x_i - x_j\|$ between the face stimulus x_i and the other face stimuli $x_j \in X_{dev}$, with $i \neq j$, exhibiting a different facial expression from the one exhibited by x_i:

$$\Delta_{x_i} = \{\delta(x_i, x_j) \mid x_j \in X \wedge x_i \neq x_j \wedge \mathscr{E}(x_j) \neq \mathscr{E}(x_i)\} \tag{6.3}$$

Since face stimuli of the same identity are perceptually close, their respective distances would be, at least on average, well below their distances from face stimuli with different identities. Then, given the mean $\mu_{\Delta_{x_i}}$ and standard deviation $\sigma_{\Delta_{x_i}}$ of the distances included in the set Δ_{x_i} it is possible to compute the set $\mathscr{I}_i^{\approx}$ described as follow:

$$\mathscr{I}_i^{\approx} = \{x_j \mid \delta(x_i, x_j) < \mu_{\Delta_{x_i}} - \beta\sigma_{\Delta_{x_i}}\} \tag{6.4}$$

where β is a parameter suggesting how many standard deviations below the mean distance would be set the maximum threshold. In this work, β was set equal to 2.5 after empirical tests with face stimuli gathered from different datasets available in face recognition literature. The resulting set $\mathscr{I}_i^{\approx}$ includes most of the training samples sharing the same identity of the sample x_i.

The weight matrix W^{Δ} can be realised as follow:

$$W_{ij}^{\Delta} = \begin{cases} \frac{1}{n_{\mathscr{I}_i^{\approx} \cup \mathscr{I}_j^{\approx}}}, & \text{if } x_j \in \mathscr{I}_i^{\approx} \vee x_i \in \mathscr{I}_j^{\approx} \\ 0, & \text{otherwise.} \end{cases} \tag{6.5}$$

where $n_{\mathscr{I}_i^{\approx} \cup \mathscr{I}_j^{\approx}}$ is the number of unique samples in the set $\mathscr{I}_i^{\approx} \cup \mathscr{I}_j^{\approx}$. The realised weight matrix W^{Δ} is clearly symmetric. Furthermore, since the weights on each

row i of the matrix W^{Δ} sum to 1, the weight matrix W^{Δ}, in a similar way to matrices $W^{\mathscr{I}}$ and $W^{\mathscr{E}}$, implements a Laplacian behaving as a block centring matrix:

$$L^{\Delta} = I_N - W^{\Delta} \tag{6.6}$$

Thus, when the matrix W^{Δ} is used in the objective function Equation 6.1, from each sample in X will be subtracted the prototypical face obtained by averaging the face stimuli with weight $w > 0$ on the corresponding row. Since on average most of these face stimuli share the same identity of the associated training sample, the resulting prototypical face will be similar to the one realised by using matrix $W^{\mathscr{I}}$ (*i.e.* the identity of sample x showing a neutral facial expression). Therefore, the objective function in Equation 6.1 realises a *norm-based space*, similarly to the objective function in Equation 5.9 (for a mathematical demonstration, please refer to details in Section 5.4 on page 159).

In a similar way to what I suggested in Chapter 5, the objective function in Equation 6.1 can be solved by the iterative algorithm proposed by Ngo et al. (2012) and presented in Algorithm 5.1 on page 162. In the following section, I will demonstrate that the mapping matrix $V^{\Delta}_{overall}$, obtained by solving the objective function in Equation 6.1 and by posing it in Equation 6.2, is enough to implement a dual face-space able to facilitate both facial expression and identity recognition with performance comparable to the model proposed in Chapter 5. Importantly, this new model does not require to label the training samples with an identity label and it is not necessary to establish an a priori number of clusters, since the weight matrices will force the topology of the resulting space to reflect the desired distances. This will provide evidence in favour of Proposition 6.8, thus being enough to validate the thesis argued in this dissertation.

6.2.2 Argument Validation

In this section I will evaluate the performance of the proposed Δ face-space in order to demonstrate that:

(i) It facilitates facial expression and identity recognition, although its implementation requires only the facial expression set $\mathscr{E}_{X_{dev}}$. This set is plausibly realised by sensory-motor embodied mechanisms;

(ii) Its facial expression and identity recognition performance are comparable or better to the ones of the dual face-space presented in Chapter 5.

Providing evidence in favour of these statements is enough to validate Proposition 6.8 and the proposed thesis argument.

Dataset

In order to maintain consistency with the previous chapter, I will evaluate the model using the Karolinska Directed Emotional Faces (KDEF) dataset (Lundqvist et al., 1998, please, refer to Section 5.5 on page 164 for more details). For the present study I used frontal face stimuli similar to what I described in Chapter 5. The facial region was extracted from the images and its resolution reduced to 80×80 pixels. Eyes and mouth were at approximately the same position. Illumination variations were reduced by applying a simple equalisation process to the images (the `histeq` function in Matlab software[2]).

The data was first pre-processed by submitting the pixels of the images in input to a PCA, as explained in Section 5.4 on page 159. In all the experiments were retained the components able to explain 95% of the variance of the original data resulting in 200 components.

Procedure

The present experiment tests the ability of the novel Δ face-space, implemented without knowing the identity labels of the training stimuli, to support subsequent processes of identity and facial expression recognition. For this experiment was used a 15-fold cross validation approach. For each iteration, the data was split by taking one fold as the test set and the rest for model training. The number of folds was increased as compared to the experiments presented in Chapter 5 to gather enough samples to run statistical analyses proposed in the remainder of this section.

By using the training data, I estimated the mapping matrix $V_{overall}$ of the dual face-space proposed in Chapter 5 as per Equation 5.9 and Equation 5.6, and the mapping matrix $V^{\Delta}_{overall}$ of the Δ face-space proposed in this chapter as per Equation 6.1 and Equation 6.2. Then, each test sample was mapped onto the dual face-space, thus obtaining the encodings $Y^{\mathscr{E}} = V^{\top}_{overall}X$ and $Y^{\mathscr{I}} = \tilde{Y}^{\mathscr{E}} = \tilde{V}^{\top}_{overall}X$, respectively used during the expression and identity recognition tasks for the dual face-space setting. Similarly, the Δ face-space encodings were obtained as $Y^{\Delta\mathscr{E}} = V^{\Delta\top}_{overall}X$ and $Y^{\Delta\mathscr{I}} = \tilde{Y}^{\Delta\mathscr{E}} = \tilde{V}^{\Delta\top}_{overall}X$, respectively used during the expression and identity recognition tasks for the Δ face-space setting. Recall that

[2]https://au.mathworks.com/products/matlab.html

	Accuracy mean (%)	Accuracy SME (%)	Best dimension
Baseline	68.16	7.08	N/A
Dual face-space	79.39	4.63	25
Δ face-space	82.04	3.4	15

Table 6.1 Expression recognition rates for the best dimensions among the compared models.

	Accuracy mean (%)	Accuracy SME (%)	Best dimension
Baseline	87.96	1.58	N/A
Dual face-space	94.90	1.16	65
Δ face-space	93.27	1.32	175

Table 6.2 Identity recognition rates for the best dimensions among the compared models.

$\tilde{V}$ is the matrix V with its columns disposed in an inverse order, and that $\tilde{Y}$ is the set of codings Y having their components sorted in an inverse order.

The classification process used a nearest neighbour algorithm. For each sample x_i, y_i and y_i^{Δ} were computed the Euclidean distances with respect to the centroids of each class (*i.e.* the prototypical identities in the case of identity recognition or the prototypical expressions in the case of facial expression recognition) in the corresponding space (*i.e.* respectively the perceptual space, the dual face-space, and the Δ face-space) selecting the class associated with the centroid having the lower distance from the sample.

For the encodings y_i and y_i^{Δ} this process was repeated for each dimension $k = [1, \ldots, 200]$ by taking only the first k components of the encodings.

In classifying each raw observation x_i for a baseline comparison, I considered all the pixel values of the input image as coordinates of points in a $\mathscr{D}$-dimensional space. This resembles a basic features matching strategy unmediated by the proposed face-space frameworks.

Results

Figures 6.1a and 6.1b show, respectively, the results of facial expression (*i.e.* the 7 facial expressions) and identity recognition tasks (*i.e.* the 70 identities). Tables 6.1 and 6.2 summarises the recognition rates for the best dimensions of the compared models. The results show that the novel Δ face-space can still achieve better facial expression and identity recognition rates compared to a baseline approach. In addition, the recognition accuracy between the dual face-space model proposed in Chapter 5 and the new Δ face-space model are really modest.

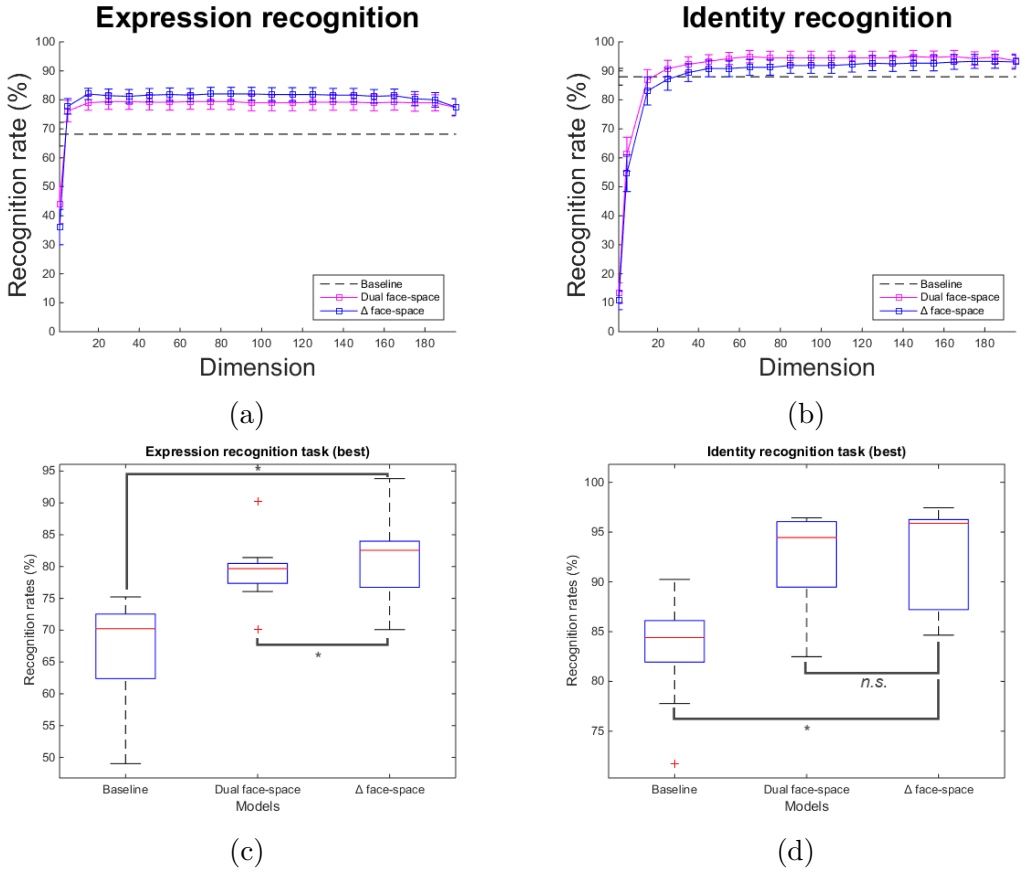


Figure 6.1 Comparative analysis of the models' performance. (a,b) The average recognition rates, among the 15-fold cross validation, for expression and identity recognition along the different dimensions of the space. The error bars are the standard error of the mean (SEM). (c,d) The recognition rates of each model considering the most frequently predicted class of each test sample among the predictions available in each dimensions k (mode) and during each of the 15 cross-validation tests for expression and identity recognition tasks. * $p < .05$.

For each of the 15 iterations the recognition rates of each dimension $k = [1, \ldots, 200]$ were computed by taking only the first k components of the encodings when classifying the test stimuli. For this reason, during each cross-validation iteration, the dual face-space and Δ face-space provided k predictions. In order to get a single prediction for each test sample during each cross-validation iteration (similarly to the baseline approach), for each test sample was selected the most frequently predicted class (mode) among the available k predictions. For each cross-validation iteration, I then computed the overall recognition rate in both facial expression and identity recognition conditions. This resulted in 15 samples for each considered model and task.

My thesis argument will be invalid if the dual face-space model and Δ face-space model show significant differences in facial expression and identity recognition rates, with the dual face-space showing better performance of the Δ face-space. Hence, I computed statistical analyses to test if there are such significant differences and, if so, these differences lead the dual face-space to act better than the Δ face-space. The distribution of the sampled recognition rates was first assessed for normality using a D'Agostino's K-squared test (D'Agostino and Pearson, 1973) finding that the samples from both facial expression and identity tasks followed a normal distribution (p-values respectively 0.7423 and 0.1198). The differences between models were evaluated by a two-tailed Student's t-test (Keppel, 1991) at a significant level of $\alpha = 0.05$. The t-test found a significant effect between the dual face-space and the proposed Δ face-space for the facial expression recognition task (p-value=0.0031), but no effect for face identity recognition task between the two models (p-value=0.1086). The effect size for the facial expression task was assessed by computing Cohen's d (Cohen, 1977). It resulted in a negative large effect size ($d = -0.92 < -0.8$). Hence, this result reveals that the observed effect refers to an *enhancement*, not an impairment, in facial expression classification performance for the novel Δ face-space compared to the dual face-space. The t-tests on the differences between the proposed Δ face-space and the baseline approach found significant effects for both facial expression (p-value=2.0e$-$5) and identity recognition tasks (p-value=0.0014), thus rejecting the null hypothesis for both the two tasks. The Cohen's ds suggested large effect size for both facial expression and identity recognition tasks (facial expression $d = 1.6 > 0.80$, facial identity $d = 1.0 > 0.80$).

6.2.3 Discussion

The present results demonstrate that to implement the proposed Δ face-space it is sufficient to interpret the sensory-motor information of face stimuli plausibly available from embodied mechanisms. This face-space can still facilitate both facial expression and identity classification. In particular, the statistical test on facial identity recognition proves that there is no sufficient evidence to demonstrate significant differences between the recognition rate of the two models, whereas the statistical test on facial expression recognition demonstrates significant differences between the two models, with the novel Δ face-space performing better than the dual face-space proposed in Chapter 5. This is an unexpected result since the weight matrix W^{Δ} used to implement the Δ face-space introduces some

misclassification errors with respect to the correct identity labels, contrary to the optimal weight matrix $W^{\mathscr{I}}$ used to implement the dual face-space, which used ground truth identity labels. Thus, I expected to find no significant differences or at most better performance for the dual face-space.

A possible explanation for this effect can be found in the intrinsic twofold structure of the face-space deriving from the used objective functions. Since the objective function in Equation 6.1 is defined as a minimisation of a ratio, it is plausible to assume that noise in the denominator would lead to an optimal solution facilitating more the constraint posed by the numerator at the expense of the one posed by the denominator (Bolker, 1966). Nevertheless, the noise in the denominator weight matrix was not enough to detect significant effects during the facial identity recognition task. At present, the proposed explanation is not definitive yet, and more investigations are needed to understand the real nature of the observations.

In addition, since the statistical test between the Δ face-space and the baseline approach found significant differences with large effect size for both facial expression and identity recognition, the Δ face-space is still capable of promoting both facial expression and identity recognition mechanisms compared to basic features matching strategies.

Therefore, these results, together with the ones in Chapter 4 and Chapter 5, are enough to validate the thesis proposed in this dissertation:

> *Sensory-motor information of face stimuli is sufficient to facilitate the acquisition of face recognition capabilities because this information is available early in life via embodiment mechanisms able to map sensory information of novel face stimuli, encountered throughout social exchanges, onto corresponding motor representations.*

In fact:

- In this chapter I demonstrated that in order to implement a face-space promoting facial expression and identity recognition capabilities to interpret the sensory-motor information of face stimuli plausibly available from embodied mechanisms, *i.e.* the set of facial expression labels exhibited by the face stimuli;

- In Chapter 4, I demonstrated that this set could be plausibly approximated by sensory-motor embodied mechanisms realising bodily formatted representations of face stimuli promoting classification;

- The proposed embodied account can be implemented by using a set of self-centred motor potentials. The literature presented in Chapter 2 provided evidence suggesting that the neural mechanisms of such process are likely available from birth, although experience and associative learning can refine them.

6.3 Embodied Mechanisms Constitute Social Cognition

In addition to the validation of the main thesis introduced in Chapter 1, I will provide preliminary evidence in favour of embodied cognition theories by means of the *constitution hypothesis* (see Section 2.3.2 on page 53 for an overview).

First, recall the definition of *constituent* as proposed by Shapiro (2010):

> *A process* **X** *is a central or important constituent of a process* **Y** *if* **Y** *would fail or be something else without* **X***'s presence.*

Then, the secondary objective of this dissertation is to show *the plausibility* of the following hypothesis:

Hypothesis 6.1 *Embodied simulation is constituent of face processing mechanisms, and, since face processing is vital for social cognition, embodied simulation is a constituent of social cognition.*

Thus, in order to advance this hypothesis, it is necessary to show that face processing mechanisms would fail or be something else without (or with impaired) embodied mechanisms.

In Chapter 4 (Section 4.4.3 on page 139), I showed that by using a non-embodied matching strategy the classification performance was worse than employing embodied simulation mechanisms. Therefore, it is likely to expect that absent or impaired embodied mechanisms would realise an altered set $\ell_{X_{dev}}$.

Proposition 6.9 *Absent or impaired embodied mechanisms would likely realise altered motor interpretations $\ell_{X_{dev}}$ of the observed face stimuli.*
Proof: *see the classification performance compared to a non-embodied matching strategy in Section 4.4.3 on page 139.*

Hence, in this chapter, I aim to investigate the effects that impaired embodied mechanisms can plausibly have on face processing capabilities and to demonstrate

that an altered set $\mathscr{E}_{X_{dev}}$, following impaired embodied simulation mechanisms, will realise a dysfunctional face-space exhibiting significantly impaired facial expression and identity classification skills:

Proposition 6.10 *By employing an altered set of facial expressions classes $\mathscr{E}_{X_{dev}}$, plausibly realised without or with impaired embodied mechanisms, it would result a dysfunctional face-space showing significantly impaired facial expression and identity classification capabilities.*

6.3.1 Simulating Embodied Mechanisms Impairments

In Chapter 3, I proposed Hypothesis 3.1 suggesting that face processing mechanisms lie on two dimensions of embodied mechanisms: the sensory-motor dimension and the visceral dimension:

> *The information provided by embodied mechanisms, shaped as bodily formatted representations, lies on at least two dimensions: a sensory-motor dimension, and a visceral dimension. The sensory-motor dimension describes perceptual aspects of the perceived action, such as the motor potentials, the viewpoint and the pose. The visceral dimension specify emotional aspects of the perceived action, such as feelings and sensations associated with the perceived social stimuli.*

In Chapter 3, I showed that autism and schizophrenia populations exhibit impaired facial expression and identity recognition capabilities, whereas psychopaths exhibit deficits on facial expression recognition only. After providing a discussion of the available clinical evidence, I suggested that face processing impairments in autism and schizophrenia patients can be explained by corrupted sensory-motor information, whereas face processing impairments in psychopaths can found an explanation in absent or insufficient attention to peripheral visceral information.

Also, in Chapter 3, I proposed that the sensory-motor dimension has a crucial role in determining the correct motor configuration exhibited by a face stimulus, whereas the visceral dimension assists this classification process by providing additional visceral information able to facilitate the discrimination of perceptually similar or opaque motor configurations (see Figures 3.2c and 3.2b on page 102). I introduced Hypothesis 3.2 suggesting that:

> *Alterations of the sensory-motor embodied process of face stimuli significantly impair both facial identity and facial expression recognition*

capabilities. Alterations of the visceral embodied process of face stimuli significantly impair facial expression recognition capabilities only.

Denote with $\hat{\ell}_{X_{dev}}$ a set of facial configurations classes realised by impaired sensory-motor dimension of embodied mechanisms, and with $\check{\ell}_{X_{dev}}$ a set of facial configurations classes realised by impaired visceral dimension of embodied mechanisms. These two sets approximate the impaired sets of facial expression classes $\hat{\mathscr{E}}_{X_{dev}}$ and $\check{\mathscr{E}}_{X_{dev}}$ respectively (see Proposition 6.7 on page 179).

Since autism and schizophrenia people show a more generalised impairment on all the facial expressions plausibly explained by altered sensory-motor information, whereas psychopaths show particularly marked impairments of their bodily responses when processing fearful, painful and sad signals compared to other emotions (Blair, 1999; Blair et al., 1997) and they present accentuated deficits in processing facial expressions of fear and sadness compared to other emotions (Dawel et al., 2012), I propose to computationally simulate the impairment of embodied mechanisms by using the following methodology:

1. Choose a percentage ε denoting the extent of the embodiment deficits;

2. Denote with $\mathscr{E}_{true}(x_i)$ the correct facial expression class exhibited by the sample x_i and included in the set $\mathscr{E}_{X_{dev}}$. Then, there is a chance of $\varepsilon\%$ that the facial expression class $\hat{\mathscr{E}}(x_i) \in \hat{\mathscr{E}}_{X_{dev}}$ would be different from the correct class $\mathscr{E}_{true}(x_i)$;

3. Denote with $\mathscr{E}_{true}(x_i)$ the correct facial expression class exhibited by the sample x_i and included in the set $\mathscr{E}_{X_{dev}}$. Then, if $\mathscr{E}_{true}(x_i) =$ *'fear'* or $\mathscr{E}_{true}(x_i) =$ *'sadness'*, there is a chance of $\varepsilon\%$ that the facial expression class $\check{\mathscr{E}}(x_i) \in \check{\mathscr{E}}_{X_{dev}}$ would be different from the correct class $\mathscr{E}_{true}(x_i)$;

4. If the class for the sample x_i is misclassified, the misclassified class would be the one perceptually closer to x_i among the available alternative facial expression classes.

Given the current state of knowledge reviewed in Chapter 3, the proposed methodology is plausible. However, there is only partial indirect support for it. Thus, I will draw conclusions significantly advancing embodied cognition research. However, these ones would not be definitive and further work will be necessary.

6.3.2 Argument Validation

In this section, I will evaluate the performance of the proposed Δ face-space against corrupted versions of it. I will demonstrate that:

(i) Simulating an alteration of the sensory-motor dimension of embodied mechanisms leads to generalised impairments in facial expression and identity recognition;

(ii) Simulating an alteration of the visceral dimension of embodied mechanisms leads to generalised impairments in facial expression recognition only.

The simulated impairments can be explained by dysfunctional or absent embodied mechanisms, and they result in face processing deficits. In other words, without or with impaired embodied simulation mechanisms, the face processing skills would fail or be significantly impaired. Therefore, the evidence presented in this section would advance the plausibility of Hypothesis 3.2 and Hypothesis 6.1, at least at the current state of knowledge and through simulated computational data.

Dataset

The experiments proposed in this section would again evaluate the model using the Karolinska Directed Emotional Faces (KDEF) dataset (Lundqvist et al., 1998) (refer to Section 5.5 on page 164 for details).

Similarly to the previous experiments, were used only frontal face stimuli. The facial region was extracted from the images and its resolution reduced to 80×80 pixels. Eyes and mouth were at approximately the same position. Illumination variations were reduced by applying a simple equalisation process to the images.

The data was first pre-processed by submitting the pixels of the images in input to a PCA, as explained in Section 5.4 on page 159. In all the experiments were retained the components able to explain 95% of the variance of the original data resulting in 200 components. The PCA output was used to estimate the 200 components of the used face-space models.

Procedure

The present experiments used repeated random iterations of the dataset's samples. In each iteration, 25 identities were randomly selected as test set among the 70 available identities in the dataset to simulate unfamiliar identities. For each of

the 25 selected identities, 2 facial expressions were randomly chosen as training observations for identity recognition task, and the remaining 5 facial expressions as the test set, leading to a total of 125 test samples for each iteration. The images of the other 45 identities, together with the 50 selected training samples, were used as the training set for the current iteration. Therefore, each iteration included:

- a training set of 365 face stimuli: 315 face stimuli of all the 7 facial expressions for 45 identities, plus 50 face stimuli of 2 randomly selected facial expressions for 25 unfamiliar identities;
- a test set of 125 face stimuli of the remaining 5 facial expressions for the 25 unfamiliar identities.

The process was repeated 35 times. This methodology and number of repetitions were chosen in order to reduce the random effects due to the alteration of the set $\mathscr{E}_{X_{dev}}$.

For each iteration were created three Δ face-space models with the following facial expression sets:

- A preserved set $\mathscr{E}_{X_{dev}}$ containing the correct facial expressions associated with the training stimuli X_{dev};
- An impaired set $\hat{\mathscr{E}}_{X_{dev}}$ including a percentage $\varepsilon = \{1, 5, 10, 15, 20\}$ of misclassified facial expressions, as explained before in this section (paragraph "Procedure");
- An impaired set $\check{\mathscr{E}}_{X_{dev}}$ including a percentage $\varepsilon = \{1, 5, 10, 15, 20\}$ of misclassified facial expressions limited to expressions of fear and sadness, as explained before in this section (paragraph "Procedure").

In the experiments I compared the classification performance between the impaired models and the control face-space model for both facial expression and identity recognition tasks. The performance was measured with respect to the classification of the test stimuli during each iteration, leading to 35 measurements for each model for both facial expression and identity conditions. I also measured the performance of a baseline approach. This approach considered all the pixels of the face stimuli when matching similar facial expressions or identities. The models' performance were assessed with the same procedure described in Section 6.2.2 (page 183).

The proposed face-space models can use the first $k = 1, \ldots, d$ components of the mapping matrix $V^{\Delta}_{overall}$ to map the face stimuli in face-space representations and perform recognition tasks. Thus, the face-space models provided d predictions for each test sample during each of the 35 iterations. To gather a single prediction for each test sample during each iteration, I selected the most frequent class (mode) predicted by the face-space model, as per a majority voting approach. For each iteration, I then computed the overall recognition rate for the unimpaired control Δ face-space and the rest of impaired face-space models for both facial expression and identity recognition settings. This process led to 35 samples for each considered model and task.

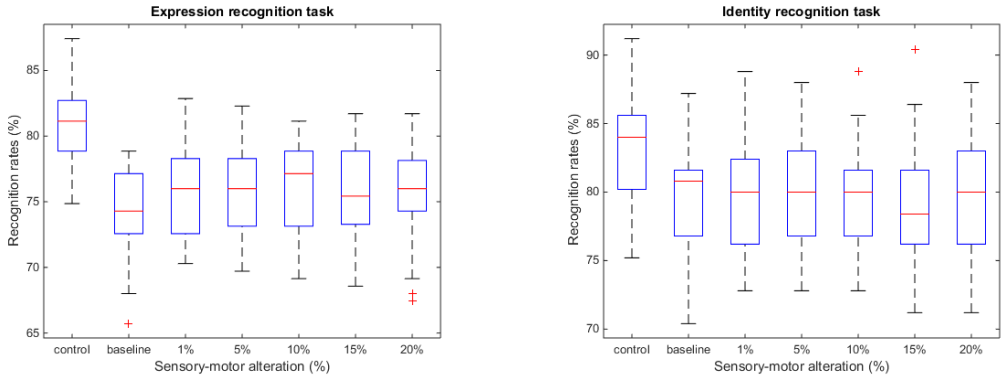
Results

The distribution of the sampled recognition rates was first assessed for normality using a D'Agostino's K-squared test (D'Agostino and Pearson, 1973) finding that the samples from both facial expression and identity tasks in both the conditions (*i.e.* sensory-motor and visceral dysfunctions) followed a normal distribution (p-values > 0.05). Thus, the differences between the control model and the impaired models were evaluated using a two-tailed Student's t-test (Keppel, 1991) at a significance level of $\alpha = 0.05$. Due to the multiple comparisons with the control model (*i.e.* 5 comparisons for each independent measure of classification accuracy), the significance level was corrected with a Bonferroni correction, leading to a significance level of $\alpha = 0.01$.

Results for impaired sensory-motor embodiment. Figure 6.2 and Table 6.3 summarise the results for simulated impairment of sensory-motor embodiment

The models with simulated dysfunctional sensory-motor embodiment resulted significantly impaired in all the considered levels of alteration and task (p-value < 0.01) compared to the control face-space model. The observed effect size was large, and the statistical tests reached high power $(1 - \beta)$ during all the considered settings ($d > 0.80$ and $\beta < 0.01$).

Importantly, by testing the differences between the baseline approach and the face-space models with impaired sensory-motor embodiment during facial expression classification, I found significant effects (p-values < 0.01) resulting in medium to large effect size and high power ($d \approx 0.7$ and $\beta < 0.17$) for all the considered levels of alteration. On the contrary, I did not find any significant dif-



(a) (b)

Figure 6.2 Recognition rates of the considered models from the 35 iterations during the simulated sensory-motor dysfunctions setting. (a) Expression recognition task. (b) Identity recognition task.

Alteration %	Is expression recognition impaired?
1	Yes (p-value=1.5e−14, d=2.15)
5	Yes (p-value=1.7e−15, d=2.32)
10	Yes (p-value=4.5e−16, d=2.44)
15	Yes (p-value=3.7e−14, d=2.09)
20	Yes (p-value=1.0e−14, d=2.19)
Alteration %	Is identity recognition impaired?
1	Yes (p-value=3.7e−10, d=1.46)
5	Yes (p-value=1.1e−8, d=1.26)
10	Yes (p-value=8.0e−8, d=1.14)
15	Yes (p-value=3.5e−12, d=1.76)
20	Yes (p-value=1.7e−8, d=1.23)

Table 6.3 The outcomes of the statistical analyses for impaired sensory-motor embodiment setting.

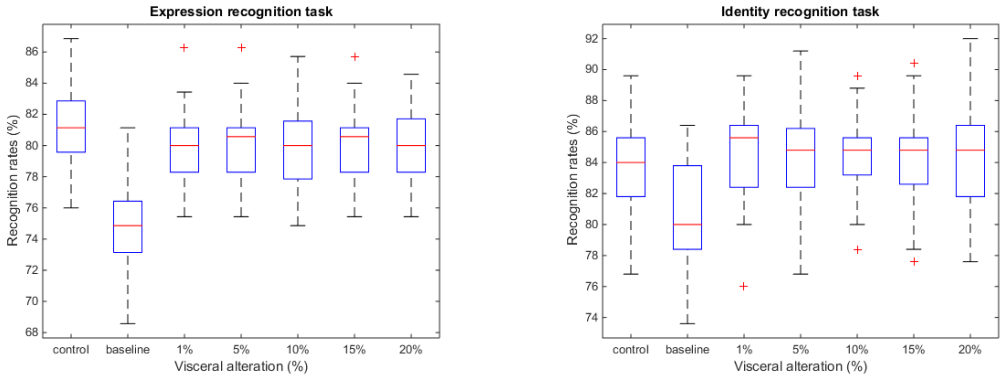
ference between the baseline approach and the impaired face-space models during facial identity discrimination task (p-values > 0.05). This evidence suggests that the altered models were still able to facilitate expression recognition compared to the baseline approach, although presenting significantly worse recognition rates than the control model. Nevertheless, the same did not happen for facial identity recognition, where simulated alterations of the embodied mechanisms significantly altered facial identity classification even compared to a baseline approach. Therefore, these results cannot be explained by simple causal interactions between the proposed alterations and the observed dysfunctions in the models. Rather, these results point out a *constitutional* relationship demonstrating that alterations of the sensory-motor embodiment negatively impact on face processing capabilities: *without or with dysfunctional embodied mechanisms facial expression and, in particular, facial identity recognition would fail.*

Results for impaired visceral embodiment. Figure 6.3 and Table 6.4 illustrate the results for the simulated impairment of visceral embodiment.

When simulating a dysfunctional visceral embodiment, the control and impaired models significantly differed during facial expression recognition task (p-value < 0.01) for all the considered levels of alteration. The effect size for facial expression recognition task was large ($d > 0.80$) and reached high power ($\beta < 0.01$) among all the considered levels of alteration.

On the contrary, during facial identity recognition task were not observed significant differences between the control model and the impaired ones (p-value > 0.01). More importantly, the impaired models performed better than the control model (Cohen's $d < 0$), similarly to what observed in the previous experiments presented in this chapter (refer to Section 6.2.2 for a brief discussion).

The visceral impaired models were tested against the baseline approach to test significant differences. In both facial expression and identity task were found significant effects (p-values < 0.05) for all the considered levels of alteration, with the impaired models presenting better performance than the baseline approach. This means that the models with simulated impaired visceral embodiment were still able to present better performance compared to a basic feature matching strategy as the baseline approach for both facial expression and identity recognition.



(a) (b)

Figure 6.3 Recognition rates of the considered models from the 35 iterations during the simulated visceral dysfunctions setting. (a) Expression recognition task. (b) Identity recognition task.

Alteration %	Is expression recognition impaired?
1	Yes (p-value=3.9e−8, d=1.18)
5	Yes (p-value=2.3e−6, d=0.95)
10	Yes (p-value=1.3e−7, d=1.11)
15	Yes (p-value=4.4e−9, d=1.31)
20	Yes (p-value=1.9e−8, d=1.23)
Alteration %	Is identity recognition impaired?
1	No (p-value=0.0712, d=-0.31)
5	No (p-value=0.0778, d=-0.30)
10	No (p-value=0.2993, d=-0.17)
15	No (p-value=0.1244, d=-0.26)
20	No (p-value=0.1147, d=-0.27)

Table 6.4 The outcomes of the statistical analyses for impaired visceral embodiment setting.

6.3.3 Discussion

The present results demonstrate that simulating an impairment of embodied mechanisms significantly affects the development of the proposed face-space. This, in turn, impairs facial expression and identity recognition capabilities.

In particular, simulating a sensory-motor dysfunction of embodied mechanisms significantly affects both facial expression and identity recognition, whereas simulating impairments of visceral embodiment results in a significant effect for facial expression recognition only. Moreover, during the visceral condition, facial identity recognition is not affected even when the visceral dimension is altered significantly (20% of alteration). Instead, for an alteration of just 1% facial expression recognition exhibited significant differences with large effect size. Furthermore, compromised sensory-motor embodiment significantly affected facial identity recognition even compared to a baseline approach. Therefore, the provided results are sufficient to validate Proposition 6.10. Here it is important to acknowledge that whereas a sensory-motor dysfunction impairs all the facial expressions (7/7), a visceral dysfunction impairs only approximately the 28% of facial expressions (2/7), therefore leading to a much lower amount of impaired samples. However, this is not a limitation of the used methodology, because this is simply simulating what expected to happen in human subjects exhibiting sensory-motor or visceral dysfunctions. These results suggest that by comparing the two types of dysfunctions there are different outcomes in the acquired face-spaces resembling the ones discussed in Chapter 3.

Importantly, the present results significantly advance Hypotheses 3.1 and 3.2 introduced previously in this dissertation (page 100), although they are not enough for drawing definitive conclusions. Specifically, since simulating a dysfunction in the sensory-motor dimension affected both facial expression and identity recognition, whereas simulating visceral impairments affected facial expression recognition only, the present results add computational evidence in favour of Hypothesis 3.2.

Finally, the results illustrated in this chapter advance Hypothesis 6.1 (page 189). Assuming that the set of facial expression classes in the training set is plausibly approximated by the embodied mechanisms proposed in Chapter 4, a failure of these mechanisms would lead to altered facial expression classes. In this chapter, I demonstrated that by altering even a small amount (*i.e.* 1%) of these facial expression classes would lead to significant impairments of both facial expression and identity recognition capabilities. Since face-to-face interactions are crucial for the development of social cognition (Grossmann, 2015), a failure of

embodied mechanisms would lead to a failure of face processing mechanisms, and these impairments would significantly impact on social cognition development.

6.4 Conclusions

The objective of this chapter was to provide evidence in favour of the thesis suggested in this dissertation. Thus, I started with providing the links between the embodied mechanisms modelled in Chapter 4 and the facial identity recognition mechanisms modelled in Chapter 5 in order to advance my argument.

However, the evidence available so far was not enough to validate the suggested argument. Therefore, in this chapter, I proposed to generalise the face-space framework provided in Chapter 5 in order to fill the missing gaps preventing the validation of the thesis argument presented in this dissertation. I showed that the dual face-space proposed in Chapter 5 could be implemented without knowing the facial identity classes of the training stimuli, but only their facial expressions. This information can be plausibly approximated by embodied mechanisms providing correct interpretations of facial motor information, as suggested in Chapter 4.

The present experimental results, gathered from computational simulations, provided additional quantitative evidence in favour of my thesis. In fact, I demonstrated that the recognition performance of the novel model is not significantly different from the one proposed in the previous chapter and that the novel model can significantly enhance classification of facial expressions and identities compared to a baseline approach.

With this ultimate computational tool available, I was able to investigate the effects that embodiment impairments would have on face processing capabilities. In particular, I wanted to advance the plausibility of Hypothesis 3.2, suggesting that alteration of sensory-visceral embodiment would realise generalised deficits in facial expression and identity recognition, whereas damages of the visceral embodiment would realise impairments of facial expression recognition only. The experimental results demonstrated the plausibility of this claim by clearly showing significant effects of sensory-motor embodiment dysfunctions on both facial expression and identity recognition performance, and by demonstrating an effect of visceral embodiment alteration limited to facial expression recognition mechanisms.

The evidence used to promote Hypothesis 3.2 can also advance Hypothesis 6.1, suggesting that embodied mechanisms are constituents of social cognition. Although the evidence provided in favour of Hypotheses 3.2 and 6.1 is limited by

plausible but not yet validated assumptions, this dissertation still provides a significant first contribution favouring embodied cognition theories.

In the remainder of this dissertation, I will provide a summary of the research gaps I advanced and of the primary and secondary contributions of this work. In addition, I will evaluate my thesis on several desirable features, namely innovation, falsifiability, parsimony, integrability and plausibility.

Chapter Bibliography

Blair, R. J. R. (1999). Responsiveness to distress cues in the child with psychopathic tendencies. *Personality and Individual Differences*, 27(1):135–145.

Blair, R. J. R., Jones, L., Clark, F., and Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, 34(2):192–198.

Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., and Lanzarotti, R. (2018). Deep construction of an affective latent space via multimodal enactment. *IEEE Transactions on Cognitive and Developmental Systems*.

Bolker, E. D. (1966). Functions resembling quotients of measures. *Transactions of the American Mathematical Society*, 124(2):292–312.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

D'Agostino, R. and Pearson, E. (1973). Tests for departure from normality. empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika*, 60(3):613–622.

Dawel, A., O'Kearney, R., McKone, E., and Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience & Biobehavioral Reviews*, 36(10):2288–2304.

De Heering, A., De Liedekerke, C., Deboni, M., and Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect. *Developmental Science*, 13(1):181–187.

Gallese, V. (2001). The 'shared manifold' hypothesis. From mirror neurons to empathy. *Journal of Consciousness Studies*, 8(5-6):33–50.

Gallese, V. and Caruana, F. (2016). Embodied simulation: Beyond the expression/experience dualism of emotions. *Trends in Cognitive Sciences*.

Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., and Goodale, M. A. (2005). The involvement of the "fusiform face area" in processing facial expression. *Neuropsychologia*, 43(11):1645–1654.

Goodman, G. S., Sayfan, L., Lee, J. S., Sandhei, M., Walle-Olsen, A., Magnussen, S., Pezdek, K., and Arredondo, P. (2007). The development of memory for own- and other-race faces. *Journal of Experimental Child Psychology*, 98(4):233–242.

Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological Bulletin*, 141(6):1266.

Grossmann, T. and Vaish, A. (2009). Reading faces in infancy: Developing a multi-level analysis of social stimulus. In Striano, T. and Reid, V., editors, *Social Cognition: Development, Neuroscience and Autism*. Blackwell Publishing, Oxford, UK.

Iacoboni, M. (2009). Do adolescents simulate? Developmental studies of the human mirror neuron system. In Striano, T. and Reid, V., editors, *Social Cognition: Development, Neuroscience and Autism*. Blackwell Publishing, Oxford, UK.

Kadosh, K. C., Luo, Q., de Burca, C., Sokunbi, M. O., Feng, J., Linden, D. E., and Lau, J. Y. (2016). Using real-time fMRI to influence effective connectivity in the developing emotion regulation network. *NeuroImage*, 125:616–626.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* . Prentice-Hall, Inc.

Leo, I., Angeli, V., Lunghi, M., Dalla Barba, B., and Simion, F. (2018). Newborns' face recognition: The role of facial movement. *Infancy*, 23(1):45–60.

Lundqvist, D., Flykt, A., and Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, pages 91–630.

Meissner, C. A. and Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, 7(1):3.

Meltzoff, A. N. and Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, pages 702–709.

Meltzoff, A. N. and Moore, M. K. (1992). Early imitation within a functional framework: The importance of person identity, movement, and development. *Infant Behavior and Development*, 15(4):479–505.

Ngo, T. T., Bellalij, M., and Saad, Y. (2012). The trace ratio optimization problem. *SIAM Review*, 54(3):545–569.

Niedenthal, P., Wood, A., and Rychlowska, M. (2014). Embodied emotion concepts. *The Routledge Handbook of Embodied Cognition*, pages 240–249.

Palermo, R. and Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*, 45(1):75–92.

Pascalis, O., de Haan, M., and Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, 296(5571):1321–1323.

Rhodes, G., Pond, S., Burton, N., Kloth, N., Jeffery, L., Bell, J., Ewing, L., Calder, A. J., and Palermo, R. (2015). How distinct is the coding of face identity and expression? Evidence for some common dimensions in face space. *Cognition*, 142:123–137.

Sariyanidi, E., Gunes, H., and Cavallaro, A. (2015). Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6):1113–1133.

Shapiro, L. (2010). *Embodied Cognition*. Routledge.

Simion, F. and Di Giorgio, E. (2015). Face perception and processing in early infancy: Inborn predispositions and developmental changes. *Frontiers in Psychology*, 6.

Trevarthen, C. (2006). The concept and foundations of infant intersubjectivity. In Bråten, S., editor, *Intersubjective Communication and Emotion in Early Ontogeny*, pages 15–46. Cambridge University Press.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.

Valentine, T., Lewis, M. B., and Hills, P. J. (2015). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, (ahead-of-print):1–24.

Yankouskaya, A., Humphreys, G. W., and Rotshtein, P. (2014). The processing of facial identity and expression is interactive, but dependent on task and experience. *Frontiers in human neuroscience*, 8.

I am turned into a sort of machine for observing facts and grinding out conclusions.

— Charles Darwin —

7

Conclusions and Final Remarks

I started this dissertation with some philosophical questions: *does the body shape the mind? Is there something beyond the brain in developing cognition?* With this work, I provided computational evidence to hint at plausible answers to these questions. The acquisition of face recognition skills can be explained by embodied mechanisms having the function of giving interpretations to observed facial motor behaviour (Chapter 4). These interpretations assist the development of a psychological face-space enabling facial expression and identity recognition skills (Chapters 5 and 6). The presented embodied mechanisms are deeply related to aspects of the body, and they crucially assist in shaping cognition. I showed that:

- The bodily mental representations assisting embodiment processes are centred in the cogniser's own body. What the subjects can interpret depends on the motor acts and the visceral reactions the subjects can correctly reproduce within their body (Chapter 4);
- Embodied simulation processes are not only used to understand others' actions and to attribute mental states to them. These mechanisms have broader functions central to other aspects of cognition, such as face recognition (Chapter 5);

- Impaired embodied mechanisms negatively impact on face processing capabilities, thus suggesting that embodied simulation is a constituent of face processing skills (Chapter 6);

- Face processing capabilities are of paramount importance for social cognition development (Section 2.2), thus embodied simulation is a central mechanism shaping cognition (Section 2.3).

Although I demonstrated the plausibility of the proposed hypotheses offering new exciting perspectives on human cognition, I did not discuss the implications of the present findings yet. Therefore, the objective of this chapter is to step back to the original intentions presented in Chapter 1 and to demonstrate that the insights offered in this dissertation are enough to reach the suggested research objectives. However, before doing so, in Section 7.1, I reinforce the validity of the discussed thesis argument by presenting its ability to be innovative, falsifiable, parsimonious, integrable with previous findings and plausible. Then, in Section 7.2, I will analyse each of the contributions and research gaps discussed in Chapter 1, demonstrate their consistent accomplishment, and explain their implications for the society and the research community, as proposed in Section 1.3. In Section 7.3, I will discuss how the present findings well integrate with face processing studies. Finally, in Section 7.4, I will conclude with the aimed future works.

7.1 Additional Values of the Thesis Argument

In Chapter 6, I summarised my thesis argument into a set of propositions and demonstrated their validity by referencing the supporting evidence. This methodology provided direct computational evidence in support of my thesis argument.

In this section, I aim to reinforce the discussed thesis by assessing it against a set of additional desirable properties, namely innovation, falsifiability, parsimony, integrability and plausibility.

7.1.1 Innovation

A novel hypothesis or theory must be innovative, otherwise there are no benefits to advance the considered hypothesis. A hypothesis (or theory) is considered to be innovative when it can potentially disrupt current understanding of the investigated topic and offer significant insights to foster the research community. Therefore, in this section I will answer to the following question:

Is the new hypothesis provocative enough to disrupt current perspectives, thus pushing face processing research forward?

Traditionally, facial expression and identity processing are described as distinct capabilities making use of distinct cognitive mechanisms. In this work, I showed through computational simulations that face processing development can be deeply linked to aspects of the body and that facial expression and identity are crucially interdependent processes. Hence, in this dissertation I demonstrated that it might be necessary to revise previous theories in face processing in order to accommodate this new embodied understanding. As I will discuss in Section 7.3, the explanations presented in this dissertation can unify traditional and modern understandings of face processing with an innovative perspective. Therefore, the discussed thesis argument offers an innovative contribution able to push forward face processing research.

7.1.2 Falsifiability

Falsifiability is a key element of what a theory is. Unless it is falsifiable, it is not a theory. A hypothesis (or theory) is falsifiable whenever it is possible to test it and gather direct evidence for its validation or rejection. This evaluation is not possible when the hypothesis (or theory) holds on thought experiments or arguments lying on abstract metaphors. Therefore, in this section I will answer the following question:

Is it possible to conduct an experiment which could falsify my argument that sensory-motor information is sufficient for the acquisition of face recognition skills?

In this dissertation, I tested the validity of the proposed thesis by gathering direct evidence from computational simulations of embodied mechanisms and face processing models. Since I provided a computational description of the mechanisms underlying these capabilities, it is possible to test the proposed thesis argument by gathering evidence from human experiments.

For example, human face processing exhibits what are called *"contingent face aftereffects"*, namely illusory properties of the observed face stimulus that are apparent only after being sufficiently exposed to another induction stimulus[1] for a

[1] An induction stimulus is a stimulus the subject is required to observe for a certain amount of time before the real test stimulus, so to generate synaptic fatigue on neurons hypothesised to correlate with the task under investigation during the test phase.

sufficient period of time (Leopold et al., 2005). Researchers in face processing use this property of face stimuli in order to investigate their potential representation in the human brain (Rhodes and Jeffery, 2006).

Therefore, it may be possible to design the following experiment using the following conditions:

- In the first condition the subject observe an induction stimulus exhibiting a face that is represented in the face-space framework in such a way that many of its expression components interacts with identity components (see the plot in Figure 5.4 on page 168 for more insights);
- In the second condition the subject observe an induction stimulus exhibiting a face that is represented in the face-space framework in such a way that just a few or no expression components interact with the identity components of the stimulus.

Thus, by using the same set of face stimuli as test samples during both the conditions, it is possible to ask the subject to perform a face recognition task. If my hypothesis is correct, I will expect to find significantly more recognition errors in the first condition rather than in the second one. On the contrary, finding no differences between the two settings would provide evidence against my hypothesis.

In addition, neuroscience experimentation can provide additional evidence in support (or against) of my thesis argument. For example, one could use functional magnetic resonance imaging (fMRI) to monitor brain activity in motor areas when a subject is performing a face recognition task on unfamiliar identities. Absent or low activation in motor areas during the training stage (*i.e.* when the subject observes and memorises the new identities) can falsify my argument.

7.1.3 Parsimony

Parsimony is another valuable property for good hypotheses and theories, and it is of particular importance to guide theories in human cognition (Farah, 1995). An hypothesis (or theory) is parsimonious when it can explain observed phenomena by providing models having lower complexity than previous existing theories. Therefore, in this section I will answer to the following question:

> *Is the hypothesis more elegant and less complex than previous theories in face processing?*

Whereas traditional understanding of face processing suggests two separate routes for processing the expression and the identity of face stimuli (Bruce and Young, 1986; Haxby et al., 2000), I demonstrated that facial expression and identity processing are more closely processed than traditionally thought. In fact, it is possible to develop a single common representation able to facilitate both facial expression and identity recognition capabilities, without the need of distinct processes, thus reducing the overall complexity.

Furthermore, I suggest that it is possible to develop facial identity recognition capabilities by recognising the facial motor configuration exhibited by novel face stimuli and that this information is plausibly provided by embodied mechanisms available from birth. This again reduces complexity, since I demonstrated that the embodied mechanisms could be implemented with a set of self-centred motor stimuli of face dynamics, which are definitely less than a set of stimuli exhibiting different facial configurations and different identities often necessary in a computational model of face identity recognition.

Therefore, the present thesis offers a parsimonious understanding of face processing. However, the present argument introduces also some questions about its integrability and plausibility with respect to previous findings in neuroscience and human cognition. I will show in the following section and in Section 7.3 that my hypothesis is indeed able to integrate with previous findings and that it is well supported by other available evidence in neuroscience and cognitive studies.

7.1.4 Integrability

A new hypothesis should conveniently offer an understanding able to integrate new available evidence with existing one. Integrability is then the property of a hypothesis (or theory) to explain the newly available evidence by incorporating it with previous findings in the same field or neighbouring fields. In this section I will answer the following question:

> *Can the hypothesis be integrated with findings supporting other pre-existing theories in face processing?*

Although Haxby et al. (2000) suggest two separate neural areas for processing dynamic and invariant features of the face, their findings can be integrated with the novel insights offered in this dissertation. Haxby et al. (2000) argue that changeable aspects of the face (*i.e.* eye gaze, expression and lip movement) are processed in the Superior Temporal Sulcus (STS), whereas the Lateral Fusiform

Gyrus (LFG) deals with invariant aspects of the face necessary to classify the exhibited identity. In Section 7.3, I will discuss how the model presented in Chapter 4 can resemble the functions of the STS, whereas the face-space models presented in Chapters 5 and 6 enables face recognition skills as the LFG. I will argue that this interpretation can unify traditional and modern understandings of face processing by proposing an innovative interpretation. For a more in-depth and comprehensive argument I refer the reader to Section 7.3.

7.1.5 Plausibility

A good hypothesis or theory needs to be plausible. This feature is connected with integrability, in the sense that the offered novel perspective should not make claims or lie on assumptions obviously false. Therefore, in this section I will answer the following question:

Is the hypothesis free from conceptual problems?

In Section 7.3, I will suggest the plausibility of the present thesis argument from a neuroscience perspective, since able to well integrate traditional and modern findings in face processing research. In addition, there is a significant amount of literature in cognitive science showing that face stimuli are represented in multidimensional norm-based spaces[2] (Rhodes and Jeffery, 2006; Rhodes and Leopold, 2011; Rhodes et al., 2005) and that invariant and dynamic features of the face interact with one another (Ganel and Goshen-Gottstein, 2004; Ganel et al., 2005; Kadosh et al., 2016; Pell and Richards, 2013; Rhodes et al., 2015). In agreement with these studies, I demonstrated that the face-space model described in Chapters 5 and Chapter 6 is a norm-based space able to integrate invariant and dynamic features of the face within a shared representation leading to interesting interactions during face processing tasks.

7.2 Dissertation Objectives

In Sections 1.2 and 2.4, I outlined the contributions and research gaps advanced by the present research. I now revisit each of them to demonstrate that I have indeed followed through on my original ambitions and that the provided contributions are of significant value for the research community, as suggested in Section 1.3.

[2]see Section 5.4 for a mathematical interpretation of a norm-based space

7.2.1 Broad Contribution

The proposed broad contribution of this dissertation was:

> *Providing a better understanding of social cognition's core mechanisms and proposing plausible hypotheses connecting social cognition to embodied cognition theories.*

This work met the proposed contribution. In fact, I provided an in-depth understanding of social cognition by providing literature investigating the chosen topic under different perspectives and introduced some significant hypotheses advancing embodied cognition research program. In particular:

- I proposed a taxonomy of the core mechanisms plausibly underlying social cognition (Sections 2.1.3, 2.1.4 and 2.1.5);
- I offered valid links between these mechanisms and embodied cognition theories (Section 2.3.4);
- I reported studies linking face processing research to social cognition domain (Section 2.2) and tested the plausibility of my hypotheses through computational tools of face processing inspired by embodied cognition theories (Chapters 4, 5 and 6).

Addressed Research Gaps

The currently available literature provides only a very limited amount of computational models explaining the underlying mechanisms of mirroring process (Cangelosi and Riga, 2006; Oztop et al., 2006), which I suggested to be at the core of an embodied understanding of social cognition. In addition, the available computational accounts often investigate embodied simulation mechanisms with respect to motor control capabilities, and they tenuously link to other higher-level cognitive skills[3].

To address these gaps, in Chapter 4, I provided a computational model suggesting an explanation of embodied simulation mechanisms underlying a mind-reading episode occurring during face-to-face interactions. In addition, as demonstrated in Chapter 6, this model links to face processing capabilities, thus providing a valid connection between embodied mechanisms and cognition. The

[3]But see the work of (Boccignone et al., 2018) for a more recent extension of the theory proposed in this dissertation.

identified cognitive capabilities (*i.e.* face processing) are of particular importance for the development of social cognition (Grossmann, 2015, but see also Section 2.1.6 on page 34).

Significance

Given the summary provided in this section, the broad contribution of this dissertation is of significant value for at least the following reasons:

- The computational explanation of embodied mechanisms offered in this work can be used to facilitate social cognition research and to advance new hypotheses;
- The offered hypotheses and computational theories can be easily assessed by experiments with human subjects, thus allowing falsifiability of the proposed hypotheses;
- The probabilistic account of simulation mechanisms and the the face-space framework offered in this work can be extended and used by artificial intelligence research community to advance the current state of the art, specifically in artificial social learning and the promotion of long-term human-machine social interactions.

7.2.2 Primary Contribution

The aimed primary contribution of this dissertation was:

> *Providing a computational understanding of face perception and processing mechanisms and investigating the inter-dependencies between facial expression and identity processing.*

The present primary contribution was met by this dissertation. In fact, in Chapter 5, I proposed a novel hypothesis able to explain recent findings in human face processing studies. I validated the suggested hypothesis by extending a computational framework widely used in face processing, the face-space, thus providing a valuable computational tool advancing face processing research. In summary:

- My face-space framework offers innovative understandings of face processing able to integrate with evidence available from face studies (see Sections 5.2 and 7.3);

- The offered face-space framework can be used as a valid tool able to assist the validation of novel hypotheses in face processing research (Sections 5.4 and 6.2.1);
- I demonstrated that a single representation of face stimuli includes expression-related and identity-related features interacting during face processing tasks (Sections 5.5.1 and 5.5.2). In addition, I showed that a correct interpretation of sensory-motor information exhibited by face stimuli is sufficient to acquire face recognition capabilities (Section 6.2.2).

Addressed Research Gaps

In Section 2.4, I reported some crucial open questions from face processing research (Calder, 2011; Fisher et al., 2016):

- How dynamic and invariant features of the face are represented and they interact?
- How the interaction between facial expression and identity recognition is affected by experience?
- Where in the face processing hierarchy representations of invariant and dynamic facial features interact?

In this dissertation, I proposed that facial identity discrimination can be achieved by mean of a twofold face-space coding both invariant and dynamic perceptual features of the face in a single multidimensional representation. These representations can be acquired by observing novel face stimuli and interpreting their exhibited motor configuration provided by embodied mechanisms described in Chapter 4. I demonstrated that:

- Dynamic features of the face can be encoded in bodily formatted representations possibly during very early stage of visuo-motor processing mediated by embodied mechanisms, as explained in Chapter 4;
- Identity discrimination is mediated by a face-space representation, which can be acquired via the motor information provided by embodied mechanisms through experience (Chapters 5 and 6);

Therefore, I concluded that the development of a psychological face-space, and with it the acquisition of face processing skills, does not lead to two distinct

representations describing dynamic and invariant facial features. Instead, both invariant and dynamic facial features are coded by a single face-space model enabling both facial expression and identity recognition capabilities interacting with each other. In addition, in Section 7.3, I will propose in which brain areas these representations may interact.

Significance

In this section, I illustrated that this dissertation addressed the proposed primary contribution. This contribution is of significant value for at least the following reasons:

- The face-space framework described in this dissertation can be used to assist face processing research and to generate new hypotheses;
- The offered hypotheses and computational theories can be easily assessed by face processing experiments with human subjects, thus allowing falsifiability of the proposed hypotheses;
- The new standpoint suggesting inter-dependencies between facial identity and expression processing offers theories and methodologies for facial identity and expression recognition communities in machine learning to work jointly for the development of machine learning algorithms exhibiting higher recognition rates.

7.2.3 Secondary Contribution

The proposed secondary contribution of this dissertation was:

> *Providing computational evidence supporting embodiment of social cognition and introducing novel hypotheses explaining how this embodiment can potentially affect face processing capabilities.*

In this dissertation, I addressed this secondary contribution. In fact, in Chapter 3, I suggested the possible links between embodied mechanisms and face processing capabilities (and consequently social cognition, see Section 2.1.6), whereas in Section 6.3 I provided an argument proposing that embodied mechanisms, as discussed in Chapter 4, can plausibly shape important aspects of social cognition. I supported my argument with data available from computational simulations

(Section 6.3.2). In addition, the provided computational evidence offered preliminary insights on the possible effects that dysfunctions in embodiment can have on face processing capabilities. This suggested compelling similarities with findings in clinical populations affected by social disorders, as reviewed in Chapter 3. In summary:

- I provided a computational account explaining embodied simulation mechanisms (Chapter 4) and a computational model explaining phenomena observed in face processing studies (Chapter 5), showing their links in Chapter 6;

- I introduced an argument suggesting a constitutive role of embodied simulation for social cognition (Section 6.3) and I validated it through computational evidence (Section 6.3.2), thus adding computational evidence in favour of embodied cognition theories;

- I introduced novel hypotheses on how dysfunctions of embodied mechanisms can plausibly affect face processing capabilities (Section 3.6) and provided preliminary computational evidence explaining these interactions in Section 6.3.2.

Addressed Research Gaps

Current computational models of embodied simulation limit their investigation to two main topics: motor control and mind-reading (Oztop et al., 2006). To the best of my knowledge, there are no computational models able to explain the mechanisms of embodied simulation processes in a way that links them to other cognitive abilities usually suggested to not depend on bodily mechanisms, such as facial identity recognition. This is a vital contribution since it offers a first computational explanation of the reuse hypothesis (Gallese and Caruana, 2016), namely how embodied simulation mechanisms can be reused by other cognitive capabilities (refer to Definition 2.26 on page 59).

In addition, the current literature does not offer computational evidence in favour of the constitution hypothesis suggested by embodied cognition research. This kind of evidence is particularly important for embodied cognition research, since the available behavioural evidence supporting this hypothesis is not yet strong enough to validate a constitutive role of the body in cognition (Shapiro, 2010). On the contrary, with the computational tools provided in this dissertation

it is possible to investigate the interactions and more easily demonstrate that the role of embodied mechanisms on cognition is constitutive and not merely causal.

Significance

The present work met the proposed secondary contribution. Hence, this contribution is of significant value for at least the following reasons:

- It promotes a closer collaboration between cognitive science research and embodied cognition research since the work offered meaningful connections harmonising both the two considered research communities;
- The proposed theories can be tested under different methodologies, thus allowing their falsifiability;
- It advances new hypotheses of mirroring mechanisms and their interactions with face processing capabilities. These hypotheses are of extreme value for fostering both cognitive science and embodied cognition research communities;
- It offers new insights on how dysfunctions of embodied mechanisms can impair face processing capabilities in individuals affected by social disorders.

7.3 Integration of the Findings with Face Studies

In this section I will discuss how the presented findings can relate to literature and models in cognitive studies. In Section 2.2.3 I reported the study by Meltzoff and Moore (1983) showing that human infants of about 40 minutes old are already able to mimic simple facial expressions, such as mouth opening and tongue protrusion. Importantly, it has been shown that this early behaviour cannot be explained as a reflex matching the observed action with the enacted one, but it encompasses a broader psychological framework (Meltzoff and Moore, 1992). Meltzoff and Moore (1992) suggested that one of the psychological functions that early imitation serves is to identify people. In this chapter, I showed how face recognition can indeed be acquired using nonverbal behaviour (*i.e.* motor interpretation of facial expressions), as suggested by the authors. Nevertheless, Meltzoff and Moore (1992) suggested that the infant may attribute identities by associating particular behaviours (*e.g.* smile or frown) to specific individuals. Here, instead, I argue that identification skills can be acquired by interpreting the

observed motor configuration (via imitation or mirroring). By doing so, the infant can access sufficient information to develop a face-space enabling face recognition. This face-space representation can be continuously adjusted through experience by observing new face stimuli.

In the literature I also mentioned the work by Leo et al. (2018) showing how presenting dynamic face stimuli can significantly contribute to the acquisition of face recognition capabilities in infants. Importantly, the same pictorial information used for the experimental condition using dynamic stimuli was presented statically frame by frame (*i.e.* preventing a fluid motion) during another experimental condition (*i.e.* multistatic condition). These new multistatic stimuli were not sufficient to facilitate face recognition skills in the newborns. In addition, the study by Turati et al. (2011) compared emotional face expressions with other visual non-emotional motions of the face, such as speech motions[4]. The authors demonstrated that face recognition skill is enhanced in infants observing the dynamic emotional face stimuli, as compared to infants presented with dynamic neutral face stimuli. Thus, it seems that *"in order to be effective, motion of face features should have an overall emotional value"* (Turati et al., 2011, page 315). These studies can help to integrate my findings with face processing studies, as I will explain soon.

Neuroscientific studies on face processing suggest that changeable aspects of the face (*i.e.* eye gaze, expression and lip movement) are processed in the Superior Temporal Sulcus (STS), whereas invariant aspects of the face necessary to classify the exhibited identity are processed in a distinct brain area, the Lateral Fusiform Gyrus (LFG) (Haxby et al., 2000). The STS presents neural connections with the amygdala and other brain areas usually associated with emotional processing capabilities (Adolphs, 2002) and interactions were observed between the STS and the LFG (Haxby et al., 2000). Recent studies propose that the STS is also associated with areas dedicated to mirroring mechanisms and imitative capabilities (Buxbaum et al., 2014) and Molenberghs et al. (2010) provided evidence suggesting that the role of the STS in imitation is not only to passively register observed biological motion, but rather to actively represent sensory-motor correspondences between one's actions and the actions of others. In addition, Schultz et al. (2013) found that the activation of the STS is modulated by the fluidity of presented facial movements. When the presentation of facial movements differs from biologically plausible motions, STS neural activation is affected and face recognition impaired

[4]In the study the stimuli did not include audio cues.

(Redcay, 2008; Xiao et al., 2014). The perception of biological motion modulates a broad network of brain regions including the STS (Vaina et al., 2001). These inputs are integrated by the STS to extract their social significance (Redcay, 2008).

Indeed, the STS is not only modulated by visual information, but also by audio information providing social information (Redcay, 2008). Therefore, Redcay (2008) suggests that the STS has a more general role of providing interpretations of dynamic social stimuli, including facial expressions. Thus, the STS, assisted by putative emotional brain areas like the amygdala (Adolphs, 2002), can provide information necessary to interpret the social meaning of the observed motor information (*e.g.* emotional expression of the observed face stimuli), as suggested in this dissertation with the model presented in Chapter 4. Here it is important to note that I did not include temporal dynamic information in my model. However, the theory underlying my model was recently extended to temporal dynamics by Boccignone et al. (2018), thus demonstrating the plausibility of the present argument. The interpretation of the observed facial motion provided by the STS can then be used by the LFG to acquire face recognition capabilities, as per the the psychological face-space discussed in Chapters 5 and 6.

Finally, as reported in Section 2.2, traditional studies in face processing propose that identity and expression processing rely on distinct brain pathways (Bruce and Young, 1986; Bruyer et al., 1983), whereas modern literature in face studies demonstrate the presence of interactions between invariant and dynamic features of face stimuli (Becker et al., 2007; Calder et al., 2001; Rhodes and Jeffery, 2006; Rhodes and Leopold, 2011). The models and theory proposed in this dissertation well integrate with both these understandings. In fact, the model provided in Chapter 4 can assist in extracting the social meaning of face stimuli, similarly to what suggested previously for the STS. In addition, the face-space provided in Chapter 5, similarly to the LFG, enables face recognition capabilities, which can be acquired by using the interpretations of the observed face stimuli extracted by the embodied mechanisms presented in this dissertation, as suggested in Chapter 6.

Therefore, the explanation proposed in this thesis can support traditional models of face processing, since suggesting a degree of separation between processes extracting dynamic information of the face (*i.e.* the model presented in Chapter 4 resembling functions of the STS and mirror neuron system) and the ones extracting invariant information of the face (*i.e.* the face-space in Chapter 5 resembling functions of the LFG). In addition, this interpretation can also support modern understandings of face processing, since the representations provided by

the face-space show interactions between identity-related and expression-related components. I argue that it is plausibly at the LFG level where invariant and dynamic codings of the face interact during face processing tasks, but that it is at the STS and mirror neuron system level where interpretations of facial emotional content may affect the acquisition of face recognition capabilities. In fact, dysfunctional embodied mechanisms at the STS and mirror neuron system level will lead to impaired interpretations, and consequently damage the acquisition of correct face recognition skills, as demonstrated by the simulations presented in Section 6.3.

7.4 Future Work

This work is the result of a doctoral research degree, necessarily constrained by resources and time. Although I largely demonstrated the validity of the proposed thesis argument and the plausibility of secondary hypotheses through appropriate computational evidence, the offered computational theories are yet to be definitive. In fact, it is necessary to test them through appropriate human experiments.

In Section 7.1, I provided two examples of studies that can be implemented with human subjects to validate my hypothesis. Thus, a first proposed future work is to create collaborations with cognitive scientists investigating face processing mechanisms in order to design human experiments able to validate (or reject) the thesis offered by this dissertation. The collaboration can also be extended to neuroscientists, in particular, researchers investigating motor, imitative and mirroring functions in the brain. The additional available evidence would contribute to providing better insights in both face processing and embodied cognition research.

The second aimed future work is to apply the insights offered by this dissertation to develop computational models having better facial identity and expression recognition performance. In particular, deep learning is achieving unparalleled success in image recognition tasks (LeCun et al., 2015; Sun et al., 2014; Taigman et al., 2014) and the model and theory presented in Chapter 4 was successfully extended to deep learning techniques (Boccignone et al., 2018). These computational models can be conveniently linked to early processing of human visual cortex (Boccignone et al., 2018) and they are, therefore, suitable candidates for potential future computational simulations.

I believe that the proposed face-space model can be adapted to a deep network too. In this dissertation I used the pixels intensities of static images as input to my

models. However, the input to the model can be any vector of features extracted by the observed face stimuli able to preserve its perceptual information using a representation with lower dimension. Therefore, a viable non-linear alternative of my face-space model can be obtained by pre-processing the input face stimuli by using a deep neural network model trained to preserve invariant and dynamic features of the face in a more compressed and smart representation (Le et al., 2013), instead of the used linear PCA. Then, it will be necessary to use the objective function presented in Chapter 6 to develop a single multi-dimensional representation facilitating both facial identity and expression recognition. Importantly, by doing this it will be possible to develop a single model for facial identity and expression recognition by training it with labels of facial expressions only. Alternatively, it may be possible to use a similar approach to derive a model promoting facial expression and identity recognition by training it with labels of facial identities only. The latter strategy is indeed more interesting, since it reduces the complexity of labelling face stimuli with the exhibited facial configuration, which is a strenuous task.

Also, the models presented in this dissertation can be extended to include temporal information. Boccignone et al. (2018) recently provided an extension of the model presented in Chapter 4 successfully including temporal aspects of the observed face stimuli. In addition, temporal dynamics can be included in the proposed face-space model by extracting features from a set of consecutive stimuli instead of extracting them from a single static image. An alternative can be using other techniques to improve temporal coherence in deep learning models (Mobahi et al., 2009), thus being able to extract more convenient features to use as input for the objective function in Chapter 6.

Finally, social robotics research can highly benefit from my discussed theories. In particular, by using the insights gathered by this study it may be possible to integrate embodied mechanisms in cognitive architectures developed for robotic platforms to extend their use to not only select the action to execute, but also to observe the action of the interaction partner and determining, with the very same architecture, the more plausible mental state to attribute to the interaction partner (Boccignone et al., 2018). By doing so, it would be possible to design socially intelligent robots, which is an important feature to support safe human-robot interactions (Vitale et al., 2014; Williams, 2012).

Chapter Bibliography

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12(2):169–177.

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K., and Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, 92(2):179.

Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., and Lanzarotti, R. (2018). Deep construction of an affective latent space via multimodal enactment. *IEEE Transactions on Cognitive and Developmental Systems*.

Bruce, V. and Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77:305–327.

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., and Rectem, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. *Brain and Cognition*, 2(3):257–284.

Buxbaum, L. J., Shapiro, A. D., and Coslett, H. B. (2014). Critical brain regions for tool-related and imitative actions: A componential analysis. *Brain*.

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

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., and Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, 41(9):1179–1208.

Cangelosi, A. and Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. *Cognitive Science*, 30(4):673–689.

Farah, M. J. (1995). Dissociable systems for visual recognition: A cognitive neuropsychology approach. In Kosslyn, S. M. and Osherson, D. N., editors, *Visual Cognition: An Invitation to Cognitive Science*, volume 2, pages 101–119. MIT Press Cambridge.

Fisher, K., Towler, J., and Eimer, M. (2016). Facial identity and facial expression are initially integrated at visual perceptual stages of face processing. *Neuropsychologia*, 80:115–125.

Gallese, V. and Caruana, F. (2016). Embodied simulation: Beyond the expression/experience dualism of emotions. *Trends in Cognitive Sciences*.

Ganel, T. and Goshen-Gottstein, Y. (2004). Effects of familiarity on the perceptual integrality of the identity and expression of faces: The parallel-route hypothesis revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3):583.

Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., and Goodale, M. A. (2005). The involvement of the "fusiform face area" in processing facial expression. *Neuropsychologia*, 43(11):1645–1654.

Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological Bulletin*, 141(6):1266.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6):223–233.

Kadosh, K. C., Luo, Q., de Burca, C., Sokunbi, M. O., Feng, J., Linden, D. E., and Lau, J. Y. (2016). Using real-time fMRI to influence effective connectivity in the developing emotion regulation network. *NeuroImage*, 125:616–626.

Le, Q. V., Ranzato, M., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J., and Ng, A. Y. (2013). Building high-level features using large scale unsupervised learning. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8595–8598. IEEE.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Leo, I., Angeli, V., Lunghi, M., Dalla Barba, B., and Simion, F. (2018). Newborns' face recognition: The role of facial movement. *Infancy*, 23(1):45–60.

Leopold, D. A., Rhodes, G., Müller, K.-M., and Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1566):897–904.

Meltzoff, A. N. and Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, pages 702–709.

Meltzoff, A. N. and Moore, M. K. (1992). Early imitation within a functional framework: The importance of person identity, movement, and development. *Infant Behavior and Development*, 15(4):479–505.

Mobahi, H., Collobert, R., and Weston, J. (2009). Deep learning from temporal coherence in video. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 737–744. ACM.

Molenberghs, P., Brander, C., Mattingley, J. B., and Cunnington, R. (2010). The role of the superior temporal sulcus and the mirror neuron system in imitation. *Human Brain Mapping*, 31(9):1316–1326.

Oztop, E., Kawato, M., and Arbib, M. (2006). Mirror neurons and imitation: A computationally guided review. *Neural Networks*, 19(3):254–271.

Pell, P. J. and Richards, A. (2013). Overlapping facial expression representations are identity-dependent. *Vision Research*, 79:1–7.

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neuroscience & Biobehavioral Reviews*, 32(1):123–142.

Rhodes, G. and Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, 46(18):2977–2987.

Rhodes, G. and Leopold, D. A. (2011). Adaptive norm-based coding of face identity. In Calder, A. J., editor, *The Oxford Handbook of Face Perception*, chapter 14, pages 263–286. Oxford University Press.

Rhodes, G., Pond, S., Burton, N., Kloth, N., Jeffery, L., Bell, J., Ewing, L., Calder, A. J., and Palermo, R. (2015). How distinct is the coding of face identity and expression? Evidence for some common dimensions in face space. *Cognition*, 142:123–137.

Rhodes, G., Robbins, R., Jaquet, E., McKone, E., Jeffery, L., and Clifford, C. W. (2005). Adaptation and face perception: How aftereffects implicate norm-based coding of faces. In Clifford, C. W. and Rhodes, G., editors, *Fitting the Mind to the World: Adaptation and After-effects in High-level Vision*, volume 2, chapter 8, pages 213–240. Oxford University Press.

Schultz, J., Brockhaus, M., Bülthoff, H. H., and Pilz, K. S. (2013). What the human brain likes about facial motion. *Cerebral Cortex*, 23(5):1167–1178.

Shapiro, L. (2010). *Embodied Cognition*. Routledge.

Sun, Y., Wang, X., and Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1891–1898.

Taigman, Y., Yang, M., Ranzato, M., and 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*, pages 1701–1708.

Turati, C., Montirosso, R., Brenna, V., Ferrara, V., and Borgatti, R. (2011). A smile enhances 3-month-olds' recognition of an individual face. *Infancy*, 16(3):306–317.

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., and Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences*, 98(20):11656–11661.

Vitale, J., Williams, M.-A., and Johnston, B. (2014). Socially impaired robots: Human social disorders and robots' socio-emotional intelligence. In *6th International Conference on Social Robotics*, pages 350–359.

Williams, M.-A. (2012). Robot social intelligence. In *Social Robotics*, pages 45–55. Springer.

Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y.-H. P., and Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology*, 5.