Social neuroscience of emotion perception and regulation : the influence of context and individual differences

Vrticka, Pascal

How to cite

VRTICKA, Pascal. Social neuroscience of emotion perception and regulation : the influence of context and individual differences. Doctoral Thesis, 2009. doi: 10.13097/archive-ouverte/unige:5169

This publication URL: https://archive-ouverte.unige.ch/unige:5169
Publication DOI: 10.13097/archive-ouverte/unige:5169

UNIVERSITÉ
DE GENÈVE
SCIENCES

DOCTORAT EN NEUROSCIENCES
des Universités de Genève
et de Lausanne

UNIVERSITE DE GENEVE FACULTE DES SCIENCES

Professeur Patrik Vuilleumier, directeur de thèse
Professeur David Sander, co-directeur de thèse

TITRE DE LA THESE

SOCIAL NEUROSCIENCE OF EMOTION PERCEPTION AND REGULATION

THE INFLUENCE OF CONTEXT AND INDIVIDUAL DIFFERENCES

THESE
Présentée à la
Faculté des Sciences

de l'Université de Genève

pour obtenir le grade de
Docteur en Neurosciences

par

Pascal VRTIČKA

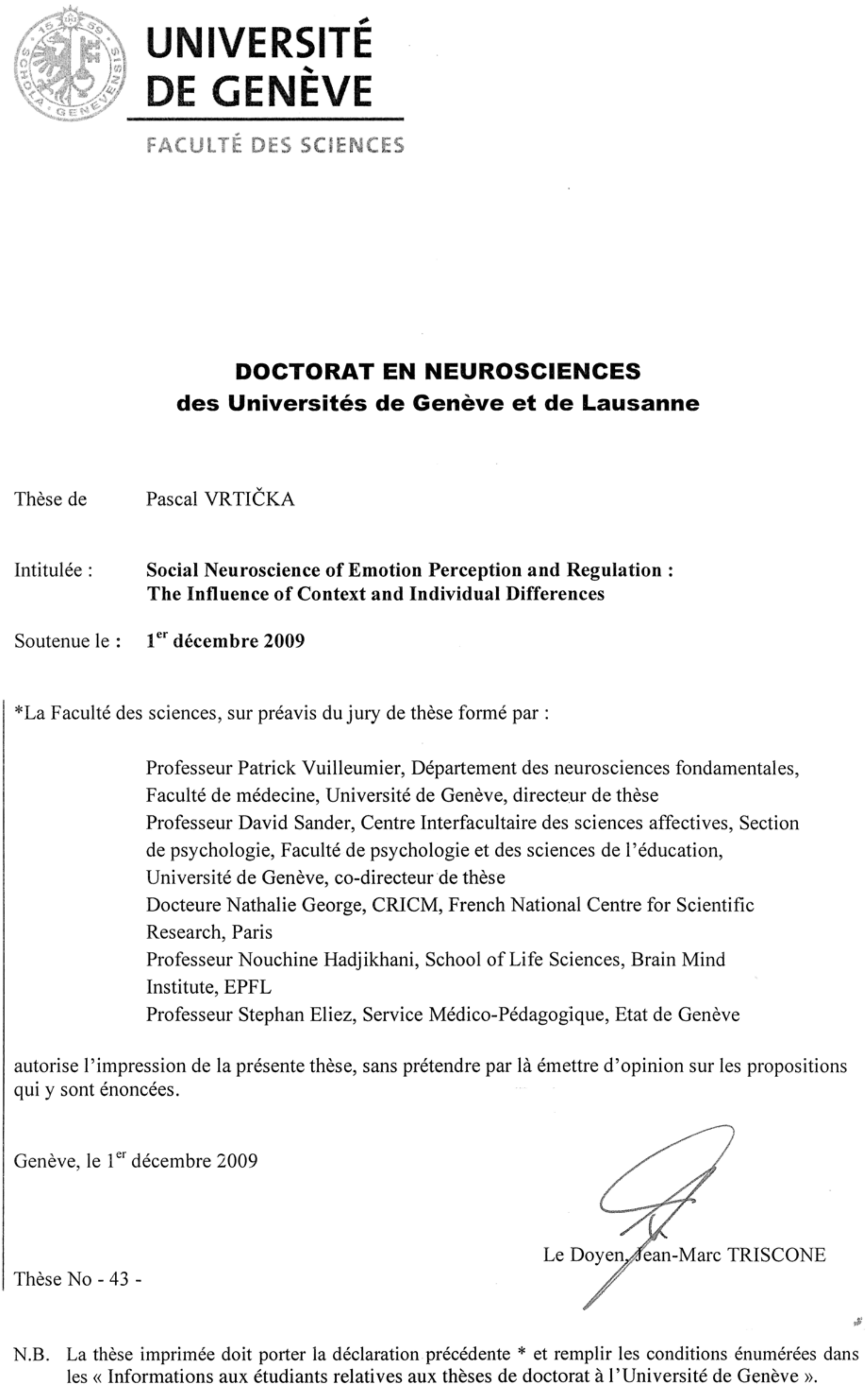
de Luzern (LU)

Thèse N° 43

Genève

Editeur ou imprimeur : Université de Genève

2010



UNIVERSITÉ
DE GENÈVE
FACULTÉ DES SCIENCES

DOCTORAT EN NEUROSCIENCES
des Universités de Genève et de Lausanne

Thèse de Pascal VRTIČKA

Intitulée : **Social Neuroscience of Emotion Perception and Regulation : The Influence of Context and Individual Differences**

Soutenue le : **1^er^ décembre 2009**

*La Faculté des sciences, sur préavis du jury de thèse formé par :

Professeur Patrick Vuilleumier, Département des neurosciences fondamentales, Faculté de médecine, Université de Genève, directeur de thèse
Professeur David Sander, Centre Interfacultaire des sciences affectives, Section de psychologie, Faculté de psychologie et des sciences de l'éducation, Université de Genève, co-directeur de thèse
Docteure Nathalie George, CRICM, French National Centre for Scientific Research, Paris
Professeur Nouchine Hadjikhani, School of Life Sciences, Brain Mind Institute, EPFL
Professeur Stephan Eliez, Service Médico-Pédagogique, Etat de Genève

autorise l'impression de la présente thèse, sans prétendre par là émettre d'opinion sur les propositions qui y sont énoncées.

Genève, le 1er décembre 2009

Le Doyen, Jean-Marc TRISCONE

Thèse No - 43 -

N.B. La thèse imprimée doit porter la déclaration précédente * et remplir les conditions énumérées dans les « Informations aux étudiants relatives aux thèses de doctorat à l'Université de Genève ».

« Where the heart lies, let the brain lie also. »

Robert Browning (1812 -1889)

CONTENTS

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER VI

CHAPTER VII

ACKNOWLEDGMENT

I would like to thank Prof. *Patrik Vuilleumier*, my thesis director, for his willingness to take charge of me as his PhD student during the last four years, despite the fact that my Bachelor and Masters background was tremendously different from what were and still are the current research topics at the *Laboratory for Neurology and Imaging of Cognition (LabNIC)*. I am also deeply grateful for his very professional, yet always proximal guidance regarding fMRI methods particularly at the beginning, and concerning the write-up and submission of articles towards the end of my studies.

I am greatly indebted to Prof. *David Sander*, my thesis co-director, who not only importantly contributed towards the development of the first experimental paradigm giving rise to two already published articles, but also to his many very constructive comments and helpful advice regarding the more psychological aspects of my doctoral thesis at the *Swiss Center for Affective Sciences (CISA)*.

My sincere thanks go to several people as collaborators or in the administrative board of the abovementioned two institutes, who importantly contributed to the successful completion of the present thesis in many different ways. This includes Prof. *Klaus Scherer*, *Daniela Sauge*, *Audrey Souchard*, and *Carole Varone* from the CISA, as well as Dr. *Sophie Schwartz* from the LabNIC.

I would also like to thank all the thesis committee members who agreed to participate as external experts at the last but very important steps of my postdoctoral qualification. These are Prof. *Stephan Eliez*, Dr. *Nathalie George*, and Prof. *Nouchine Hadjikhani*.

I am particularly grateful to Dr. *Frédéric Andersson*, who took care of me as a postdoc at the very beginning of my PhD thesis and taught me everything about fMRI, from scanning over data transfer to SPM analysis.

The completion of my doctoral thesis would not have been possible without the help of the many other former and actual fellow collaborators of the LabNIC and CISA, whom I would like to thank here as well, particularly Dr. *Sylvain Delplanque* for the help with the French Resume.

Finally, I would like to thank my family and family in law, my wife and baby boy for their unconditional support during the last four years (or a little less...), both at times of success but also failure. I hope that, with the successful completion of my PhD thesis, it will become evident that the many necessary sacrifices were worth the trouble.

ABSTRACT

The present doctoral thesis investigated the neural correlates of two core processes in social neuroscience, namely emotion perception and regulation, and their modulation by external (social context) as well as internal (individual differences – and in particular adult attachment style) factors.

The results of the performed experiments revealed for the first time that the social (or non-social) nature – and thereby conveyed specific meaning – of presented stimuli not only differentially modulated brain activity during emotion perception (fMRI study N° 1a) and subsequent recognition (fMRI study N° 1b) of facial expressions, as well as during natural emotion experience induced by emotional scenes (fMRI study N° 2a), but also differentially influenced brain activation during emotion regulation by either cognitive re-appraisal or behavioral suppression of emotion expression (fMRI study N° 2a). Moreover, all of these relations (except for emotion recognition) were crucially dependent on individual differences in adult attachment style (fMRI studies N° 1a and 2b), disclosing so far unknown direct interactions between personality, social context, and emotion regulation strategy use, and corresponding underlying brain activity. Finally, adult attachment style was also found to influence brain activation during self- versus other-perception (fMRI study N° 3), thereby providing new clues on the neural correlates of self- and other models related to individual differences.

Our new data therefore not only significantly extend the already available literature in each research domain of interest – emotion perception, emotion regulation, adult attachment style –, but also provide important clues on the inter-dependence of the latter fields. This has decisive implications on future emotion research by suggesting that a more comprehensive experimental approach is needed to further elucidate the neural mechanisms of social perception. Even though exactly this claim has initially lead to the emergence of the research field of social neuroscience, more efforts are still necessary to properly implement it into the paradigms applied to study human social functioning.

RESUME

La recherche concernant les neurosciences dans le domaine des sciences sociales a pour objectif d'élucider les mécanismes biologiques qui sous-tendent les comportements sociaux humains. Elle intègre des concepts issus à la fois de la biologie, de la neurophysiologie et de la psychologie. En s'inscrivant dans cette démarche, les expériences menées dans le cadre de cette thèse utilisent autant des méthodes de neuro-anatomie fonctionnelle, comme l'imagerie par résonnance magnétique fonctionnelle (IRMf), que des apports de la psychologie sociale, prenant en compte les influences du contexte social et les différences individuelles. L'ambition est de révéler des relations par trop méconnues entre la personnalité, le contexte, le comportement, l'émotion et la cognition.

Dans le cadre de cette thèse, les recherches se focalisent sur deux processus fondamentaux des neurosciences sociales : la perception et la régulation émotionnelle. Leur intérêt porte sur les diverses modulations de l'activité cérébrale associées aux deux processus susmentionnés lors de manipulations du contexte social d'une part et en fonction des différences individuelles d'autre part.

Dans un premier temps, nous partons du contexte théorique associé à la perception des émotions sociales, en décrivant, et en évaluant (appraisal) les différents mécanismes par lesquels les émotions sociales sont hypothétiquement déclenchées (imitation, empathie, attitudes / préjudice, connexion / rejet social) et comment elles peuvent influencer la cognition sociale (théorie de l'esprit, prise de décision sociale). Nous reprenons ensuite la théorie de la régulation émotionnelle principalement basée sur le modèle de développé par James J. Gross (voir Figure M4) qui différencie la ré-évaluation cognitive et l'inhibition comportementale de l'expression (ou suppression) émotionnelle.

Puis, après avoir introduit brièvement quelques aspects du développement du cerveau social d'un point de vue phylogénétique et les processus neuro-anatomiques fonctionnels mis en exergue par les neurosciences sociales, nous abordons le rôle potentiel du contexte dans la perception sociale. Cette partie inclut une argumentation sur les mécanismes sous-tendant, i) la modulation directe d'une stimulation émotionnelle (une expression faciale) par une information verbale concomitante qui concerne la signification sociale de l'expression, ii) l'intentionnalité, iii) la présence (ou l'absence) d'une audience et, finalement, iv) la présence ou absence de contenu social basique de la stimulation. La fin de l'introduction reprend quelques exemple des concepts psychologiques utilisés pour étudier l'influence des différences individuelles sur l'activité cérébrale associée à la cognition sociale, tels que l'anxiété, le névrotisme, l'intro-/ extraversion et la

sensibilité au rejet. Toutefois certains éléments laissent supposer que ces derniers modèles psychologiques ne sont probablement pas parfaitement appropriés pour expliquer le rôle des traits de la personnalité sur l'activité cérébrale induite par des stimuli émotionnels sociaux. Il est alors proposé d'appliquer, dans le cadre de cette thèse, des styles d'attachement (sécure, évitant et anxieux) comme concepts psychologiques.

La section suivante de la thèse présente les résultats expérimentaux obtenus et ouvre la discussion dans le cadre des concepts psychologiques susmentionnés.

Lors de l'étude d'IRMf N° 1ab, la signification sociale de visages heureux (souriant) ou en colère a été manipulée par un jeu qui rendait personnellement pertinents ces visages, certains devenant amis et d'autres ennemis. Les conséquences des interactions entre l'expression du visage et son statut social sur la perception émotionnelle ont été étudiées lors de la phase d'encodage et lors de la reconnaissance. Enfin, le style d'attachement a été introduit comme une variable additionnelle modulatrice.

Les résultats montrent que les visages émotionnels ne sont pas codés comme tels dans le cerveau humain, mais en fonction de leur statut social : l'activité dans le striatum ventral augmente lors de la récompense sociale, celle de l'amygdale lors de la punition sociale (menace / peur), et celle du sillon temporal supérieur lors de l'incongruité en général. De plus, l'activation dans les deux premières régions susmentionnées est modulée de manière importante par le style d'attachement des individus : une diminution de l'activité liée à la récompense est observée en fonction de l'évitement, et une augmentation de l'activité liée à la peur / menace est observée en fonction de l'anxiété.

Dans une étude comportementale initialement prévue pour sélectionner les stimuli de l'étude IRMf subséquente N° 2ab, nous avons utilisé des scènes positives et négatives, sociales ou non-sociales, qui ont été jugées par les participants sur les dimensions d'agréabilité, d'activation et de contrôle. Nos données révèlent quelques effets liés à la perception des stimuli per se. Mais surtout, ils confirment les résultats obtenus lors de l'étude d'IRMf N°1ab en montrant i) une corrélation négative spécifique entre l'agréabilité des stimuli sociaux positifs et l'évitement, ii) ainsi qu'une augmentation d'activation et une diminution des jugements de contrôle des images sociales négatives en fonction de l'anxiété.

Lors de l'étude d'IRMf N° 2ab, des scènes positives ou négatives, sociales ou non-sociales, ont été présentées aux participants dans des conditions de visionnage normal, ou avec l'instruction d’utiliser une stratégie de ré-évaluation ou de suppression, afin de diminuer l’impact émotionnel subjectif des images.

Les résultats montrent que, i) des activités préfrontales différentes sont mises en place pour les stratégies de ré-évaluation ou de suppression et ceci indépendamment de l’agréabilité ou du contenu social des images, ii) différentes régions corticales et sous-corticales sont activées en fonction de la valence et du contenu social des scènes visuelles, iii) la ré-évaluation est plus efficace pour réduire l’activité dans différentes régions cérébrales associées à l’aspect social des images, alors que la suppression diminue de manière plus marquée les changements d’activités observés lors de la comparaison des images positives et négatives. De plus, iv) les activités des amygdales sont latéralisées, l’amygdale gauche répondant préférentiellement à la ré-évaluation, la droite à la suppression. Enfin, v) le style d’attachement d’évitement module l’activité cérébrale dans toutes les conditions de visionnage, particulièrement en augmentant le conflit cognitif et émotionnel en présence de scènes sociales négatives si aucune stratégie de régulation ne peut être mise en place, et en maintenant une telle activité durant la ré-évaluation, ce qui signifie un manque d’efficacité de cette stratégie cognitive de régulation émotionnelle.

Enfin, lors de l’étude d’IRMf N°3, le rôle modulatoire du style d’attachement sur la perception de soi et des autres a été étudié au travers de l’attribution de traits (grâce à l’utilisation d’adjectifs) à soi-même ou à des amis proches. Cette manipulation était justifiée par l’hypothèse selon laquelle le style d’attachement d’évitement entraîne une vision négative des autres (et probablement une vision plus positive de soi), alors que le style d’attachement anxieux est lié à une vision négative de soi (et probablement à une vision positive des autres).

Les résultats montrent que l’attachement évitant entraine un biais attribuant plus de traits positifs à soi-même et implique plus d’activité cérébrale pour cette condition spécifique dans le sillon temporal supérieur, le fusiforme, et dans le cortex préfrontal ventro-médial. Par contre, l’attachement anxieux est caractérisé, en général, par plus d’activité cérébrale pour l’attribution des traits à autrui, et pour l’évaluation des traits négatifs des autres et de soi-même en particulier.

L’ensemble des résultats empiriques présentés dans cette thèse prouve de manière innovante l’existence d’importantes nouvelles relations entre les mécanismes cérébraux, tels qu’étudiés par les neurosciences sociales, incluant la perception des émotions et leur régulation, et des facteurs modulateurs comme le contexte et les différences individuelles. Ces résultats constituent un apport significatif à la littérature existante, en démontrant que les traitements de l’information émotionnelle ne sont pas circonscrits aux caractéristiques intrinsèques du stimulus, mais dépendent aussi de sa signification sociale et des prédispositions individuelles à interpréter ce stimulus d’une manière particulière.

GUIDELINES

The doctoral thesis manuscript at hand is divided into four main sections.

Chapters I – IV: Introduction

The aim of the introduction is to provide the reader with the necessary background information for a smooth reading of the subsequent chapters devoted to experimental work and to the general discussion of the obtained results.

Chapter I starts with shortly describing the origins of social neuroscience by mentioning two psychological accounts that marked this research domain. Subsequently, the main part of Chapter I supplies a detailed explanation of two core processes in social neuroscience, namely emotion perception and regulation. This information is crucial, because exactly the latter two processes were extensively investigated during the experimental work included in the present doctoral thesis.
Chapter I: Origins and Core Processes.

Chapter II offers insight into the refinement of social perception during the course of evolution and its connection to the disproportionate increase in human brain size, particularly the neo-cortex. Even though not central for the working hypotheses and experimental work of the present doctoral thesis, Chapter II nonetheless contains valuable information on the social brain and the development of its unique properties in humans.
Chapter II: An Evolutionary Perspective.

Chapter III lists the most important human brain regions involved in social perception from a functional neuro-anatomical perspective. This is to provide the reader with the necessary knowledge about the "*social brain*" for a straightforward understanding of any anatomical accounts mentioned in the experimental results and general discussion sections.
Chapter III: Functional Neuro-Anatomy.

Chapter IV provides specific information about the influence of social context and individual differences on social perception, the two variables manipulated in the experimental work within the scope of the present doctoral thesis. With the help of literature examples already known connections are illustrated, their limitations are pointed out, and the reasons for the choice of the particular manipulations that were employed for the experimental work are explained.
Chapter IV: The Role of Social Context and Individual Differences.

Chapter V: Main Questions and Aims

After having explained the theoretical background in Chapters I - IV, Chapter V focuses on the specific research questions asked of each fMRI and behavioral experiment in the context of this doctoral thesis. To do so, methods and results of similar previous research are mentioned in more detail to emphasize the novel aspects addressed in the current paradigms.

Chapter VI: fMRI and Behavioral Studies

Chapter VI comprises the reports of the experiments conducted in the context of this doctoral thesis and their respective results. It includes two already published papers (fMRI studies N° 1a & b), three papers in preparation (fMRI studies N° 2a & b and 3), and one unpublished manuscript (behavioral study).

Chapter VII: General Discussion

Because the reports of the experimental work (Chapter VI) only contain discussions of specific research questions, the general discussion aims at synthesizing all these individual contributions and to analyze them from a more global perspective. This includes the explication of common and distinct brain activation patterns, limitations, and future perspectives.

Information concerning Figures and References

To avoid confusion between figures depicted in the thesis manuscript and figures as integral parts of the articles included in Chapter VI, the following numbering will be employed. Pictures as part of the manuscript will contain the letter “M” (e.g. Figure M7), whereas figures integrated in the articles will only be referred as to by numbers (e.g. Figure 3ab).

References mentioned in the main thesis manuscript are listed at the end in the separate Reference section. References as integral parts of the articles (Chapter VI) are listed at the end of each article separately.

CHAPTER I

Origins and Core Processes

I.1. Emergence of a new Research Field

The new research field of *social neuroscience* aims at elucidating the biological mechanisms that underlie social processes and behavior by integrating biological, neuro-physiological as well as social psychological concepts.

Such an assumption that neuro-physiological processes and psychological phenomena are tightly linked has already been brought forward more than a century ago in the famous *"Principles of Psychology"* by William James [1]. He recognized that developmental, environmental and socio-cultural factors have an impact on neuro-physiological processes, which in turn underlie psychological phenomena, and that such influences could be studied as neuro-physiological transactions [2].

Until the end of the 20th century, the investigation of brain structures and functions associated with psychological events was limited to animal studies, post-mortem examinations, and observations of specific cases involving selective brain lesions in humans. A well-known example of the latter is the restricted damage to the ventro-medial frontal lobe of Phineas Gage by a railroad accident, causing emotional instability and inappropriate social behavior but no changes in intellectual functioning [3, 4]. Only after the introduction of new brain research techniques, and in particular functional magnetic resonance imaging (fMRI) in the 1990ies, a more direct investigation of the connection between social and emotional functions and neuro-cognitive systems became possible.

In the context of the emergence of this new research field combining psychology and cognitive neuroscience to investigate the neural correlates of human social behavior, the term *social neuroscience* first appeared in an article by John Cacioppo and Gary Berntson in 1992 [2]. Since then, the exact denomination has been refined several times, including variations like *social cognitive neuroscience* [5, 6] or even *social cognitive affective neuroscience* [7]. Apparent from such designations is the fact that social neuroscience is a highly interdisciplinary field, combining cognitive, emotional and motivational, as well as personal and social contextual mechanisms. Moreover, the chronological evolution of the different denominations for the field of social neuroscience to first include the term "*cognitive*" and only subsequently "*affective*" testifies for the consecutive use of distinct underlying conceptual constructs and psychological theoretical models used to define and study social processes.

I.2. Origins of Social Neuroscience

The most recent imaging experiments aimed at deriving neural models of human social fnunctioning were crucially influenced by thoughts and findings from different behavioral paradigms dating back as early as to the 1930ies. These originated from two rather distinct psychological domains, namely developmental and social psychology. To understand the latest theoretical social neuroscience account [6] used to explain the functional neuro-anatomy (Chapter III) and the role of context and individual differences in social perception (Chapter IV), as well as the experiments and results of the present thesis (Chapters V and following), the most important concepts of the latter two psychological domains are briefly outlined below.

I.2.1. Developmental Psychology

The first neural models of *social perception*[1] were significantly inspired by *developmental psychology* [8, 9]. The primary concern of this domain is the description of the psychological changes that occur in humans over the course of the life span. In this regard, the developmental finding that children only become able to represent the states of other minds at an age of about four years [10], and that such abilities are likely not to be acquired by people suffering from autism disorder [11] was central for the genesis of so-called *social cognition* accounts in the context of social neuroscience. Subsequent imaging studies were able to localize specific cortical brain substrates of theory of mind and related mentalizing processes (see Chapter III and [9, 12, 13]), which were later on extended and integrated in a more comprehensive social cognition theory framework. Accordingly, social cognition is defined as *"… the cognitive processes used to decode and encode the social world …"* [14], and is thus primarily concerned with deliberate, explicit social perception mechanisms.

Even though not directly related with developmental changes in human brain functioning over the life span, theories of social cognition also address the issue of brain evolution across different mammalian species. After the discovery of specific neurons in the infero-temporal cortex of monkeys responding selectively to faces [15, 16], a quest has started to identify processes and brain structures that evolved specifically to subserve social functions. Culminating in the social brain hypothesis [17], such accounts state that there are at least two uniquely human processes related to social cognition (theory of mind and the representation of triadic relations between two

---

[1] In the context of this doctoral thesis, the term *social perception* is referring to all processes involved in the neural representation of human social behavior. It comprises *social cognition* – including deliberate, explicit mechanisms investigated by developmental psychology – as well as *perceptual, motivational and emotional functions* – which are important automatic, implicit contributors to social perception, as emphasized by social psychology (see following pages of Chapter I for more details).

minds and an object [13]) which have fuelled human brain evolution, accounting for the disproportionate increase of particularly the neo-cortex size in the human species (see Chapter II).

I.2.2. Social Psychology

Only a few years after the rise of developmental psychology, *social psychology* started to have an impact on the views of social neuroscience researchers and on the neural models of human social behavior [5-7, 18]. Rather than being interested in the developmental changes over the lifespan within a single person – the main focus of developmental psychology (see Chapter I.2.1. above) –, social psychology is primarily concerned with processes that occur between two people or within groups. With the words of Allport [19] – one of the founding figures of personality psychology –, *"… social psychology studies how the actual, imagined, or implied presence of other people influences an individual's thoughts, feelings and behaviors …"*.

Accordingly, a principal contribution of social psychology to social neuroscience was the inclusion of *social context* as well as *social motives and goals*[2] as important modulatory factors into neural models of social perception [7]. This comes from a number of famous social psychology studies already conducted as early as in the 1930ies demonstrating that social context differentially affects behavior elicited by the same stimulus or event. For example, Levin [20, 21] has observed that common objects would evoke distinct reactions in different people depending upon the presence of others, and concluded that *"… behavior is a joint function of person and situation, and the situation is primarily social …"*. Similarly, Trope [22] showed that a given emotional facial expression can be perceived either as positive or negative, only depending on the social context.

In addition, another central notion in social psychology is the distinction between *automatic* versus *controlled* processes of social perception [23], forming *"… a substantial theoretical basis for cognitive psychology ... including categorization, memory, reasoning and decision making …"* [24]. Therefore, in contrast to developmental psychology mainly focusing on deliberate, explicit mechanisms related to social cognition, social psychology emphasizes the importance of automatic and very rapid inferences about other people as basic components of social perception. Because many such fast processes are mediated by emotional and/or motivational processes, *"... social neuroscience ... can never by solely cognitive but will inevitably entail emotion as well …"* [7]. Along these lines, social neuroscience inspired by social psychology attaches great importance to

---

[2] As proposed in the title of this doctoral thesis, the main focus of attention was directed towards *individual differences* as internal determinants of *social motives and goals*. However, because there are also external factors and social psychology theory refers to *social motives and goals* as a whole [7], this expression will be used within the first sections until further specified in Chapter IV.4.

the interplay between cognition and emotion / motivation as well as other more basic perceptual processes like familiar face recognition and the representation of perceived action (see Chapters I.3. and III).

Since the year 2000, an ever growing number of fMRI studies have investigated the role of social context as well as social motives and goals on brain responses related to social perception. The results and implications for further research in social neuroscience – also in the context of this doctoral thesis – are discussed in Chapter IV.

I.2.3. Contemporary View and Neuroscience Approach

A general consensus on the theoretical underpinnings of social neuroscience, as suggested in recent reviews by different authors [6, 18], is that it is important to employ well established concepts of social psychology theory to derive neural models of social perception. Consequently, the following description of the fundamental processes of social neuroscience (see Chapter I.3.) will be geared to such social psychology accounts, and particularly to the model put forward by Lieberman (2007).

I.3. Social Neuroscience of Emotion Perception and Regulation: Fundamental Mechanisms and Core Processes

Lieberman (2007) suggests dividing the fundamental mechanisms and core processes of social neuroscience into four social information processing units. These include: 1) *understanding others* and 2) *understanding oneself* (similarly to [14]), as well as 3) *controlling oneself* and 4) *processes that occur at the interface of self and others*. Carrying on with the social psychology tradition, he then differentiates between automatic *versus* controlled processes within each of these four units. Finally, he attempts to map each of them onto different corresponding anatomical brain regions (see Chapter III).

As the main focus of the present doctoral thesis was directed towards emotion perception and regulation in a social context, greatest attention will be attributed to the first (*"understanding others"*) and third (*"controlling oneself"*) unit of Lieberman's (2007) model. For reasons of simplification, the first and fourth (*"processes that occur at the interface of self and others"*) unit will be combined to assess the role of emotion perception in social neuroscience more comprehensively. Finally, the two categories with the titles *emotion perception* (Chapter I.3.1.) and *emotion regulation* (Chapter I.3.2.) will be extended with theoretical aspects of the corresponding research domains. This should serve as a theoretical basis for understanding the intentions and obtained results of the experimental work of the present thesis (see Chapters V and following).

I.3.1. Emotion Perception

Emotion perception[3] represents a crucial skill for effective social functioning, importantly contributing to the understanding of others, as well as processes that occur at the interface of self and others [6]. It involves automatic and implicit mechanisms such as imitation / mimicry (Chapter 3.1.2.), empathy (Chapter 3.1.3.), attitudes, stereotypes, or prejudices (Chapter 3.1.4.), feelings of social connection / rejection (Chapter 3.1.5.), as well as more voluntary, explicit functions such as theory of mind (TOM; Chapter 3.1.6.) and social decision making (Chapter 3.1.7.).

Before briefly describing the implication of emotion perception in the domain of social neuroscience, a general introduction into the fundamental mechanisms of emotion perception will be provided (see Chapter 3.1.1. below).

3.1.1. Fundamental Mechanisms of Emotion Perception

Even though emotion perception in a social context can rely on several different modalities, most previous research – particularly in the domain of neuro-imaging including fMRI – has concentrated on the visual modality (as was also the case for the present thesis). Moreover, most human lesion and neuro-imaging studies investigating social perception used emotional faces as prevailing stimulus category. Consequently, the theoretical aspects listed below are mainly concerning the perception of emotion from facial expressions. These include three central elements, namely (1) the general organization or structure of emotions, (2) the perceptual stages of stimulus processing, and (3) eye gaze.

One of the central facets of emotion perception theory is tightly linked with the question about the **general organization or structure of emotions** [25, 26]: Are emotions perceived as discrete states representing different emotion categories? Does emotion recognition occur by the

---

[3] *Emotion perception* in the context of this doctoral thesis does not only refer to early visual emotional processes, but also includes associated knowledge representation important for emotion recognition [26]. This term is therefore used to describe the processing of emotions in general (similar to the term *social perception* [see Chapter I.2.1.]).

detection of basic stimulus properties? Are emotions triggered by stereotypical stimulus-behavior associations? Or does emotion recognition depend on subjective appraisal?

A first line of argumentation [27] proposes that emotions are expressed and perceived as universal categories with a biological origin, corresponding to *basic emotions*, including anger, disgust, fear, happiness, sadness, and surprise (see Figure M1a). Similar to this approach put forward by Ekman, other theorists such as Russell [28] and Rolls [29] also derived categorical models of emotions, but defined the latter by low-dimensional spaces determined by elementary stimulus properties such as valence *versus* arousal (see Figure M1b) or reward *versus* punishment (see Figure M1c), rather than discrete emotional states. In addition, Panksepp [30] likewise argued that emotions are organized in categories, but conceptualized them in relation to the stimuli that serve as triggers and in terms of the behaviors that are associated with them, including seeking, panic, rage, and fear (see Figure M1d). All the above-mentioned emotion theories thus suggest that emotions are organized in distinct categories, which are somewhat independent of external or internal influence like context or individual biases, respectively. However, the last three theoretical propositions already take into account that *"… our everyday emotion categories [i.e. the ones put forward by Ekman] are probably not the best suited for scientific investigation …"* [25], and suggest different means (i.e. dimensions) by which emotion could be assessed.

Taking one step forward, a different approach developed by Scherer [31] proposes that it is the subjective evaluation or appraisal according to a set of *stimulus evaluation checks* that determines the perception and expression of a certain emotion. In this view, an emotion (and associated facial expression) arises always as a combination of synchronous changes in different bodily systems resulting from the output of the stimulus evaluation check sequence (see Figure M2). A crucial component of this appraisal model of emotions is the fact that individual factors such as goal and need relevance or coping potential are also taken into account, making emotions more dependent on contextual factors as well as individual characteristics (see Chapter IV).

Another important aspect of emotion perception theory is concerned with **perceptual stages of stimulus processing** (i.e. diagnostic features of facial expressions [25]). Thereby, in the domain of face processing, two main issues are usually considered, namely the question about the dissociation between early visual information processing *versus* more abstract associated knowledge representations, and the distinction between configuration- *versus* feature-based information encoding [32-34].

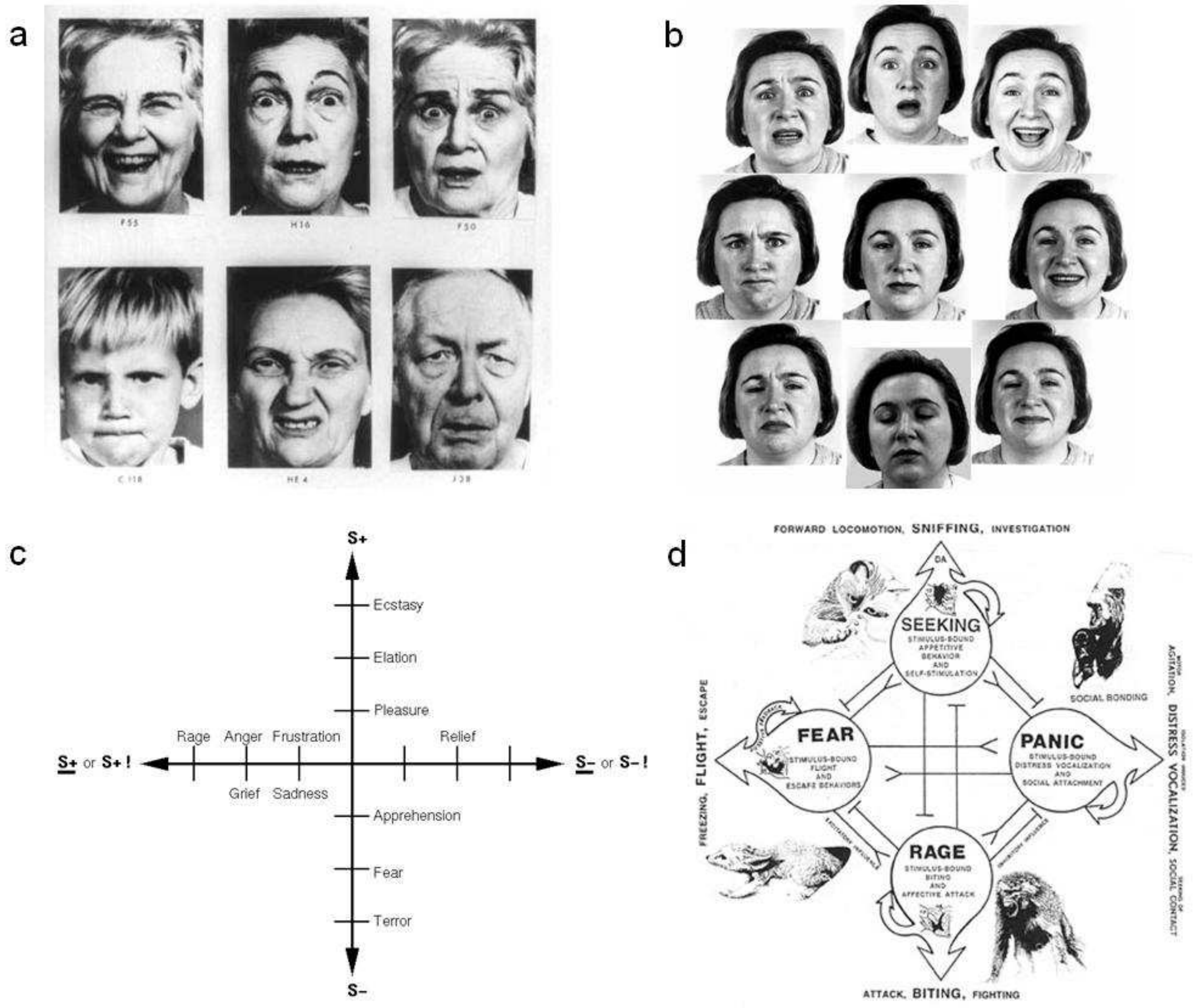


Figure M1: Different Models of general Emotion Structure. a) *The six basic emotions proposed by Ekman (original in [35]). Top: happiness, surprise and fear; bottom: anger, disgust and sadness (f.l.t.r.).* ***b)*** *Categorical Model by Russell (taken from his laboratory website: http://www2.bc.edu/~russeljm/), with valence (negative to positive) on the horizontal and arousal (low to high) on the vertical axis.* ***c)*** *Categorical Model by Rolls (original in [36]), the horizontal axis describing emotions associated with the non-delivery of an expected reward (left) or an expected punisher (right), and the vertical axis describing emotions associated with the delivery of a reward (up) or punisher (down).* ***d)*** *Categorical model by Panksepp (original in [30]) distinguishing four emotion categories – namely fear, rage, panic, and seeking – with corresponding prototypical triggers and/or behavioral responses.*

Basically, the perception of distinct emotions in faces could be achieved by simply comparing visual geometric properties with each other, along the lines: smiling mouth = happy, and wide open eyes = fearful expression. However, such a purely perceptual approach becomes

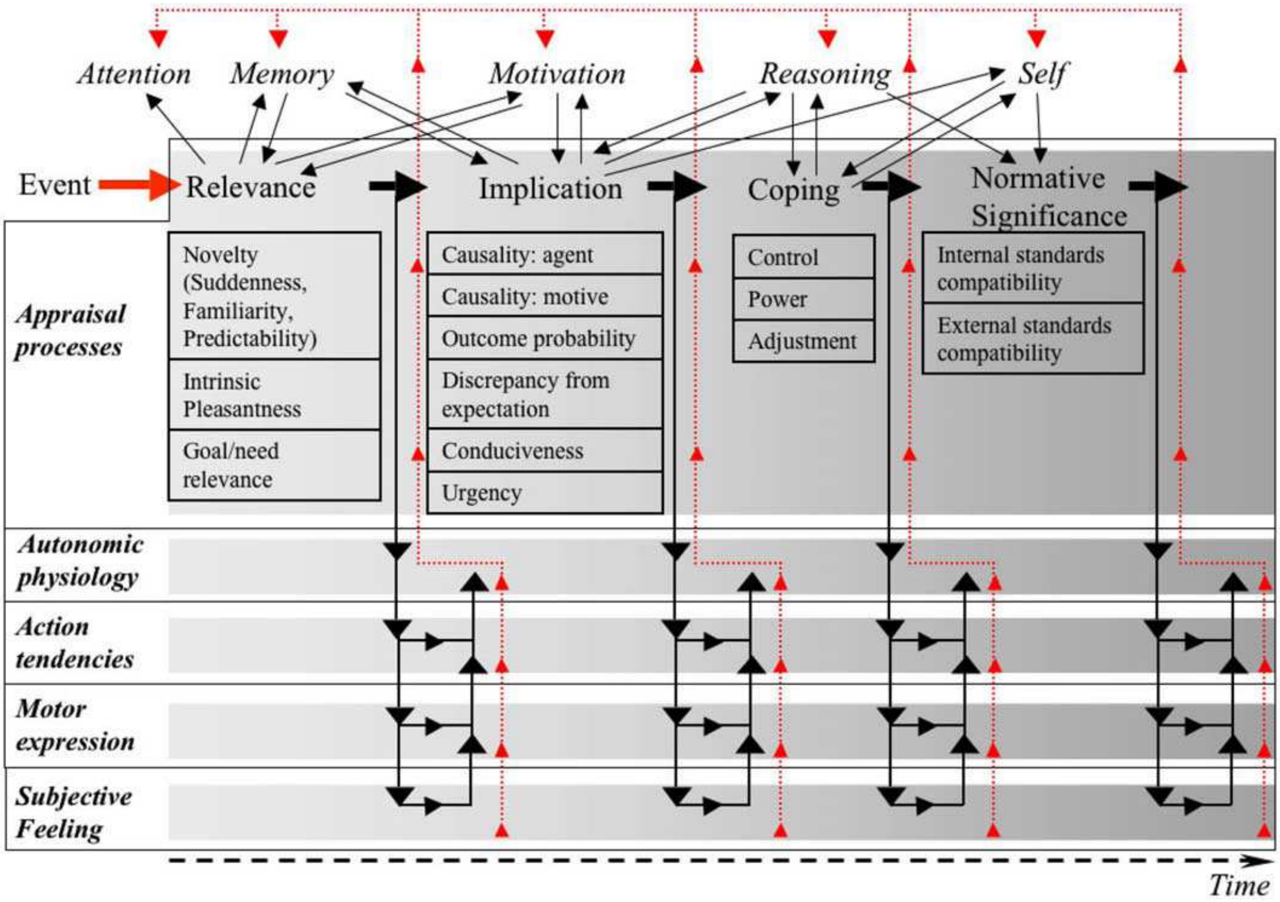


Figure M2:* *Component Process Model of Emotion *(taken from [37], original in [38]). Emotion is conceptualized as a sequence of evaluation checks – namely relevance, implication, coping, and normative significance –, each of which leads to changes in different bodily systems together with corresponding behavioral and affective responses.*

increasingly difficult with more complex social (so called moral or self-conscious) emotions, like i.e. shame or pride. For a correct recognition of the latter, it may be important to consider not only additional information already present in the stimulus itself, i.e., age, gender, sex etc., but also *associated knowledge* from memory about the person and/or a situation where this specific emotional facial expression has been seen before.

On a fundamental perceptual level, an important distinction in face processing has to be made between the perception of emotional expression *versus* identity, or the separation between changeable, configuration-related *versus* rigid, feature-related information encoding. Derived from the initial "*cognitive model*" by Bruce & Young [32], several models of emotion recognition from faces (or face recognition in general) have been proposed during the last two decades, emphasizing these two distinctions either from a functional neuro-anatomical or a sequential (time-course) point of view [25, 33]. This will be described in more detail in Chapter III.1 (see Figures M7 and M8 and).

Finally, the display of similar facial features (e.g. specific muscular patterns or “action units”) can sometimes be perceived as conveying different intensities [39], or even different types [40, 41] of emotions, depending of other concomitant cues (see Chapter IV). Thereby, one important aspect of faces with implications for emotion perception theory partly similar to the above-mentioned distinction between feature- *versus* configuration-related encoding is **eye gaze**.

In this respect, two major points should be borne in mind, either related to the use of eye gaze to successfully recognize emotional facial expressions (internal), or concerning the detection of eye gaze in others as important additional source of social information (external; see Figure M3ab).

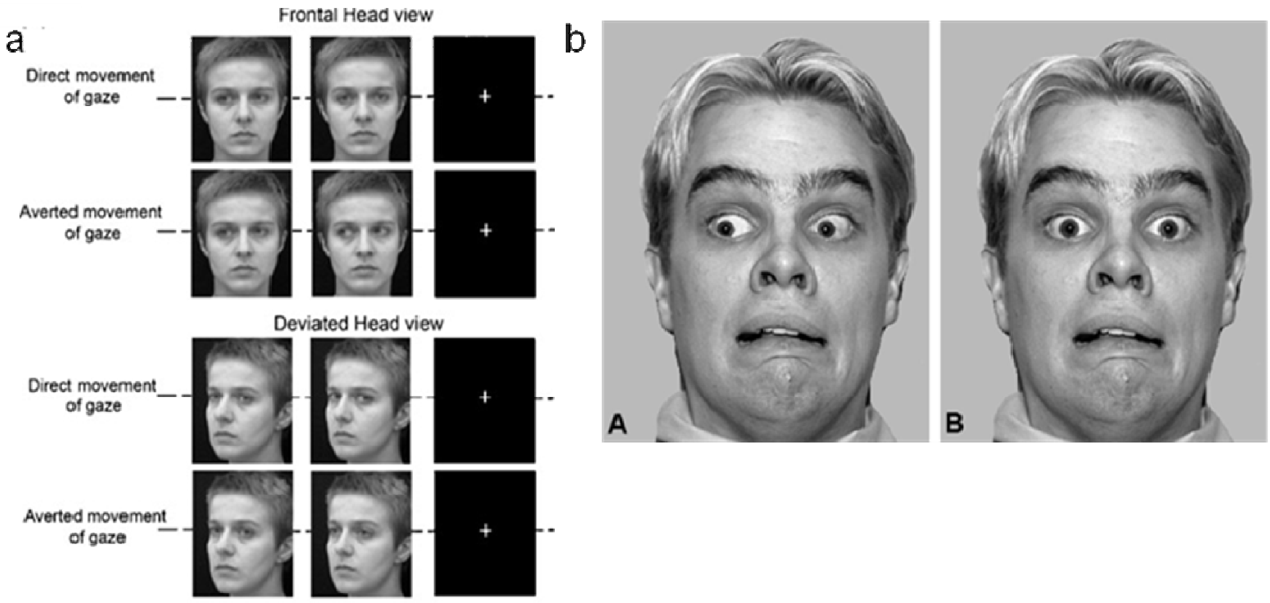


Figure M3: Examples of stimuli used for the investigation of eye-gaze effects. a) *Different combinations of eye-gaze movements with initially frontal head view (top) or deviated head view (bottom; original in [42]).* ***b)*** *Fearful emotional facial expression with averted (A, left) and direct eye-gaze (B, right; original in [43]).*

Firstly, the accuracy of emotion recognition from facial expressions seems to depend on the direction of attention towards the eye region of the face. Such an assumption builds on findings made in a patient called SM with bilateral amygdala damage (see Chapter III.2.1. for more functional anatomical details on this brain area) who showed a selective impairment in recognizing fear due to a lack of spontaneous fixation on the eyes during free viewing of faces [44], as well as on evidence from multiple studies conducted in the context of autism and other brain pathologies linking emotion recognition and theory of mind deficits with abnormal gaze patterns. [45]. Similarly, it has been shown that eye-gaze alters brain activity during emotion regulation [46].

Second, emotion perception from facial expressions is also crucially modulated by the eye-gaze position of the observed other. Several experiments have shown that different brain activation patterns are detected if people are exposed to faces establishing gaze contact *versus* showing an averted gaze, implying either increased early resource recruitment for direct gaze in general [42], or differential activity to specific emotions such as fear [43] and anger [47, 48] as a function of eye gaze .

Taken all together, emotion perception in general and face processing in particular have turned out to be organized in a less straightforward manner than initially thought. On one hand, emotion perception seems not to represent a simple perceptual matching process (basic emotion theory) or the encoding of a restricted amount of distinct stimulus features like valence or arousal (bi-dimensional models), but instead a more complex sequential integration of intrinsic stimulus properties with various subjective factors (component process model / appraisal theory). On the other hand, even though different stimulus features such as identity or emotional connotation are processed separately at early stages (see Chapter III.1. for more details also concerning the functional neuro-anatomy aspects), they ultimately both seem crucial for successful emotion perception and recognition. This implies that it is not appropriate to completely separate the processing of basic from extended stimulus features. Finally, the direction of eye gaze on different areas of the face of the observed person can importantly modulate emotion perception and regulation, and the eye gaze of the perceived face can crucially alter emotion perception in the observer.

After having explored some fundamental factors modulating general emotion perception, the following sections will continue with describing the implication of specific processes entailing emotions in the domain of social neuroscience.

3.1.2. Imitation / Mimicry

The easiest way for understanding the behavior, intentions and emotional experience of others could be achieved by imitation or mimicry of somebody else’s comportment. Based on the discovery of so called “*mirror neurons*” in primate motor cortex that fired during active grasping but also when the same action was only observed [49], a theory has been put forward suggesting that representations of somebody other’s actions – and possibly also mental states and emotions – could be encoded in the brain by an internal “*replay*” of similar self-related past experience. In other words, by i.e. activating the same neural network when seeing somebody smile as during an actively performed own smile, we could get insight into the mind of the smiling person through inferences according to our own feelings and/or mental states associated with it [6]. However, there is a catch. Even though mimicry has already been shown to play an important role in nonverbal communication by i.e. influencing emotion perception during the exposure to emotional facial expressions [50], no study up to date has been able to empirically demonstrate that such processes are tantamount with understanding the other person's mind. This limitation is corroborated by fMRI findings showing that imitation is associated with brain activations representing externally-focused attention (i.e., the processing of external clues such as visual input) [51], and not with internally-focused processes such as mental state representation or empathy [6].

3.1.3. Empathy

Rather than just mimicking or imitating somebody else's behavior or emotional expression, empathy refers to moments when we feel like we are experiencing the world from the point of view of another person [6]. This requires some forms of perspective taking as well as self-knowledge, even though the initial triggering of empathic responses is thought to happen rather automatically [52]. Such mental states have been extensively studied by fMRI related to pain perception in various social settings [52, 53].

One problem with empathy research as a function of pain perception is the difficulty to distinguish between the "*feeling what another person is feeling*" (also termed *empathic distress*), and *personal distress* related to emotion contagion / emotion sharing [52, 53]. This causes trouble not only from a conceptual viewpoint, but also in functional anatomical terms, because both empathic as well as personal distress have been found to activate the same sensory and sub-cortical brain areas (see Chapter III). Therefore, more research is needed to elucidate the precise neural signature of empathy independent of any processes related to mimicry / emotion contagion. A first step into this direction has already been taken by a research group investigating brain activations to images depicting other people in pain in healthy subjects as compared to patients with congenital insensitivity to pain [54], revealing differential effects of arousal and cognitive processes on empathic responses.

3.1.4. Attitudes / Prejudice

As demonstrated by social psychology [7], emotion perception in social contexts crucially relies on fast and implicit evaluations of others, including in particular information from their faces such as emotional facial expressions. With the help of different fMRI paradigms, a major role of attitudes and prejudice has been ascribed to processes related to race discrimination [55-57] or social status [58, 59]. The findings from such paradigms suggest that attitudes are an integral part of social functioning, because they automatically inform us about motivational values or potential dangers of social interactions, and thus guide the selection of appropriate behavior. These effects might reflect evolutionary determined biological mechanisms or learned stigmatizations and cultural stereotypes.

Important for the present thesis is the fact that, even though attitudes and prejudice operate on an automatic, implicit level of processing, the impact of these rapid evaluations on social perception can nonetheless be cognitively modulated or regulated with the help of different strategies [58]. This can be achieved, for example, by shifting the focus of attention from judging the race of people to the evaluation of their food preferences [60], or by processing race targets verbally (pairing a picture with verbal descriptions of race) rather than only visually (perceptual matching of two images) [61]. The same effects have also been reported by simply prolonging the exposure time to faces during race discrimination [62].

3.1.5. Social Connection / Rejection

Emotion perception is also importantly modulated by the intention to establish, maintain, or avoid social relationships – processes that Lieberman (2007) refers to as “*social connection*" and “*social rejection*”. In other words, an emotional facial expression of a beloved person, or someone we care about, should be perceived and interpreted differently from the same emotional facial expression seen in somebody we know from distance or do not know at all, and the same principle should apply for the processing of emotions displayed by friends or foes (*versus* strangers).

Social connection paradigms in neuroscience have mostly focused on the brain correlates of maternal and romantic love by presenting pictures of own children or romantic partners [63]. Both of the latter stimulus categories have been found associated with increased activations in reward-related areas of the brain coding for the pleasurable experience of seeing somebody one is in love with or one loves as a parent, which in turn represents the positive hedonic value of love in general. At the same time, a decrease in activity of the theory of mind network while seeing beloved individuals has been described, which was attributed to the fact that such people don't have to be evaluated as strongly as others or strangers regarding their intentions and mental states [33].

Concerning *social rejection*, it has been found by recent neuro-imaging experiments that the perception of social exclusion or rejection not only causes cognitive and emotional conflict, but also seems to positively correlate with subjective emotional distress ratings [64, 65]. This is explained by the originally biological function of pain induced by the separation of cubs from their mother (representing social rejection), because this typically resulted in death of the concerned animals. It is therefore thought that this ancient form of social pain evolved to represent a more elaborated form of emotional distress to social rejection in humans [6].

Additional functional anatomical details related to social bonding and conflict will be described in the corresponding sections of Chapter III.

3.1.6. Appraisal

Finally, emotion perception has more recently been linked with very early stimulus evaluation processes, as suggested in the appraisal theory of emotion [66] (see Chapter I.3.1.1. and Figure M2). Accordingly, a (social) emotional stimulus undergoes a series of sequential appraisal checks which determine the final physiological, behavioral, as well as subjective feeling responses to the latter, which in turn are modulated by different neural outputs to the same underlying stimulus. In other words, the appraisal theory suggests that there will only be a neural and subsequent bodily reaction to a given (social) emotional stimulus if the latter is appraised as relevant, and these reactions will be further modulated by the three subsequent appraisal checks.

Within this theoretical framework, the first appraisal check – termed *relevance* – has been experimentally investigated the most up to date. This particularly concerns the temporal unfolding of the included appraisal processes – namely *novelty*, *intrinsic pleasantness*, and *goal / need relevance*. Thereby, it could have been shown that novelty assessment indeed precedes the appraisal of pleasantness, and that the goal / need relevance check is computed subsequently [67, 68]. In addition, several fMRI studies have begun to investigate the functional-anatomical correlates of relevance detection, according to a seminal article by Sander and colleagues [69]. The latter attributes a major role to the amygdala (see Chapter III.2.1.) as the central structure in relevance detection.

3.1.7. *Theory of Mind*

More elaborated processes influencing emotion perception are referred as to theory of mind (TOM), which is defined as the representation of the minds of others through mentalizing capacities [9, 12].

TOM mainly involves the representation of the psychological states as well as traits of a person. Whereas the psychological states allow inferences about how a typical individual would respond to particular events, the traits concern information about how a particular person with specific kinds of dispositions would act in the same context [6]. To keep track of such information, it is helpful to distinguish between people that are already known and others that have not been met before, because the former are more predictable in terms of their potential behavior in a specific situation than the latter. This is attained through person recognition via face memory [33], but is also possible via familiar voice recognition [70]. In addition, TOM also comprises the recognition of the current intention a person is having at the moment of interaction, which can be achieved by reading his/her face and/or body posture through the interpretation of biological motion or perceived action [71].

Along these lines, the perception of an emotion from somebody's face can be influenced by the mental representation of a previous reaction of the same or another person in a comparable context according to general (psychological states) as well as specific situational knowledge (psychological traits). This process has been found particularly efficient when the self and the other are in some way perceived as similar [72], implying that the understanding of another person's mind is facilitated if more knowledge about oneself can be incorporated in the derivation of a mental "*other model*".

3.1.8. *Social Decision Making*

Finally, the most complex and advanced mechanisms influenced by emotion perception in an interpersonal context are attributed to social decision-making, including *moral reasoning, fairness,* and *trust* [6].

Moral reasoning, or "*knowing right from wrong*" [73], has mostly been investigated through moral dilemma experiments. The latter involve scenarios where the study participants are asked to choose the morally better way out of two possibilities to resolve a life-threatening situation – i.e. the famous *trolley problem* [74, 75] – revealing the generation of both cognitive and emotional conflict

[76]. The importance of such an interplay between emotion and cognition in moral dilemmas was directly shown with the help of data from patients with selective prefrontal brain lesions [77]. If, due to prefrontal brain lesions, the emotional component of a decision cannot be integrated anymore for choosing between the more logical (utilitarian) or the more emotionally aversive option, the former is selected significantly more likely. This suggests that (implicit) emotional inputs activated by the moral dilemma in prefrontal systems subserving emotional functions (e.g. OFC or VMPFC; see Chapter III) are crucial for deriving an appropriate moral decision.

On the other hand, experiments investigating the neural correlates of *fairness and trust* [78, 79] have found that reciprocal social interactions and responses to cooperative partners are hedonically linked with positive experience and neurally encoded with differential reward-related brain activity. Even though many – if not most – of the fairness and trust studies have been conducted from a neuro-economic point of view involving monetary gains and losses, researchers agree that *“money is not the only motivator”* [79]. In contrary, they argue that genuine social preferences (fairness, trust and altruism) are organized in a very similar way to economic preferences (monetary gain) in the human brain, and that *“... social sciences might benefit from social neuro-economics because of the potentially unifying force of neural data for choice-based approaches ...”* [80].

In sum, social decision making is importantly modulated by emotion perception, determining not only whether a certain situation entailing a moral dilemma will be resolved in a more logical or affective way, but also how interactions with others in terms of fairness and trust are integrated into emotional responses during social interactions.

I.3.2. Emotion Regulation

Social perception is not only concerned with perceiving and interpreting emotions from others, but also with processes occurring within the self as a response to such external affective inputs. Thereby, a central role is accorded to emotional self-regulation capacities.

Self-regulation is a crucial skill for personal as well as social functioning, particularly in the context of controlling emotional impulses in favor of obtaining a long-term goal [6]. This is also known as delay gratification [81] or delay discounting [82]. Moreover, the capacity to cognitively regulate (re-appraisal) or behaviorally modulate (suppression; see Chapter 3.2.1. below) one’s own emotional responses has turned out to be an important predictor of physical, psychological as well as social functioning, and has been linked to both positive (if used appropriately) and negative (if malfunctioning) health outcomes [83, 84].

3.2.1. Re-Appraisal versus Suppression

Emotion regulation (ER) theory [84, 85] distinguishes between two main ER strategies, namely cognitive re-evaluation or *re-appraisal* (antecedent-focused emotion regulation; REAP), and behavioral inhibition of emotion expression or *suppression* (response-focused emotion regulation; SUP). These are conceptualized to occur at different time-points during emotion perception and to lead to different psychological and physiological outcomes (see Figure M4 below).

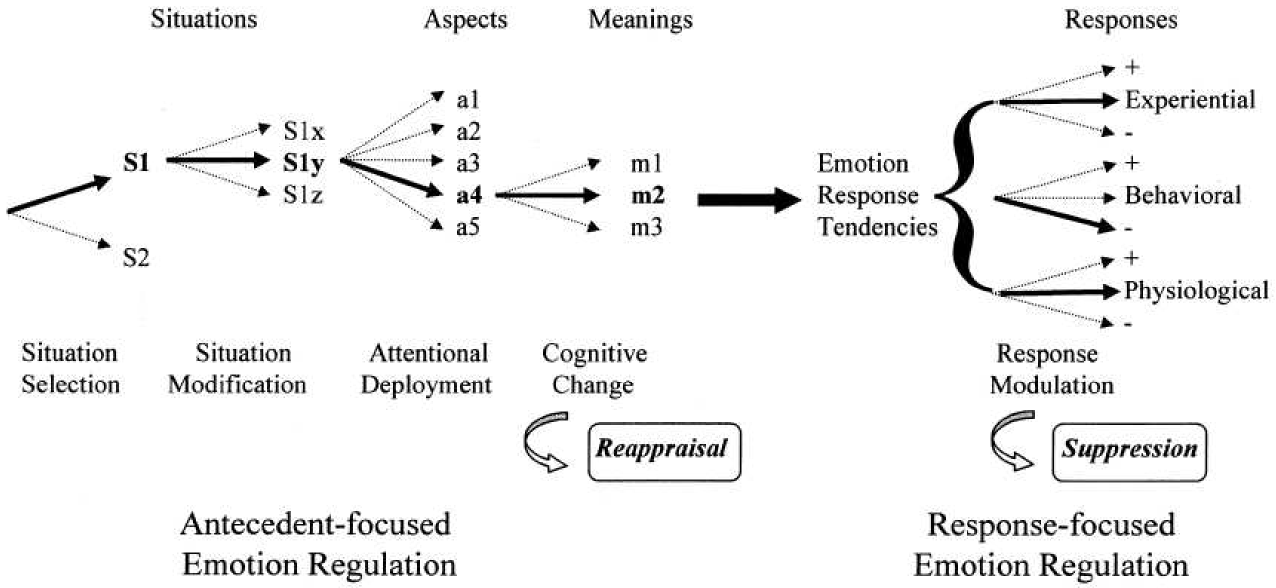


Figure M4: Time-course of emotion perception and the respective onsets of REAP and SUP (original in [84]). The simplest and earliest possible means of ER is to use attentional control to either avoid or direct attention towards an emotional event. Still before the arising of any emotional response, REAP can be applied to cognitively alter the meaning of an attended situation. Finally, already after the occurrence of an emotional reaction, SUP can be used to suppress the behavioral and/or physiological responses to the latter.

REAP is aimed at changing the impact of an emotional event by altering the meaning of a certain emotional stimulus in the first place. This can be achieved by either down- or up-regulating both negative as well as positive emotions, mostly applying strategies during which the emotional scenario is 1) imagined to get better or worse, 2) treated as if it was real or unreal or 3) altered in terms of personal closeness or distance [86, 87]. Generally, REAP is associated with decreased negative but increased positive emotion experience and expression, and seems not to entail any increases of physiological parameters, nor to have any aversive cognitive consequences [84].

In turn, SUP only comes into play once an emotional reaction has already occurred and is aimed at inhibiting the latter on a behavioral and/or physiological level so that negative (or positive) emotion can still be experienced inside but should not be visible from the outside or bodily expressed, for example through emotional facial expressions. Quite in contrast to REAP, SUP has been found to increase sympathetic activation, to have little impact on negative emotion experience, and to even decrease positive affect. Moreover, SUP has been associated with negative cognitive consequences such as impaired memory formation due to heightened cognitive resource consumption [84].

3.2.2. Intentional versus Un-Intentional Self-Regulation

Besides differentiating between REAP *versus* SUP, Liebermann (2007) proposes an additional distinction between intentional *versus* unintentional self-regulatory strategies.

Intentional self-regulation is mentioned in conjunction with *impulse control*, as the process of volitionally overriding a pre-potent response. This involves 1) the detection of conflict between a current goal and probable interfering impulses; 2) the employment of top-down control to inhibit any inappropriate response; as well as 3) working memory engagement to keep track of any existing long term objective [88, 89]. Implicated processes are, as already mentioned above, REAP [90] or SUP [91, 92]. REAP is mostly used in the context of decreasing negative affect by imagining a possible scenario to be less aversive with the help of different strategies (see Chapter 3.2.1.), but can also be applied to increase negative emotions, in addition to modulate positive emotions in either direction [87]. SUP in turn is mostly employed to specifically inhibit or suppress any emotional bodily reaction to a given stimulus.

Concerning *unintentional self-regulation*, Lieberman (2007) refers to *verbalizing* or *labeling* (“*putting feelings into words*” [93]) as well as to *expectancy-based placebo effects* [94], which have been shown to produce similar effects and involve the same brain areas (see Chapter III) like REAP, even though occurring more or less automatically without explicit control.

I.4. Summary

Social neuroscience is a highly interdisciplinary new research field, integrating cognitive, emotional and motivational, as well as personal and social contextual mechanisms, with a neurobiological perspective.

The derivation of neural models of social perception as the principal domain of social neuroscience has been inspired by two main psychological inquiries, namely developmental and social psychology. The second account has had a determining influence on the conception, implementation, and interpretation of the experimental work carried out as a part of the present thesis.

Firstly, the focus of attention was directed toward emotion perception and regulation, which are emphasized as two core processes in social neuroscience.

Second, within these processes, the influence of social context and individual differences (representing social motives and goals) on neuro-physiology was investigated. This comes from the fact that social psychology highlights these situational effects as key determinants of human social behavior.

And third, both implicit, automatic as well as explicit, controlled processes were considered and finally integrated to provide a best possible comprehensive view of the obtained results.

For more details on the distinction between automatic *versus* controlled processes, please refer to Chapter III. More information on social context and individual differences will be provided in Chapter IV.

CHAPTER II

An Evolutionary Perspective

As mentioned in Chapter I, the derivation of neural models of social perception is a very new development in psychology and cognitive neuroscience. Even though social psychology theories are now considered to provide a more appropriate framework for many experimental findings [6, 18], the field was initially founded through the investigation of explicit mechanisms related to social cognition – more precisely theory of mind and mentalizing –, and thus based on developmental psychology accounts [9, 14, 95]. The latter also emphasized the uniqueness of certain human social skills as compared to non-human primates and other mammals, and thus fuelled the interest in elucidating the possible evolutionary mechanisms of social processing. Such dynamics culminated in the proposal of the *social brain hypothesis* [17].

II.1. Unique Human Social Skills and Mammalian Brain Evolution

Regarding the most fundamental capacities of social cognition, there are not many differences between young preverbal infants, apes and monkeys. All three of them readily recognize conspecifics, monitor actions of others, and engage in contingent interactions. However, as adults, our species displays at least two uniquely human aspects of social cognition: the capacity for theory of mind and the representation of triadic relations. Such skills are anatomically linked with a widespread network of (neo)cortical brain areas, including – amongst others – the temporo-parietal junction (TPJ), superior temporal sulcus (STS), and medial prefrontal cortices (MPFC; see Chapter III) [13]. Because the size of the human brain and particularly the neocortex has disproportionally increased in size during evolution from small mammals to non-human primates and finally to our species, one intriguing hypothesis suggests that it is the very social nature of humans that accounts for their large brains [17].

II.1.1. Social Brain Hypothesis

Traditionally, evolutionary processes have been attributed to natural selection, favoring the "*survival of the fittest*" [96]. Accordingly, the evolution of brains was ascribed to an improvement of the capacity to process ecologically relevant information. This seems plausible if one considers the fact that some particular non-human primates and apes acquired complex problem-solving capacities (also called *extractive foraging* [97]), but fails to explain why *all* primates – including the folivores – have larger brains than other mammals [17]. This is particularly important when keeping in mind that the adult human brain only makes up about 2% of body weight but consumes about 20% or total energy intake [98], leading potentially to a very disproportionate balance between costs and benefits. Consequently, an alternative brain evolution hypothesis was brought forward by Robin Dunbar (1998), named the *social brain hypothesis*. The latter argues that *"… the competition for social skills led to the evolution of cognitive mechanisms for outsmarting others, and fuelled the expansion of the human brain and perhaps the elaboration of certain neural systems …"* [95].

Of central importance in the social brain hypothesis is the fact that only the neocortex ratio has been found to positively correlate with different indices of social complexity – such as social group size for the anthropoid primate species [99] or the mating system in primates [100]. No such relations could be established for other brain areas, like the hippocampus [101], primary visual cortex, or sub-cortical areas involved in emotion processing like the amygdala [102]. There is even the notion that "*emotional*" brain areas have progressively reduced in their relative size in favor of more "*executive*" brain centers, reflecting a shift away from emotional to deliberate control of behavior [103]. Taken all together, it is unlikely that the human brain has increased in size during evolution due to purely visual or emotional mechanisms, or because of an increase in memory capacity *per se* [17].

Along these lines, an appealing alternative for the driving force of human brain development is seen in the ability to manipulate information about social relationships [17]. This notion is corroborated by several lines of evidence, including – amongst others – 1) a positive correlation between the use of tactical deception rates and neocortex size in primates – a behaviour that is also indicative of mentalizing and theory of mind skills [104]; 2) a negative relation between neocortex size and male rank correlation with mating success in polygamous primates – meaning that the bigger the neocortex of lower ranking males, the more mating success they have, presumably because they can deploy more sophisticated social skills to circumvent dominant males [105]; 3) a positive correlation between adult neo-cortex size and the length of the juvenile period in primates – implying that it is not the hardware embryological development, but the "software programming" during social learning that accounts for large brains in primates [106]; and 4) a positive correlation between grooming clique size and relative neo-cortex size in primates, including humans (see Figure M5) – being interpreted as *"… a direct cognitive limitation on the number of individuals with which an animal can simultaneously maintain a relationship of sufficient depth that they can be relied on to provide mutual support when one of them is under attack …"* [17].

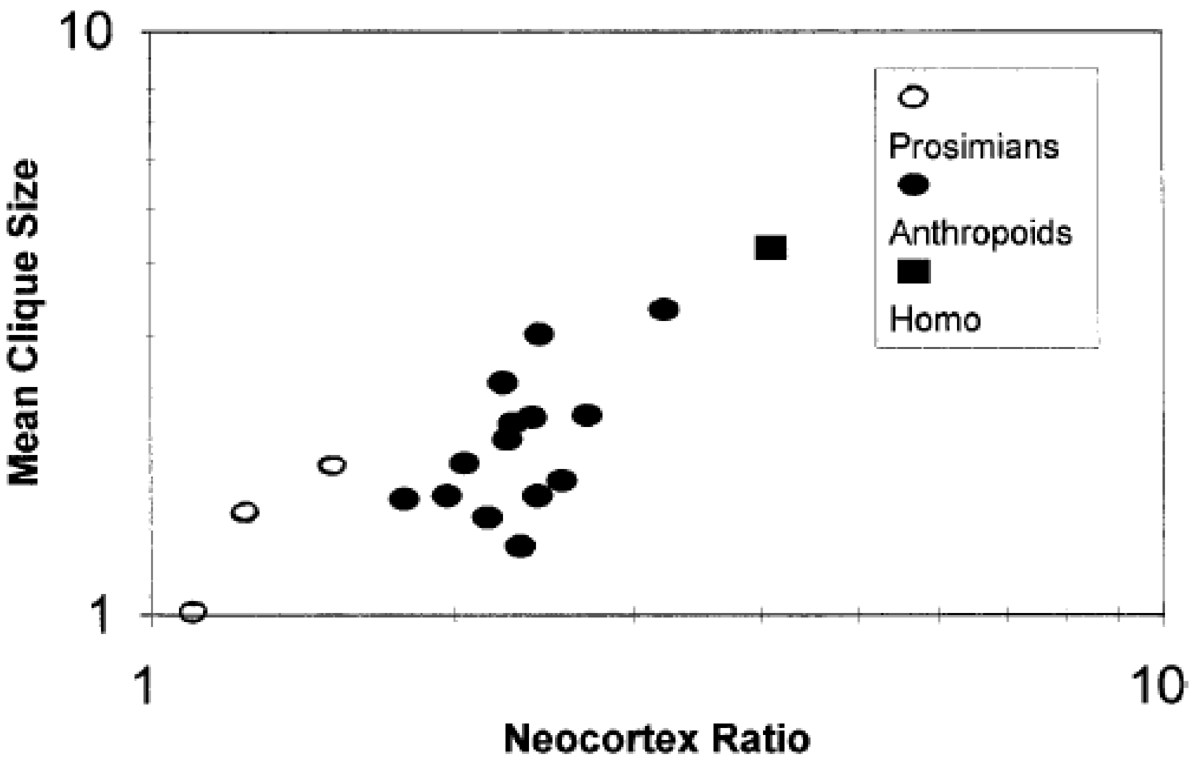


Figure M5: Mean grooming clique size plotted against mean neo-cortex ratio for individual primate genra *(reproduced from [17]).* The square is Homo sapiens. Prosimians include primates that are not monkeys or apes (i.e. lemurs), whereas anthropoids comprise the group of great apes (like gorilla, chimpanzee, and orang-utan).

The account of Saxe (2006) on the two uniquely human social cognition skills (the capacity for theory of mind and the representation of triadic relations) converges with the social brain hypothesis [17] that also stresses theory of mind as being the primary driving force of evolutionary neocortex size increase, and suggest that specific social cognition skills might mark the most significant difference between humans and non-human primates and other mammals, and that they evolved due to increasing computational demands as a function of social complexity.

For the present thesis, the social brain hypothesis does not represent a crucial account regarding experiment design or interpretation of obtained results. Nonetheless, it provides valuable information on the development of some higher-order cognitive processes involved in social perception, such as theory of mind and mentalizing. Because the latter will be mentioned repeatedly in the remaining part of the thesis like, i.e., in the results and discussion chapters, it is useful to know which biological functions they are thought to subserve and how these functions emerged during human (brain) evolution.

CHAPTER III

Functional Neuro-Anatomy

III.1. The Social Brain

After having explored the fundamental mechanisms and core processes of social perception and emotion processing in general in Chapter I, and discussed the emergence of specific social skills as the driving evolutionary force of human neo-cortex development in Chapter II, the question now is where these social computations are localized within the human brain. As already mentioned several times, social perception includes a multitude of processes, such as cognitive, emotional and motivational, as well as personal and contextual mechanisms. Hence, it is not surprising that almost the whole brain is involved in perceiving, interpreting, and modulating social emotional information.

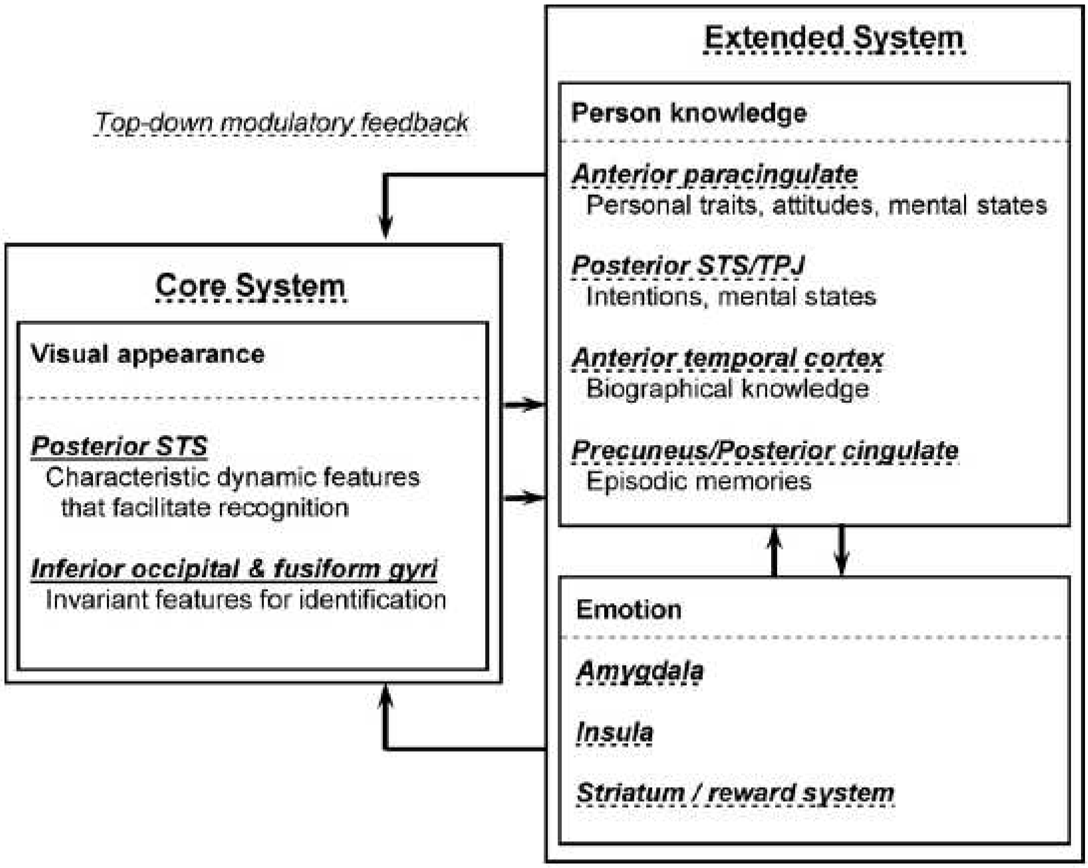


Figure M6: Familiar Face Recognition Model (original in [33]). It includes brain areas coding for the invariant and dynamic visual features of faces, the retrieval of associated person knowledge, as well as emotional components regarding social behavior and social decision making (see Chapters I.3.2 and III.2.).

When it comes to delineating a functional anatomical model of specific brain areas implicated in social perception and emotion recognition, two major models are traditionally considered, representing accounts of either face or emotion expression recognition. As already pointed out in Chapter I.3.1, one of the most elaborated models of face recognition was described by Gobbini & Haxby [33] (see Figure M6 above), while the most comprehensive model for processing of emotional facial expressions was put forward by Adolphs [34] (see Figure M7 below).

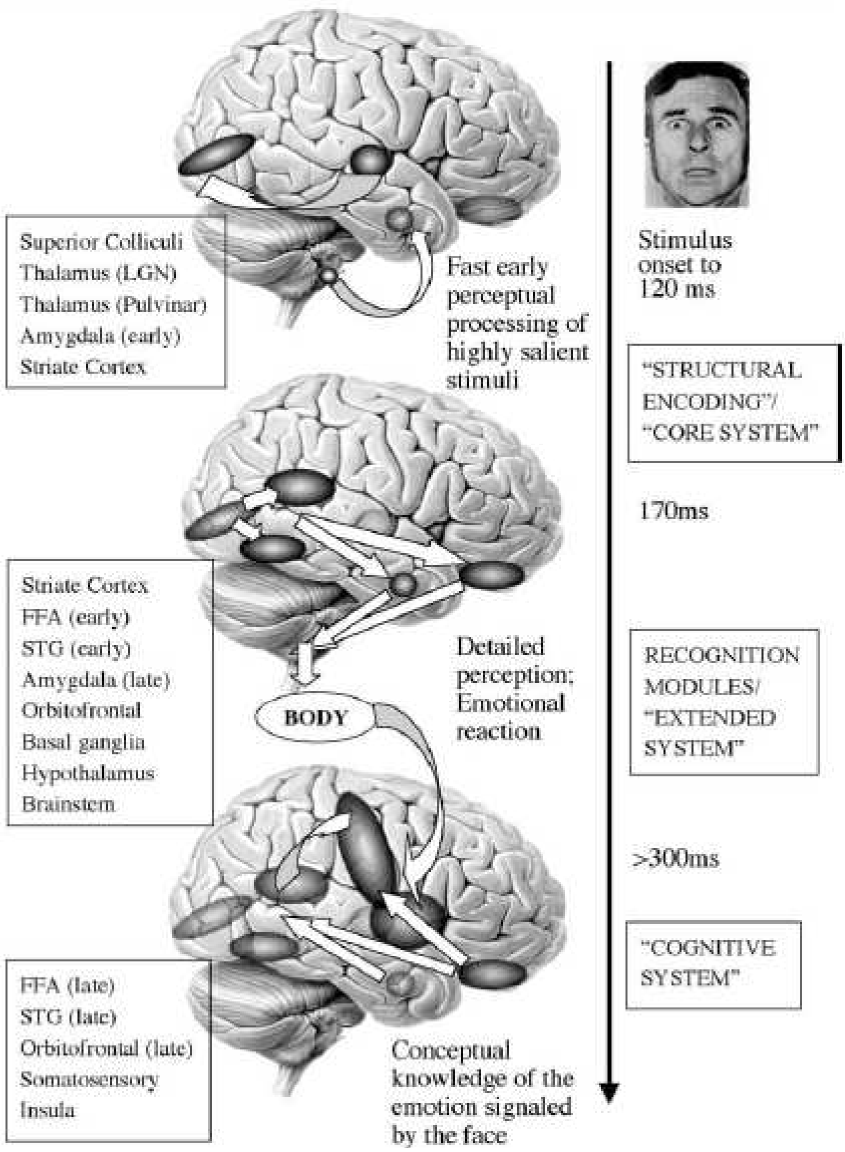
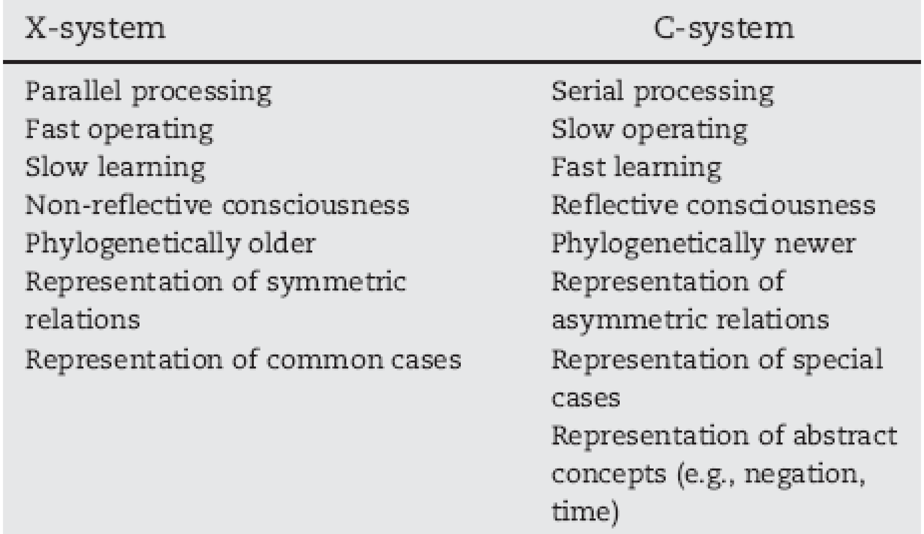


Figure M7: Emotional Facial Expression Processing Model (original in [25]). This model is mostly concerned with early processing steps separating structural *versus* conceptual representations, including the initial perception of an emotion at the top, and the identification of an emotion at the bottom.

Even though these models provide a sound basis for describing the neural substrates of social perception, separating "*core*" and "*extended*" systems that represent the differential encoding of emotion *versus* identity and featural information about social stimuli (see Chapter I.3.1.), they seem somewhat restricted to basic perceptual stimulus properties, particularly in the case Adolphs' (2002) model. Moreover, both accounts do not take consider the distinct pre-frontal cortical and cingulate contributions to social perception. Because the present doctoral thesis was not only concerned with the recognition of emotional facial expressions, but also investigated more complex social perception processes such as the influence of context and individual differences on responses to social partners as well as the regulation of social emotions (see Chapter IV and following), another approach might be more appropriate to describe the brain networks important for social neuroscience.

An alternative approach (see Figure M8 below) distinguishes between different neural systems for social perception and separates between automatic (refle***x***ive, ***x***-system) *versus* controlled (refle***c***tive, ***c***-system) mechanisms [6, 18] (see Figure M8 below and Chapter I.2.2.). The most important conceptual distinctions between these two systems are summarized in Table I.



X-system	C-system
Parallel processing	Serial processing
Fast operating	Slow operating
Slow learning	Fast learning
Non-reflective consciousness	Reflective consciousness
Phylogenetically older	Phylogenetically newer
Representation of symmetric relations	Representation of asymmetric relations
Representation of common cases	Representation of special cases
	Representation of abstract concepts (e.g., negation, time)

Table I: Basic features associated with x- and c-systems (original in [6]).

Even though such a dual-process model separating automatic from controlled mechanisms does not necessarily represent the complexity of social behavior in an ideal manner, it still provides a very useful bridge between already well established behavioral work and new evidence from neuro-imaging. In addition, the dual-process model can readily be transcribed into the context of the present thesis, where emotion perception could be attributed to processes mainly involving the x-system, whereas emotion regulation could be linked with preferential processing within the c-system. As a matter of course, such a distinction will not be applicable without any exceptions, as will become evident by reading the next two sections on the neural substrates of automatic and controlled social perception, respectively.

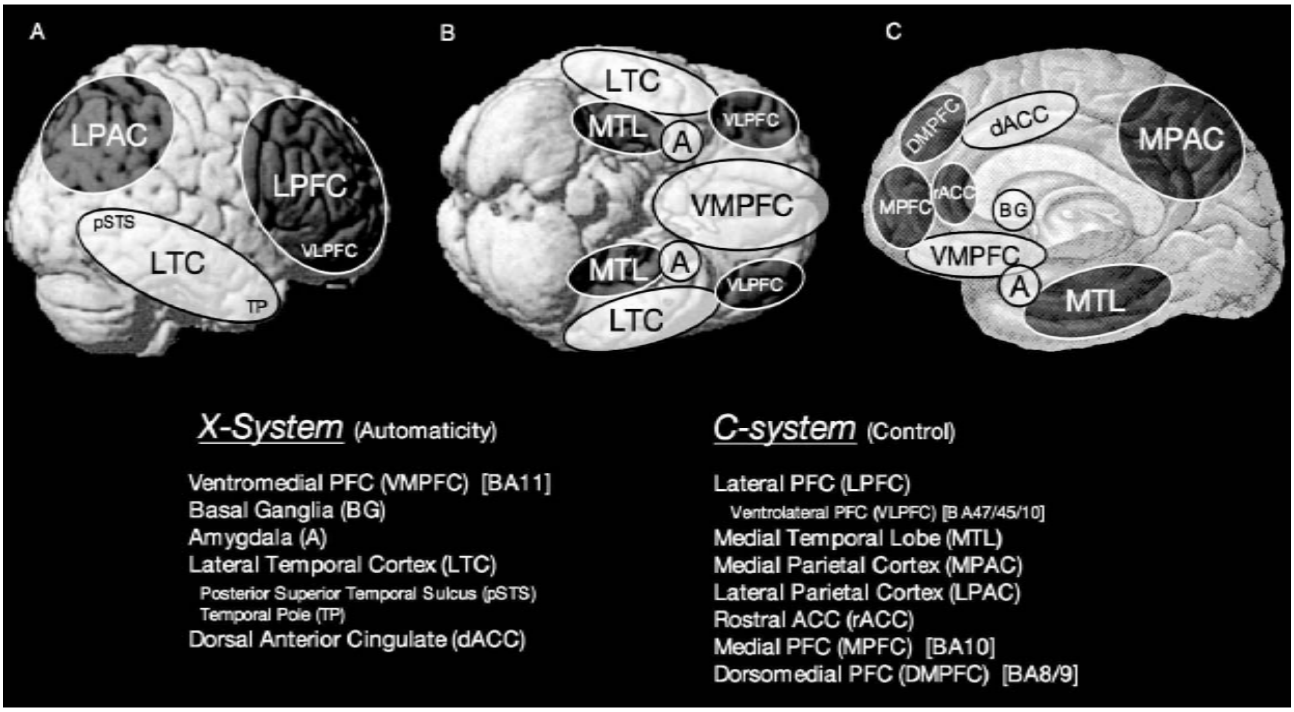


Figure M8: Reflexive (x-system; left) and reflective (c-system; right) brain systems *(reproduced from [6], original in [24]). For more details, please refer to text in Chapter III.*

III.2. Neural Substrates of Automatic Social Perception

III.2.1. Amygdala

The original description of amygdala function related to emotion perception was established more than seventy years ago by Kluver and Bucy [107]. By performing temporal lobectomies in rhesus monkeys, these authors observed reduced fear-responses, increased exploratory behavior, and hyper-sexualism in the lesioned animals [108], all constituting important alterations of fundamental social behaviors (today also known as *Kluver-Bucy syndrome*). In humans, evidence for amygdala involvement in social emotional processes became only available from the 1990ies, when lesion studies showed impairments in negative emotion recognition from the face in general, and fear in particular [25]. Such findings were later extended by a large body of research using functional neuro-imaging, which demonstrated a key role for the amygdala in a range of emotional and social domains, including recognition of facial [34] and vocal [66] emotions, discrimination of mutual gaze contact [47, 48], but also personal relevance [69], as well as processing and learning of various emotional stimuli [108]. Amygdala dysfunction has also been found in various disorders with emotional and social disturbances, including autism [109], sociopathy [110], depression [111], or neuro-genetic syndromes such as the 22q11.2 deletion syndrome (see Figure M9 below; [112]).

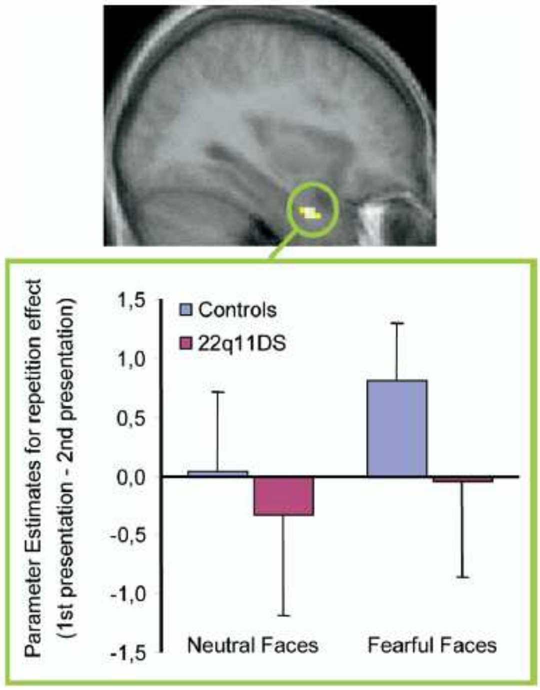


Figure M9: Abnormal response to fearful faces in amygdala for subjects with 22q11.2 deletion syndrome *(reproduced from [112]). A response to fearful faces on first presentation with subsequent repetition effect suppression for fearful faces in the amygdala is observed only in healthy control subjects (blue) but not in patients (purple). No such effect is present in both populations for neutral faces.*

Behavioral animal as well as human functional investigations of amygdala function have particularly revealed its implication in the encoding and storage of memories acquired through fear conditioning, and thus *implicit* negative emotional learning and memory [113]. Subsequently, the amygdala was also attributed an important role in the modulation of *explicit* memory consolidation as a function of arousal, by "… *favoring stimuli that are predisposed to lead to an emotional reaction …"* [108]. By demonstrating that amygdala activation to fearful (versus neutral) faces could be found even if such faces were not attended consciously [114], and that enhanced activation in visual brain areas to fearful faces was absent in patients with amygdala damage [115], a more general function subserving the preferential processing of emotional and potentially threatening information was proposed for the amygdala. Further, the amygdala has been conceptualized as a "*relevance detector*" [69], involved in the appraisal of a wide variety of negative as well as positive [116] and even neutral [117] stimuli according to their significance for the goals, needs and well-being of an organism at a specific moment in time.

More specifically related to social perception, the amygdala has been shown to contribute to the formation of first impressions about people based on subjective preference [118]. Such research was fuelled by findings that amygdala activity reflects social inferences based on trustworthiness [119] and implicit attitudes in the context of racial prejudice [55] and other social stigmas [58, 59]. Because such inferences not only involve information about others but also depend on the self – which is presumably stored as automatically accessible knowledge in the form of strong self-schemas –, the amygdala has also been found to be activated during self-descriptiveness judgments [120].

Therefore, besides mediating basic emotion recognition and implicit as well as explicit emotional memory formation and retrieval, the amygdala is thought to play an important role in automatic social evaluations, by means of evolutionary determined biological mechanisms or representations stored from previous experience.

III.2.2. Basal Ganglia

The basal ganglia are an assembly of nuclei including the *putamen, caudate, globus pallidus*, and *nucleus accumbens*. Generally, these structures are activated during implicit and probabilistic learning, motor skill acquisition, as well as the evaluation and prediction of both negative and positive outcomes, but their ventral division is most prominently described to represent reward-related signals during positive motivational events, such as addictive drug-administration or monetary gains [24].

Regarding social perception, increased activity in the basal ganglia has been found during the exposure to pictures of a romantic partner or own child [63], consistent with a role in coding for the positive hedonic valence of romantic and maternal love, as well as during presentation of unknown happy faces [121], thus representing social reward in general. Moreover, the same brain regions have been found activated in response to cooperative partners or cooperative play, thereby representing fairness and trust [79], also measurable through altruistic punishment [122]. Finally, basal ganglia activation has been observed as a function of social hierarchy, either when increasing social status [123], or when winning more money than somebody referred to for social comparison [124].

In sum, the basal ganglia seem to code for the rewarding value of various positive social interactions and scenarios. These include intimate social cues such as romantic and maternal love, fairness and trust during cooperation, as well as more general social rewards as conveyed by happy facial expressions or increases in social rank.

III.2.3. Lateral Temporal Cortex

The lateral temporal cortex includes two areas important for social perception, namely the posterior superior temporal sulcus (pSTS) and the anterior temporal pole (ATP).

The pSTS has been ascribed an important role in the detection of biological motion and is thus thought to contribute to the decoding of perceived action in others [71], which is crucial for subsequent higher-order theory of mind processes.

The ATP has been found activated during the exposure to a variety of familiar stimuli, including faces, names and landscapes, implying these areas in the retrieval of autobiographical or episodic memory information [33]. In addition, the (superior) ATP has been linked with the representation of abstract conceptual knowledge of social behaviors [125].

Together, the pSTS and ATP therefore importantly contribute to familiar face perception, theory of mind, and flexible evaluation of social behavior.

III.2.4. Dorsal Anterior Cingulate Cortex

The dorsal anterior cingulate cortex (dACC) is located in the posterior region of the rostral medial frontal cortex (prMFC), comprising Brodmann areas 8, 9, 24 and 32, and has been associated with cognitive tasks related to *action representation*, such as attention and error/conflict monitoring [88, 126, 127], but also implicated in the so-called pain-matrix and affective responses during nociception [52].

Accordingly, social neuroscience paradigms have found dACC activation in situations of social separation and rejection, presumably coding for both the cognitive conflict caused by expectancy violation [65], as well as subjective distress (or “pain”) experienced during social exclusion [128, 129]. Moreover, the dACC has been implicated in affective processes related to empathy, especially in paradigms using the observation of another person feeling physical pain [52, 130]. Such results led to the hypothesis that *“… mammalian social pain may have evolved out of the existing system for physical pain …”* [6, 64].

III.2.5. Ventro-medial Prefrontal Cortex

The only cortical prefrontal area – ventromedial prefrontal cortex (VMPFC) – listed as part of the automatic x-system lies at the intersection between Brodmann areas 10 and 11, and also include areas in the orbital medial frontal cortex (oMFC; [126]).

Originally, the VMPFC has been implicated in monitoring of rewards and punishments [131]. More recently, its function has been extended to the guidance of behavior in terms of updating the value of possible outcomes and the monitoring of externally guided actions [126], particularly through reinforcement-guided decision making and in terms of context-sensitive stimulus evaluation [127].

In social neuroscience paradigms, VMPFC activity has been found during the formation of implicit attitudes and behavioral preferences [132], as well as counterfactual judgments and regret [133]. Similarly, VMPFC damage was linked with empathic deficits [134] because VMPFC activity is thought to *"… contribute to the additional processes involved by empathy over direct feeling …"* [6], which is probably related to the fact that VMPFC is strongly connected to sensory association areas (see below – Chapter III.2.6. – and [126, 127]).

In sum, the VMPFC is thought to represent *"… certain aspects of over-learned associated social knowledge …"* [135] and might thus automatically guide social behavior through rapid evaluations of social emotional contextual cues.

III.3. Neural Substrates of Controlled Social Perception

III.3.1. Lateral Parietal Cortex

Lateral parietal cortex (LPAC) comprising the temporo-parietal junction (TPJ) – particularly on the right side – has been found activated in situations where *"… the external visual representation of one's body conflicts with one's internal experience of oneself …"* [6]. This is for example the case during so called out-of-body experiences, where somebody feels and or sees him/herself to be outside his/her physical body [136]. The LPAC is also thought to be importantly implied in agency processing as an integral part of self-perception.

The medial parietal cortex (MPAC) overlaps with the posterior cingulate cortex (PCC) and precuneus (PC). These brain areas have been shown to code for several different aspects of social perception, including long-term memory retrieval with self-centered perspective, mental imagery, familiarity, as well as emotion processing, particularly when related to self-generated emotions [33]. In addition, MPAC activation has been reported during moral judgments of one's own or someone else's behavior [75]. The MPAC thus represents an important interface between the evaluation of current and past experience, both for self- as well as other-related processes.

III.3.2. Medial Temporal Lobe

The medial temporal Lobe (MTL) comprises the hippocampal and parahippocampal memory areas. Concerning social perception, these areas are important for memory retrieval particularly regarding autobiographical knowledge [137]. Thus, they may be linked to *"… one's internal sense of self …"* [6] and memory processes involved in self-perception.

III.3.3. Rostral Anterior Cingulate Cortex

In contrast to the dACC implicated in cognitive conflict and pain (see Chapter III.2.4.), the ventral / rostral part of the ACC (rACC) has been associated with the processing [126, 138] and resolution [139] of emotional conflict in general, as well as with emotional conflict during social rejection [65].

Although a division was proposed between the cognitive (dorsal) *versus* emotional (ventral / rostral) regions in ACC, this has been recently suggested to be replaced by a non-symbolic versus symbolic functional dissociation. This is supported by the findings that the rACC is activated by negative emotions with a particular object (i.e. "*being angry at someone*") whereas the dACC responds to fear and anxiety in general. Further, the rACC was involved in the representation of anticipated pain whereas the dACC coded for the more "*bottom-up*" sensory aspects of pain [24].

III.3.4. Prefrontal Cortical Areas

The c-system also comprises the ventro-lateral prefrontal cortex (VLPFC; Brodman areas 45, 47, and 10), the lateral and medial prefrontal cortex (LPFC and MPFC, respectively; Brodman area 10; also named anterior region of the rostral MPFC [arMPFC] [126]), as well as the dorso-medial prefrontal cortex (DMPFC; Brodmann areas 8 and 9; also termed posterior region of the rostral MPFC [prMFC] [126]) cortices. Each of these regions is ascribed a different role.

3.4.1. VLPFC

The VLPFC has been found to display increased activity during the exertion of different kinds of inhibitory control, including in emotional tasks.

On the one hand, it has been linked with the *"… inhibition of one's own experience during the consideration of another's state of mind …"* [6], an important prerequisite for the development of the elaborated human theory of mind skills. In addition, VLPFC has also been found activated during cognitive re-appraisal [140, 141] and suppression [91] of various emotional stimuli.

On the other hand, VLPFC involvement has also been demonstrated in more unintentional and thus automatic regulation processes, such as affect labelling [93] or expectancy-based placebo effects [94]. Interestingly, such VLPFC activation may display an inverse relationship with activity in the amygdala [142].

In sum, the VLPFC seems to play an important role in emotion regulation through both deliberate as well as more implicit mechanisms.

3.4.2. DMPFC

The DMPFC has mainly been implicated in continuous action monitoring, together with the posterior rostral ACC (Brodmann area 24; [88]), where the so called "*error-related negativity*" has been detected by EEG [126]. In addition, increased DMPFC activity has been found during the processing of ambiguous response feedback [143].

Regarding social perception, the DMPFC has been associated with theory of mind and mentalizing [144], as well as with psychological trait attribution to a person [145]. It has however also been found activated during cognitive re-appraisal [86], and in relation with self-referential memory encoding of social information [145]. Consequently, the DMPFC seems to play an important role in social perception by computing internal representations of one's own and the minds of others, monitoring potential conflicts or errors, and thus *"… continuously updating the value of possible future actions in order to regulate behavior …"* [126].

3.4.3. MPFC and LPFC

In contrast to DMPFC being activated when thinking about private intentions, increased MPFC activation has been linked with thinking about communicative intentions [126]. Along these lines, the MPFC has also been described to play an important role in moral decision making [75], during which one has to integrate both the affective as well as the utilitarian contribution to a possible outcome. Moreover, MPFC activity has been linked with conscious self-reflection (rather than with more automatic self-knowledge [6]) and when judging the psychological traits of somebody who is perceived similar to oneself [72]. In sum, the MPFC seems to be involved in three related processes of social perception, namely mentalizing / theory of mind, self-reflection, and the understanding of others through similarity judgments.

Cncerning LPFC, its main implication has been described in terms of cognitive top-down control. [6]. Such processes include working memory to keep track of a long-term goal over a prepotent response [89] or to remember the task instructions in the context of cognitive re-appraisal [141]. In addition, LPFC activation has been linked with behavioral inhibition or suppression [91].

III.4. Additional areas

III.4.1. Anterior Insula

Not included in the social cognitive neuroscience model of Lieberman (2007) – but mentioned in Adolphs' (2002) and Gobbini & Haxby's (2007) perceptual models – is a brain region that has recently obtained a great deal of attention, particularly in the context of empathy research, namely the anterior insula [52].

The anterior insula is an important part of human interoceptive cortex [146], neurally representing internal bodily states like temperature, hunger, arousal, and information from the gut [52]. The anterior insula has therefore been hypothesized not only to cortically represent such bodily states, but also to serve as a neural substrate for empathic responses [130]. These involve *"… subjective representations of our own feelings … as well as … anticipated emotional states to our bodies …"*, which can *"… serve as visceral correlates of a prospective simulation how something may feel for others …"* [52]. However, as already mentioned in Chapter I.3.1.3., further research is still needed to separate anterior insula activation related to empathy *versus* personal distress and/or emotional contagion related to the monitoring of autonomic changes as a function of arousal or other cognitive processes [54].

Interestingly, as the activation of empathic responses in the anterior insula is thought to happen automatically [52], this region could be added to the x-system proposed by Satpute and Lieberman [6, 24], and it is likely to contribute to automatic affective processes of social perception (see Chapter III.2).

III.4.2. Fusiform Face and Body Area

Part of the neural models proposed by Adolphs (2002) and Gobbini & Haxby (2007), the fusiform gyrus (FG), and more precisely the fusiform face area (FFA), is not explicitly mentioned in the social psychology account of Lieberman (2007). However, as underscored by Adolph's (2009), the FFA constitutes a specialized cortical module of the social brain networks, dedicated to face processing. Originally thought to uniquely code for faces [147], the FFA has later been shown to also represent non-face objects when the latter had been acquired with substantial expertise [148]. New data however indicates that the FFA still shows some specificity for faces if there are competing stimuli present as in a natural visual environment [149].

Some years after the first description of the functional anatomical properties of the FFA, an adjacent and partly overlapping area of the FG was also attributed selective activation to views of bodies and/or body parts. This regions is now known as the fusiform body area (FBA; [150, 151]). It is therefore thought that the human brain also contains a social "module" for body processing, in addition to the fusiform "module" for faces.

As the FFA and FBA process information related to static features of face and body perception, they are seen as important systems for familiar person perception in the context of face recognition [33], and thus likely to contribute to basic processes involved in the understanding of others. Because such computations most likely occur automatically, the FFA and FBA might also be included in the x-system by Satpute and Lieberman [6, 24].

III.5. Summary

In the present thesis, the experimental investigations related to emotion perception (see Chapters and following) mainly aimed at revealing particular activation patterns in reflexive brain systems – such as the amygdala, basal ganglia, or ACC – related to implicit, automatic emotional stimulus processing. However, activity in these brain regions was nonetheless expected to be rather specific and importantly modulated by social context and individual differences, in accordance with appraisal theories of emotion (see Chapter I.3.1.1.).

What is concerning emotion regulation, the focus of attention was directed towards activations in both reflexive as well as reflective brain systems. Thereby, more deliberate cognitive control processes were expected to be found in cortical brain areas, in turn modulating activity in more automatic emotional brain regions. The influence of social context and individual differences was also investigated for processes related to emotion regulation, and thought to have effects on both levels of processing.

CHAPTER IV

The Role of Social Context and Individual Differences

IV.1. Social Context and Individual Differences: Dissociation between Internal and External Factors

As seen in Chapter I, social psychology theories show that social perception 1) is governed by both automatic *versus* controlled processes, and 2) is strongly modulated by emotional / motivational factors through the incorporation of situational context, motives and goals.

Thus, at the roots of social psychology, we encounter the central "*principle of situationism*", or in other words, "… *the power of the situation over behavior …*" [7] (see Chapter I.2.2.). Such an account has been derived from several seminal behavioral studies revealing that *social context* can differentially affect certain behaviors elicited by the same stimuli. This was often compared to visual perception where inherent stimulus features like from, size and color provide the central, more rigid aspect of information, which can be modulated by additional input from the visual background and the rest of the visual field. Along these lines, social context can be conceptualized as an external variable, representing i.e*. "… the actual, imagined, or implied presence of other people …"* [19] or the presence of other social clues in the proximate surrounding of a certain social scenario.

Quite in contrary, *social motives and goals* intuitively imply that internal factors can also influence social perception. Even though such internal factors can be generated from outside through specific conditions or task instructions (see below), they are more readily related to stable personality traits or *individual differences.*

On these bases, one of the central assumption motivating the present work is that social perception, at both the behavioral and underlying neural levels, can be modulated by *external* – context – as well as *internal* – individual – factors, eventually determining how the exact same stimulus is perceived and encoded by a given person in a different surrounding, or by different people with varying degrees of a certain personality trait in the same setting.

Before summarizing the most important aspects of the modulatory role of social context and individual differences from a social neuroscience perspective, two studies by Kim and colleagues will be mentioned [40, 41], because they constitute very good examples of the influence of both external as well as internal factors on social perception.

IV.2. Is Surprise Negative or Positive? Contextual Information *versus* Subjective Judgment

In the years 2003 and 2004, Kim and colleagues published two articles investigating the impact of contextual information [41] (Chapter IV.2.1.) and subjective judgment biases [40] (Chapter IV.2.2.) – and showed strong effects of external (social context) as well as internal (social motives and goals) factors on amygdala and prefrontal cortical activations during the exposure to identical stimuli, namely surprised facial expressions. Because surprise is an emotion associated with ambiguous meaning (e.g. positive *versus* negative), it is likely to be particularly sensitive to these modulatory factors.

IV.2.1. Contextual Information

In the 2004 study, subjects passively viewed surprised faces which were either preceded by a negative (i.e. "*she just lost $ 500*") or a positive (i.e. "*he just found $ 500*") sentence, while undergoing fMRI (see Figure M10a). The results revealed that activity in the left amygdala was increased for surprised faces that were primed with a negative *versus* a positive sentence, whereas there was no such effect for the sentences alone (see Figure M10b). Conversely, activity in two prefrontal brain areas was found to differentiate between positive *versus* negative sentences, with increased activity for negative sentences in VLPFC, but more activation for positive sentences in VMPFC (see Figure M10c). Moreover, connectivity analyses revealed activity in DMPFC temporally linked with brain regions responsive to both faces and sentences. Besides suggesting a valence-based dissociation between VLPFC (negative) and VMPFC (positive), and a modulatory role of DMPFC (i.e. dorsal ACC) integrating regulatory control with affective information, this study could thus demonstrate than the same visual stimulus (i.e. a surprised face) may elicit differential activation in the amygdala, only depending on the simultaneous presence of social contextual cues. Hence, external factors have indeed an influence on neural activity underlying social perception, including on areas such as the amygdala that has traditionally been associated with rapid and automatic processing (i.e. part of the x-system).

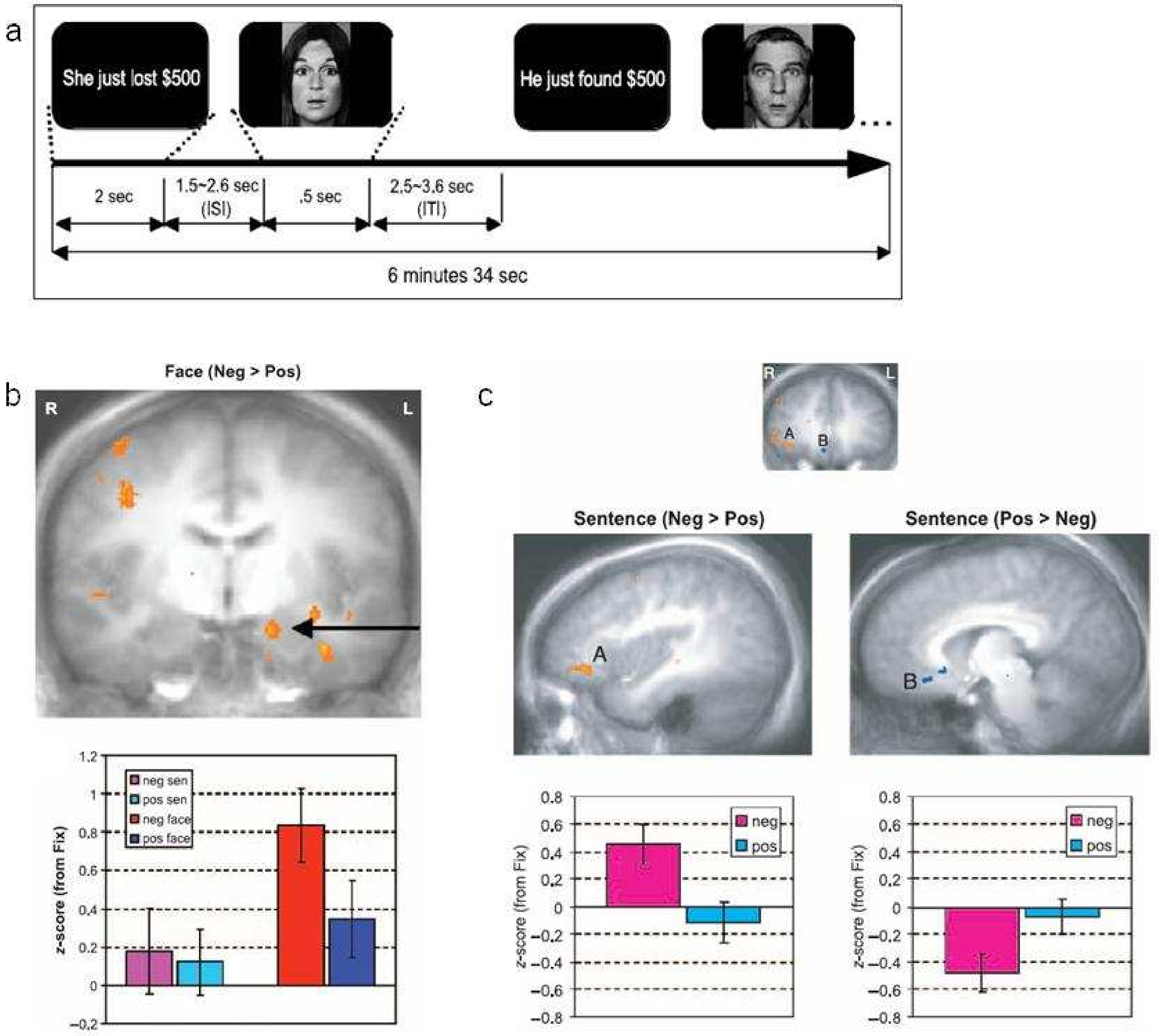


Figure M10: Modulation of amygdala and PFC activity to surprised faces by contextual information *(original in [40]). a) Layout of the experimental paradigm. b) fMRI results for the contrast "Face (negative > positive)" showing greater left amygdala activation for faces preceded with a negative sentence, but not for sentences alone. c) fMRI results for the contrasts "Sentence (negative > positive)" and "Sentence (positive > negative)" displaying increased activity in PFC during sentence viewing. Neg = negative; Pos = positive. For more information, please refer to text in Chapter IV.2.1.*

IV.2.2. Subjective Judgment Biases

In the 2003 study, participants passively viewed surprised and neutral facial expressions whilst undergoing fMRI scanning, and were then re-shown the same stimuli once they exited the machine and asked to provide valence ratings (from 1 = very positive to 9 = very negative; see Figure M11a). Data analysis revealed that in right amygdala, BOLD signal change to surprised –

but not neutral faces – was positively correlated with valence ratings (see Figure M11b). In other words, the more negatively subjects rated surprised faces in a separate rating session after scanning, the more activity in the right amygdala was detected to the same surprised faces during fMRI. Interestingly, the study also revealed an inverse relationship between valence ratings of surprised faces and activation in left VMPFC, where more positive ratings were related with increased activity (see Figure M11b).

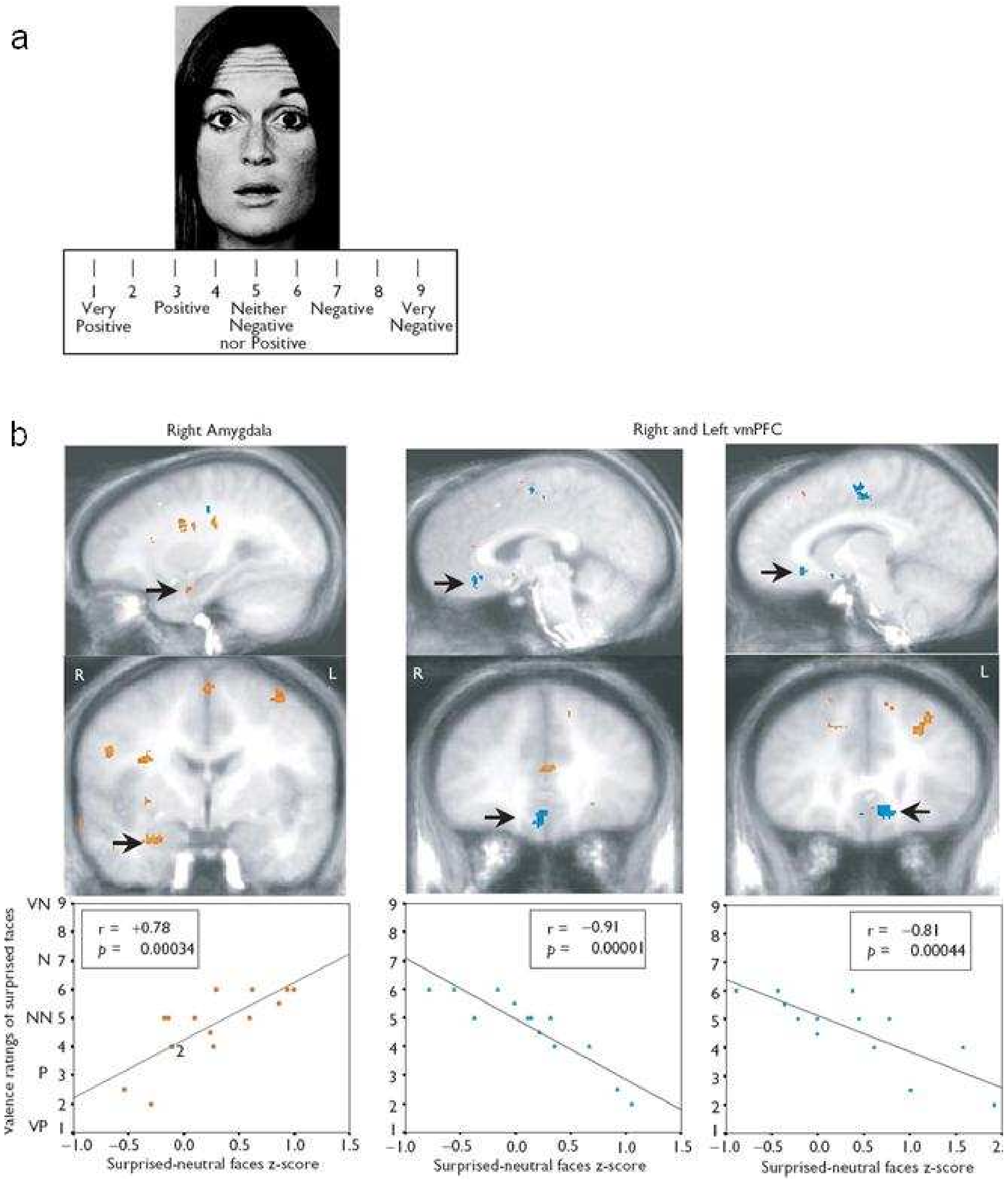


Figure M11: Modulation of amygdala and PFC activity to surprised faces by subjective valence judgments *(original in [40]). a) Rating setting of a surprised face outside the fMRI scanner. b) fMRI results as a function of subjective (negative) valence ratings displaying a positive correlation in right amygdala (left) and a negative correlation in bilateral PFC (right). VP = very positive; P = positive; N = negative; VN = very negative. For more information, please refer to text in Chapter IV.2.1.*

According to the functional neuro-anatomy of the amygdala and VMPFC (see Chapter III), such activation patterns were interpreted as follows: Individual differences in the perception of potential threat from surprised faces determined the amount of signal change in the left amygdala, consistent with a role for this brain area in fear / threat perception. Along the same lines, individual differences in the perception of potential positivity from surprised faces influenced VMPFC activation, as this brain region is known to compute the value of possible outcomes. In sum, it was found that subjective judgment – reflecting internal factors associated with personal motives and goals – modulated activity in two brain areas representing the negative as well as the positive value of an ambiguous surprised facial expression.

IV.3. Social Context: Extended Evidence

Even though the experiment of Kim and Colleagues (2004) very elegantly revealed a direct influence of social context on the encoding of the same otherwise neutral (or at least ambiguous – see Chapter V) facial display, there is, to our knowledge, no other study that used the same approach to further elaborate on context effects in social perception (but see our own experiment [152] mentioned later on).

However, there are other studies employing different approaches to investigate the role of context on social perception. These include *intentionality* in economic games (Chapter IV.3.1), the *presence / absence of an audience* during economic games or imagined moral transgressions (Chapter IV.3.2.), or a more basic *social* versus *non-social stimulus content distinction* (Chapter IV.3.3.). Interestingly, all such studies only investigated context effects on stimulus encoding and memory retrieval, but not regulation (i.e. re-appraisal or suppression), although this is also a crucial component of social perception (see Chapter I.3.2.).

IV.3.1. Intentionality

Besides direct social context comparison (see Chapter IV.2.1.), one set of experiments on social context effects in relation to social perception used the comparison of fair *versus* unfair offers during economic games. The social context effect was generated by either receiving offers from intentional or unintentional partners (see Chapter 3.1.2.) or a computer (see Chapter 3.1.1.). In this sense, the offer received during the economic game constitutes a constant stimulus – like the surprised face in Kim and colleagues' studies –, whereas the fact whether the offer came from a computer, an unintentional human player, or an intentional human player, represents the social context – like the positive or negative sentences in Kim and colleagues' experiments.

3.1.1. Human Player versus Computer

In a first study of this kind [153], the subjects participated in 30 subsequent *ultimatum games*[4], of which 10 were played with 10 people subjects met beforehand, 10 with a computer, and 10 presented as control trials (roulette wheel). Every trial started with the presentation of the interaction partner (human, computer, or roulette wheel), followed by the monetary offer from the partner, which the subject then could either accept or reject (in this study, the participant was therefore always the second player).

The fMRI results revealed that unfair offers from human partners (as compared to similar offers by the computer or during the control trials) elicited highest activity in bilateral anterior insula, ACC, and DLPFC, representing negative emotion processing, cognitive conflict, and executive control, respectively (see Chapter III for more details on the functional anatomy of the these brain regions). Such higher activations to intentional human players making unfair offers suggest that the anterior insula, ACC, and DLPFC did not only represent the amount of money offered during unfair trials, but also the social context – namely the degree of unfairness – that was associated with any unfair offer.

3.1.2. Intentional versus unintentional Human Player

A similar experiment made use of the *sequential prisoner's dilemma game*[5] to establish a moral status for previously unknown faces [154]. The game partners either acted freely and thus intentionally, or were told how to behave, and were thus acting unintentionally. Subsequent to the game, the participants saw the faces of the game partners again, intermingled with new faces, and had to perform a gender decision task while undergoing fMRI scanning. Finally, after scanning, the subjects participated in a memory task and provided likeability ratings for each face.

---

[4] During the *ultimatum game*, two players interact by making decisions upon the division of a given amount of money. The first player proposes how to divide the sum between the two players, and the second player can either accept or reject this proposal. If the second player rejects, neither player receives anything. If the second player accepts, the money is split according to the proposal. The game is played only once so that reciprocation is not an issue.

[5] During the *sequential prisoner's dilemma game*, a first player can trust a second player by sending his/her starting points (transferred to money at the end of the game) to the other player, knowing that each point sent will be tripled. The second player then has the opportunity to reciprocate (by sending money back, which is also tripled), resulting in equal payoffs for both players. The second player, however, can decide to keep the money and maximize his own earnings at the expense of the first player.

The results showed an increased activity in a number of brain areas important for social perception (including amygdala, ventral striatum, and pSTS) in response to the co-operators *versus* defectors, and that the latter effects were stronger for intentional *versus* unintentional agents. The study could demonstrate that reward-related activity during reciprocal trials was not only representing the positive experience of monetary gain *per se*, but was also related to the intrinsic social (intentional) value of cooperation.

IV.3.2. Presence / Absence of Audience

Besides the use of economic games where the game partners can either act in- or unintentionally, two other studies used the imagined (Chapter 3.2.1.) or real presence / absence (Chapter 3.2.2.) of an audience during different social scenarios.

3.2.1. Imagined Presence / Absence of an Audience

In one experiment [155], subjects were shown a set of written social stories and told to imagine them from a first-person perspective while they underwent fMRI scanning. Importantly, all stories had three different endings, which could either be neutral, or include a moral or a social transgression. Finally, these three scenarios could either be experienced alone (i.e. "… *no one else is around to watch you …*"), or with the presence of an audience (i.e. "… *people across the street watch you …*"). Along these lines, the constant stimulus was the imagined social scenario, and the presence or absence of an audience represented the difference in social context.

The results revealed that a single brain region, namely the left amygdala, was more activated for stories with the presence of an audience regardless of the story ending condition (neutral, or moral or social transgression). Such an activation pattern was interpreted by the fact that the description of the presence of an audience was always linked with the notion of *"… staring at the participant …"*, and that the amygdala has previously been shown to be implicated in gaze processing and anticipation of direct gaze [155].

3.2.2. Real Presence or Absence of an Audience

Another experiment used a simple *monetary betting task* where participants had to bet on the color of a card, which could either be red or green with a probability of 50%, so that they could win or lose the amount betted on each trial [156]. The difference in social context was created by the fact that some trials were played by the subject in the fMRI scanner alone, whereas other trials were played by the subject *plus* another person sitting in the control room. Brain activity during the time window between the color choice and the feedback presentation was investigated.

The experimental data revealed that two brain areas, namely bilateral amygdala and right DLPFC, were more activated during the feedback expectation for social *versus* non-social trials. Whereas the DLPFC effect was difficult to interpret, amygdala activation was linked with heightened vigilance, reflecting a warning signal associated with the detection of an unknown "*social other*", promoting avoidance behavior. Moreover, an inverse effect was detected in the ventral striatum, which was more activated for non-social *versus* social trials. This was also interpreted in terms of a warning signal by the presence of a "*social other*" mitigating approach behavior, which is normally encoded by the ventral striatum during reward experience promoting the approach of a positive stimulus [156].

IV.3.3. Basic Social *versus* Non-Social Stimulus Content Distinction

Finally, instead of manipulating intentionality (Chapter IV.3.1.), by involving subjects into one- *versus* two-player economic games (Chapter IV.3.2.1.) or instructing them to mentally imagine social scenarios with or without the presence of an audience (Chapter IV.3.2.2.), one can simply modulate the nature of stimuli themselves by exposing the subjects to social *versus* non-social images in the fMRI scanner. This could represent a basic human animate *versus* inanimate distinction, but also entail a more complex separation of processes related to the computation of complex social interaction scenarios involving social semantic meaning and intentionality as opposed to the mental representation of more basic biological drives, such as appetite/food, desire and disgust [157, 158].

Despite the fact that a social *versus* non-social distinction is rather straight-forward, not many experiments have employed it up to date. However, what is evident from the studies available in the literature is the fact that the amygdala was consistently found to show greater response for social as compared to non-social images, besides an emotional (negative *versus* positive) *versus* neutral effect [157-160]. Thereby, especially the superficial amygdala has been linked with the selective extraction of the social value of incoming stimuli [160]. Other regions preferentially computing social *versus* non-social information have been identified in the STS (theory of mind through perceived action), hippocampus (autobiographic memory), posterior cingulate (evaluation of current and past experience), dorsal ACC (only activated for social sadness / rejection and thus related to conflict / pain), FFA (face perception), and MPFC and DMPFC (mentalizing in general and action monitoring more specifically). In contrary, non-social images were found to induce increased BOLD signal change in the insula (interoception) and in visual cortex (attention). For more details on the functional neuro-anatomy of the abovementioned brain areas, please see Chapter III).

One possible caveat of such a basic social *versus* non-social distinction by means of image content is the fact that there are important discrepancies in stimulus properties. Firstly, social images normally entail a higher complexity and thus require more cognitive computation capacities. Second, social *versus* non-social images tend to differ on the arousal / intensity dimension, especially **if** social images include displays of mutilated bodies (negative) or erotic couples (positive). Finally, inanimate pictures are normally associated with greater emotional control than their animate counterparts [161]. However, when controlling for social complexity, arousal, and emotional control, as well as for other basic visual image properties such as luminance and/or spatial frequency, manipulating the social *versus* non-social stimulus content can be an effective approach (see following Chapters).

In the experimental work of the present thesis, social context was manipulated either directly by pairing different facial emotional expressions (i.e. smiling/happy and angry faces) with verbal information (comparable to Kim and colleagues [2003, 2004] studies) or by using the distinction between basic social versus non-social stimulus properties (see Chapter V). The information provided above about these two methods of context manipulation should thus be particularly kept in mind for the reading of the remaining parts of the present thesis.

Because intentional stimuli were always found to produce larger effects than their non-intentional counterparts (see above), no difference between intentional and non-intentional agents was employed for the experimental work within the present thesis. Likewise, the concept of audience was not applied in the experiments, because previous studies using audience manipulation only found effects related to vigilance or eye gaze (see above), which are not very specific in terms of their impact on social and emotion perception.

IV.4. Individual Differences: Extended Evidence

To measure the influence of social motives and goals on emotion perception, research in social neuroscience has started to make use of several psychological concepts to better characterize *individual differences* in social-affective behavior and the underlying neural correlates. This can either be achieved by means of self-report questionnaires assessing stable personality traits, or by including patient populations that have been diagnosed with a certain mental disorder. In the first case, the questionnaire scores are correlated with brain data in addition to any behavioral measures, which may thus also provide another source of interpretation for the obtained results. In the second case, brain measures are compared between patients and healthy controls, allowing the detection of functional or anatomical differences.

Below, some psychological concepts and individual personality traits that have been examined in a social neuroscience perspective are briefly discussed, including the brain areas and processes implicated. Please note that this list is by no means exhaustive, but rather includes psychological concepts relevant for the present thesis. However, before listing personality traits, a short notice on external guidance effects in relation to social motives and goals will be made.

IV.4.1. External Guidance

The simplest way to assess the influence of social motives and goals on social perception is to use external guidance by giving different task instructions to people evaluating the same stimuli [7]. This has for example been demonstrated in paradigms investigating implicit attitudes in the context of race differences [55, 162]. These studies found that an increased amygdala response in white participants seeing black faces (representing covert racial prejudice) was only present if the participants were performing a one-back recognition task, a categorical same/different task, or a gender categorization task. However, when the participants were instructed to make individuating judgments (i.e. what the depicted person likes or dislikes), or to perform a visual search task, the amygdala effect representing race prejudice diminished below the significant threshold. In other words, by changing the nature of processing and thereby the social motives and goals of the participants through external guidance, the interpretation of the same social stimuli was fundamentally changed, leading to a different neural signature. This means that experiments involving social processes should always be carefully planned in terms of their task instructions, because the latter can crucially determine possible outcomes.

Yet, as social motives and goals are conceptualized as *internal* factors, they should influence social perception also during the absence of any external guidance. In fact, such individual differences seem more interesting from an experimental but also therapeutical point of view, because it is more common to be exposed to challenging social scenarios without being told beforehand how to behave.

IV.4.2. Anxiety: Fear Sensitivity

Differential brain activations as a function of anxiety are normally assessed from two distinct perspectives, either by recruiting sub-clinical "*healthy*" participants classified on their level of trait (and sometimes also state) anxiety through self-report questionnaires (i.e. the Spielberger State and Trait Anxiety Inventory [STAI-S; STAI-T] [163, 164]), or by including clinical populations with different forms of anxiety disorders (i.e. post-traumatic stress disorder [PTSD], social anxiety disorder [SAD], and specific phobia [165]).

Common for both approaches is the assumption that anxiety generally involves an abnormally elevated (pre-attentive) fear or threat response mediated by the amygdala (and insula), with additional emotion regulation deficits originating from impoverished prefrontal control mechanisms [165-167]. This might explain why anxious individuals have particular problems with regulating attention to external threat-related stimuli as well as to internal worries and concerns.

One particular difficulty in the neuro-scientific investigation of the different forms of anxiety is their rather unspecific neural representation. A recent review on the functional neuro-anatomy of anxiety [165] found hyperactivation of the amygdala and insula in all three forms of anxiety assessed (i.e. PTSD, SAD, and specific phobia) – similar to paradigms applying fear conditioning in healthy subjects (see Figure M12) –, suggesting a common engagement of the fear/threat circuitry resulting in shared symptoms among the disorders. Only PTSD was associated with some hypo-activations in medial prefrontal cortex, rostral and dorsal ACC, as well as thalamus (see Figure M12).

Yet, in the context of social neuroscience, it would be particularly interesting to detect differential neural encoding between social forms of anxiety (including SAD) and other, more general anxiety conditions (such as PTSD, specific phobias, and healthy participants with high anxiety scores). One study providing some evidence for the possible existence of such a differential neural representation of social anxiety showed that amygdala activation in adolescents viewing fearful faces was selectively predicted by the social but not the remaining other subscales of the

Multidimensional Anxiety Scale for Children (MASC) [168]. Unfortunately, this result may have a reduced generalizability because of the constraints of subject age, the use of a specific questionnaire (MASC) not applied in adult populations, and the restriction of the analysis to the amygdala. It therefore remains to be further elucidated whether some brain areas code more specifically for social as opposed to other non-social forms of anxiety.

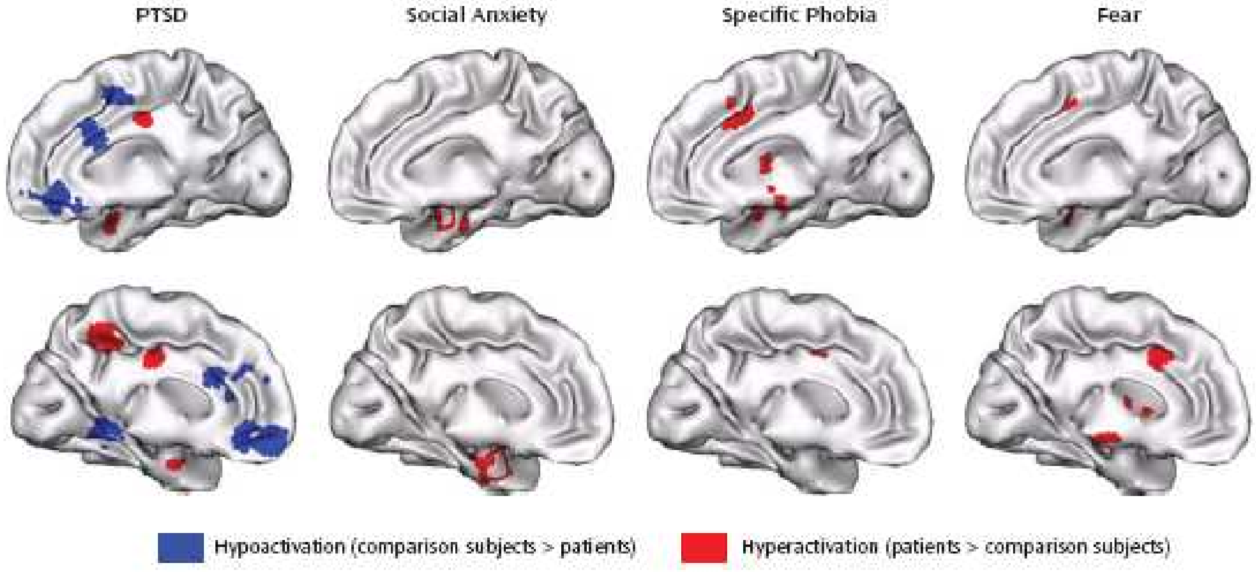


Figure M12: Hyper- and Hypoactivations related to different forms of anxiety disorders as compared to fear conditioning *(original in [165]. For more information please refer to text in Chapter IV.4.2.*

IV.4.3. Neuroticism: Punishment Sensitivity

Neuroticism is defined as *"… a pervasive sensitivity to negative or punishment cues in the environment …"* [169], and constitutes a common feature of depressive as well as anxiety disorders, associated with self-reported negative affective states [170] and increased readiness to assess situations as threatening [171].

Accordingly, neuroticism has been found to positively correlate with amygdala and rostral ACC activation during emotional conflict, especially in the anxious but not the depressive form [172]. This was interpreted by either greater emotional arousal, or increased vigilance / attention toward emotional conflict. Another study found a positive correlation between neuroticism and amygdala activity during the exposure to unpleasant pictures as well as scenes of moral violations [173]. In addition, the same study revealed increased DLPFC activity as a function of neuroticism during cognitive regulation task requiring participants to decrease the impact of unpleasant pictures.

These data were interpreted as stronger emotional responses to negative stimuli in combination with increased difficulties to regulate / decrease the latter, proportional to the neuroticism score of the participants. Moreover, it was shown that neuroticism positively correlated with neural activity in the amygdala and hippocampus during fear learning, which is in accordance with increased sensitivity to punishment for this personality trait [174]. Finally, a very recent experiment demonstrated that neuroticism was positively correlated with the amount of time subjects spent looking at fearful faces, which was attributed to trait congruent behavior involving an attentional bias towards highly arousing (threatening) stimuli [175].

Even though the proposed generative mechanism of neuroticism – punishment sensitivity – is thought to slightly differ from the one underlying anxiety – fear sensitivity (see Chapter IV.4.2.) –, both conditions produce comparable neural patterns of activation. This is also mirrored by the fact that neuroticism is described as a common feature of anxiety disorders. As a consequence, it might be very difficult to dissociate between these two aspects of personality on a neural level.

IV.4.4. Intro- / Extraversion: Reward Sensitivity

As opposed to anxiety and neuroticism (see Chapters IV.4.2 and IV.4.3.), extraversion is commonly associated with feelings of subjective well-being, in combination with more frequent laughter, smiling, and the propensity to tell jokes [176]. Accordingly, extraversion was hypothesized to display a generally increased sensitivity to reward and/or reward incentives [174]. This notion was corroborated by functional neuroimaging findings showing that amygdala activation was positively correlated with extraversion during the perception of happy, but not fearful facial expressions [116], and that the OFC and VLPFC displayed increased activity during humor appreciation as a function of extraversion [176]. However, the latter study also found an intriguing inverse relation between amygdala activation during humor appreciation and extraversion, a pattern that was confirmed by a recent investigation of observational reward-, but not fear-learning in the context of extraversion [174]. Thus, the relation between extraversion and general reward sensitivity has been called into question. The resolution of this discrepancy might lie in the different nature of the investigated processes: social reward associated with happy faces acquired through operant conditioning and thus contingent on behavior, *versus* processes learnt through classical conditioning like humor appreciation and the acquisition of reward contingencies independent of behavior [174].

In sum, extraversion as a personality trait might provide an interesting perspective to investigate individual differences related to positive emotional states and scenarios involving reward, even though some of the observed effects might be difficult to disentangle from basic influences of arousal. Moreover, no effects of extraversion on emotional memory and processes related to emotion regulation have been described up to date.

IV.4.5. Rejection Sensitivity

Rejection sensitivity is defined *as "… the tendency to anxiously expect, readily perceive, and intensely react to rejection …"*, and is supposed to result from a natural learning process through rejection experiences with caregivers and significant others during childhood [177]. Accordingly, increased dorsal ACC activation – implicated in pain perception and social rejection (see Chapter III) – has been found in response to disapproving facial expressions and this effect positively correlated with rejection sensitivity [178]. Another study reported increased prefrontal cortical activations in low versus high rejection sensitive subjects during the exposure to rejection-themed images, meaning that low rejection-sensitive participants could exert more cognitive control over such scenarios than high-sensitive subjects [177].

Interestingly, despite the fact that rejection sensitivity behavior is conceptualized to rely on similar mechanisms as the anxious and avoidant attachment style described by Mary Ainsworth and John Bowlby's attachment theory (see Chapter IV.4.6. below for theory and following chapters for experimental evidence), rejection sensitivity is nonetheless thought to have unique predictive utility. The same claim is also made regarding anxiety, neuroticism, and intro- / extraversion scores [177]. This notion is corroborated by functional neuro-imaging findings showing correlations between ACC activation and rejection sensitivity, suggesting increased pain and associated personal distress during the experience of social rejection scenarios, but no such relations for activity in the amygdala, which is more commonly related to threat or punishment sensitivity as a function of anxiety and/or neuroticism (see above).

IV.4.6. Adult Attachment Style

For the present thesis, adult attachment style (AAS) was selected as a psychological model of choice in order to investigate the modulatory role of *internal* factors representing *individual differences* in social perception, in relation with personal social motives and goals. Therefore, this concept will be introduced in more detail than the previous personality measures (see below). In addition, some reasons for the selection of AAS over other measures of anxiety, neuroticism, extraversion, and rejection sensitivity will be provided at the end of this chapter.

4.6.1. Origins

Attachment theory [179-181] has been developed by Mary Ainsworth and John Bowlby in the late 1970ies and early 1980ies, based on the so called “*strange situation*” paradigm. In the latter, a child is allowed to play for 20 minutes in a laboratory room while his/her primary attachment figure (i.e. parent) and strangers enter and leave. Thereby, the child experiences the following situations:

1) Child and Parent enter the room; 2) child and parent are left alone. The child explores the room, the parent does not participate; 3) a stranger enters the room, talks to the parent, and approaches the child. The parent inconspicuously leaves the room; 4) *first separation episode*: stranger is alone with child; 5) *first reunion episode*: parent enters room and comforts the child while the stranger leaves the room. The parent then leaves the room as well; 6) *second separation* episode: Child is first alone. Then, the stranger enters the room again and is alone with the child; 7) *second reunion episode*: parent enters room and comforts the child. The stranger leaves.

By measuring 1) the amount of time the child explores / plays throughout the 20 minutes; 2) the child’s reaction to the departure and return of its caregiver; and 3) the child’s behavior towards the stranger, the child can be categorized into three possible groups of “attachment style”, according to the following characteristics:

- **Secure Attachment Style (SAS):** The child plays / explores freely in the presence of the mother, it is not afraid of strangers, gets upset when the mother leaves, and is happy to see her come back.

- **Avoidant Attachment Style (AV):** The child does not play / explore a lot and does not show much emotional reaction, regardless of who is present in the room. He/she tends to avoid or even ignore the mother, even when she returns. The child does not treat strangers much differently from the mother.

- **Anxious Attachment Style (AX):** The child is anxious during playing / exploration and of strangers, even when the mother is present. When the mother leaves, he/she becomes extremely distressed. When the mother returns, the child displays ambivalent behavior by wanting to stay close but being resentful at the same time.

Later on, a fourth attachment style, namely disorganized/disoriented was also proposed, which included features of both avoidant and anxious attachment style. However, the latter is not relevant for the thesis at hand (see below) and will thus not be further discussed.

4.6.2. Translation from Children into Adults

During the last decades, attachment theory has been extended from mother-child relationships to adult romantic relationships [182] and even beyond [183, 184], serving today as a more general model of social interactions.

A central part of the adult attachment theory (ATT) is the so called "*secure base script*", which described the most adaptive coping mechanisms used in stressful situations by SAS individuals. This goes along the following lines: *"… If I encounter an obstacle and/or become distressed, I can approach a significant other for help; he/she is likely to be available and supportive; I will experience relief and comfort as a result of proximity to this person; I can then return to other activities …"* [185].

Once this "*secure base script*" is disrupted, the attachment system becomes continuously activated, leading to so-called "*secondary attachment strategies*". The latter can be understood as either "fight" or "protest" reactions leading to the attachment system *hyperactivation* as a hallmark of AX style, or considered as more defensive "flight" responses with an attachment system *deactivation* that is linked with AV style. The most important features of these hyper- and deactivating strategies in relation with AX and AV are depicted in Figure M13 below. For a more thorough review, readers can refer to Mikulincer & Shaver [185].

4.6.3. Measuring Adult Attachment Style

In the context of the standard behavioral psychological assessment, AAS has normally been measured with self-report questionnaires. Initially, the latter included separate attachment categories or styles, ranging from the three categories proposed by Ainsworth (see above) to four [183] or even five [186, 187] different attachment styles.

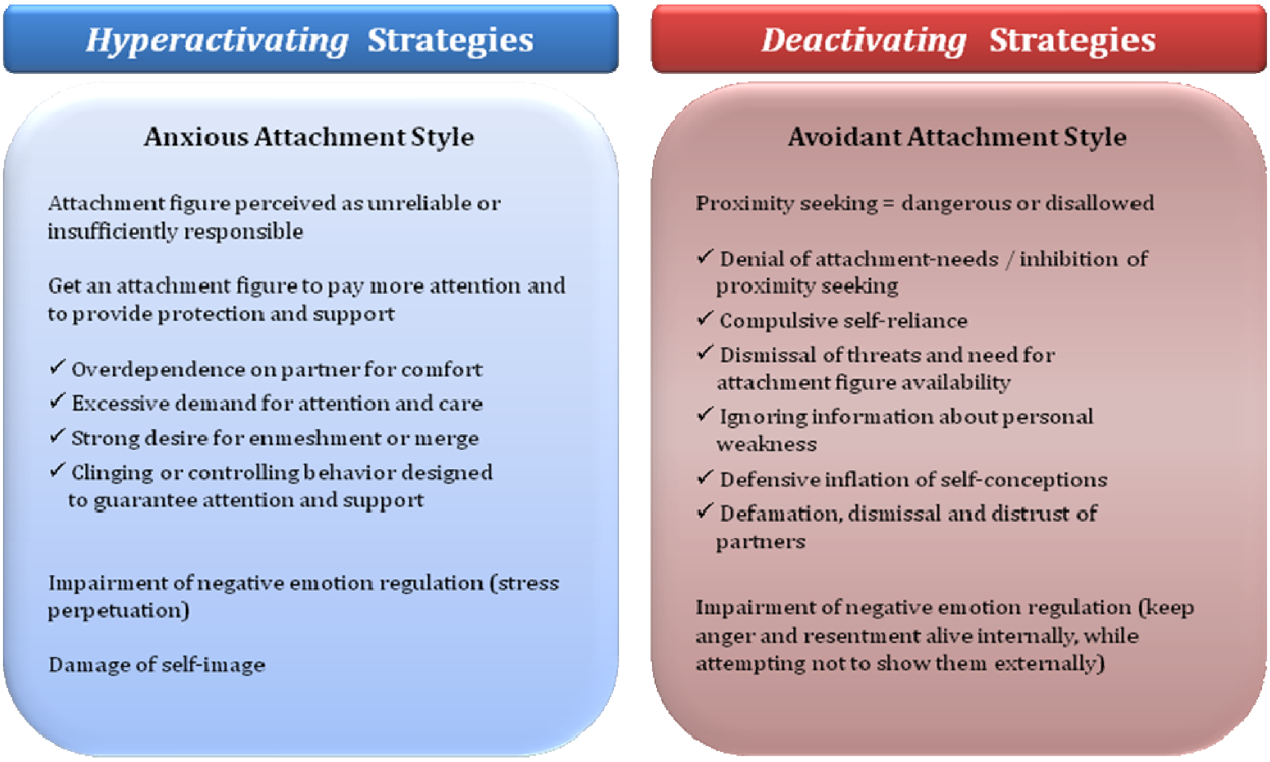
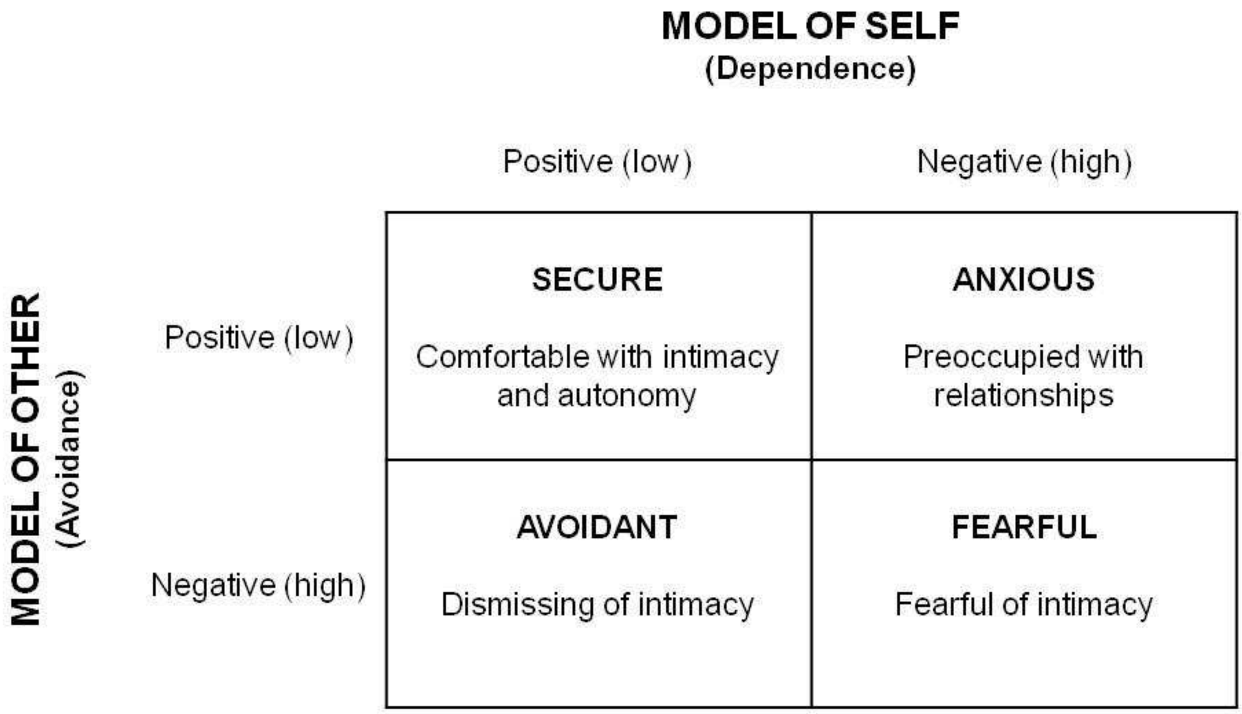


Figure M13: Most important hallmarks of hyperactivating strategies underlying AX, and deactivating strategies linked with AV *(adopted from [185]). Whereas AX is linked with hyper-vigilance or hyper-attention to signals of social reproach or threat, AV in turn is characterized by social distance and even the dismissal of threatening, but also positive social information. Both attachment styles are associated with emotion regulation deficits [185].*

An important point regarding the categorical AAS models, and particularly the one including four different categories, is the fact that AAS can also be conceptualized along two dimensions describing either the self in relation to dependence or the other as a function of avoidance (see Figure M14 below). Thereby, avoidance represents the *"… avoidance of intimacy reflecting the degree to which people avoid close contact with others as a result of their expectations of aversive consequences …"*. In turn, dependence is used in the context of a *"… strong dependency on others to maintain a positive self-regard …"* [183]. Accordingly, AX is characterized by a negative model of self and a positive model of others, whereas AV includes a positive self but a negative other model (see also Chapter V.2.4.).



MODEL OF SELF
(Dependence)

MODEL OF OTHER
(Avoidance)

	Positive (low)	Negative (high)
Positive (low)	**SECURE** Comfortable with intimacy and autonomy	**ANXIOUS** Preoccupied with relationships
Negative (high)	**AVOIDANT** Dismissing of intimacy	**FEARFUL** Fearful of intimacy

Figure M14: The four-category model of AAS as proposed by Bartholomew *[183]. Secure individuals are thought to have positive models of both themselves and others. Whereas anxious persons are conceptualized with negative self- but positive other-model, avoidant persons are described with positive self- but negative other-attributes. The fearful category is modeled as the opposite of secure attachment style.*

More recently, therefore, AAS has been mapped in a two-dimensional space made up by continuous measuring axes of AV and AX, with SAS corresponding to the area were both AV and AX are low, because such models were found to have the best validity [188]. In the present doctoral thesis, for reasons of data specificity, only the three classic categories of SAS (low AV *and* AX), AV, and AX were assessed, thereby leaving out the "*fearful*" category corresponding to the space were AV *and* AX are both high (called AVAX).

What is concerning the discriminant validity of AAS as measured by self-report questionnaires, it has been shown in various studies that the AX and AV constructs rarely display more than 25% shared variance with other with personality constructs derived from other theoretical or descriptive frameworks. This holds true for AX in association with neuroticism, trait anxiety, threat appraisal, or rejection sensitivity; and also for AV in relation with defensiveness, coping by distancing, or cognitive openness [185]. Nonetheless, as it is not uncommon to find positive correlations particularly between measures of AX and general anxiety (i.e. with the help of the STAI-T – see above), some caution is advised before making any interpretations regarding AX and other forms of anxiety.

4.6.4. Adult Attachment Style in Social Neuroscience

During the last decades, behavioral research on AAS has produced a very rich body of data. However, many of the conducted studies were purely observational and psychologically inspired. Therefore, not much was known about the neural correlates of AAS at the time the experimental work of the present thesis was carried out.

Yet, a few previous experiments used stimuli employed in social neuroscience research, such as emotional words, facial expressions, or movie-clips [189-193]. The latter have demonstrated that AAS influences processes related to vigilance and attention, including detection of the onset and offset of emotions in faces of strangers. In addition, it has been shown that AAS modulates the subjective experience of intensity, linking AX with increased arousal to negative images but AV with decreased arousal to positive images. In addition, one fMRI study [194] has revealed that AAS also modulates emotion processing, in particular during the suppression of negative relationship-related thoughts. This is in line with an experiment demonstrating that the avoidance of any social emotional involvement is a hallmark of AV, presumably linked with conscious cognitive top-down control suppressing the initial automatic emotional reaction [195]. For more details regarding the abovementioned experiments, please refer to Chapter V.2.2.

4.6.5. Adult Attachment Style as Psychological Concept of Choice

The experiments conducted as a part of the present doctoral thesis used AAS as a measure of key interest to assess individual differences in emotion perception for several reasons.

Firstly, the selection of a psychological model specifically assessing social personality traits related to social motives and goals had a great importance. In this regard, AAS seemed to be a valuable choice, because it has repeatedly been shown to rather selectively reflect the social dimension of behavior as well as mental processes, and this in a rather stable manner over the lifespan [185, 188]. Therefore, any observed effects of AAS on social perception should be readily differentiable from other measures normally used in social psychology to explain individual differences, i.e. anxiety, neuroticism etc. This also comes from the fact that AAS uses different theoretical constructs than the latter personality measures, such as vigilance, hyper- *versus* hypo-activity, mental working models linked with emotion regulation, as well as self- *versus* other dimensions.

Second, AAS includes predictions for different emotion processing steps such as encoding, retrieval, and regulation. It may therefore represent a more comprehensive approach than, say, extraversion or rejection sensitivity which only make predictions regarding emotion experience. Moreover, AV and AX are characterized by more than just one particular psychological trait – i.e. AX includes sensitivity to rejection, but also makes predictions about self- and other-models –, allowing for a more detailed assessment of behavioral and neural effects.

Third, AAS can provide sound interpretations for many observed effects in terms of developmental history of a person and his/her social relationships. This theoretical background is therefore ideally suited to serve as a basis for interpreting functional anatomical differences observed in paradigms exploring the neural correlates of social (*versus* non-social) perception.

And finally, as already mentioned above, only very few studies have used AAS in combination with fMRI up to date. Many relations thus still remain unclear or even unknown, providing unique opportunities for future neuro-imaging research.

CHAPTER V

Main Questions and Aims of the Doctoral Thesis

V.1. Social Context

In Chapters IV.2.1. and IV.3., previous experiments investigating the impact of social context on social perception have been described. These included the modulation of social context by 1) positive and negative sentences preceding surprised faces (Chapter IV.2.1.) – and thus external factors directly changing the meaning of the same facial stimuli –; 2) intentionality (Chapter IV.3.1.); 3) the presence / absence of an audience (Chapter IV.3.2.); and 4) a more basic social *versus* non-social stimulus content distinction (Chapter IV.3.3.).

One main aim of this doctoral thesis was to further extend the knowledge on such contextual mechanisms and their neural correlates by using **different methods of context manipulation**.

V.1.1. Direct Influence of Social Context on Emotion Encoding

For the fMRI study N° 1a (see Chapter VI.1.), the basic idea of Kim and colleagues' (2004) experimental paradigm was assumed, but importantly extended: the facial expressions were chosen to display emotions of happiness and anger, as opposed to surprise in the original experiment. Because surprise is an intrinsically ambiguous emotion, there are no specific neural correlates representing its encoding in the brain. This is different for happiness and anger, which have been previously linked with reward- and threat- or fear-related activity, respectively (see Chapter III). By using these basic positive and negative emotions, there was the possibility to not only identify brain areas representing the main effects of emotion (encoding for valence and/or arousal), but also the interaction of emotion and context (feedback – see below).

Moreover, the positive and negative sentences providing the different social contexts in Kim and colleagues' (2004) study modulated the properties of the facial displays independent of any subject involvement. This means that the social context only indirectly changed the way how the subjects perceived the surprised faces, along the lines: "*I learned that he just found $500, therefore I assume that he must be happily surprised*". This was changed in fMRI study N° 1a by integrating the social context as feedback in a pseudo-interactive social game, thereby making it personally relevant for the participants. Such an approach was inspired by the component process model / appraisal model of emotion (see Chapter I.3.1. and [37, 69]) and the fact that emotion encoding related to social perception was found to rely to a big extend on the significance of the stimuli for the perceiver, particularly regarding neural encoding in the amygdala [69]. This was also an important for considerations related to individual differences (see Chapter V.2.)

Regarding the above-mentioned extensions of Kim and Colleague's (2004) experiment, the following main question was addressed with study N° 1a:

- **Are happy (smiling) and angry faces always encoded in the same way, or does personally relevant social context influence their neural correlates?**

V.1.2. Direct Influence of Social Context on Familiar Face Recognition

fMRI study N° 1b investigated the representation of any learned emotion x social context (feedback) interaction induced during fMRI study N° 1a by means of a familiar face recognition paradigm (see Chapter VI.2.). Thereby, two main improvements in the experimental design as compared to previous studies on the same topic were made.

Firstly, fMRI study N° 1b used face identities that were unknown to the subjects prior to the experiment and only created a sense of familiarity for these through the pseudo-interactive social game used for study fMRI N° 1a. This is different from i.e. work by Gobbini, Leibenluft and colleagues [196, 197] that used personally familiar (romantic partners or own children) and famous (i.e. Hollywood actors) faces to assess any familiar face recognition effects underlying their model of involved processes and brain areas (see Chapter III.1 and [33]). Because personally familiar and famous faces are prone to pre-exposure effects inducing unwanted data variability, it was considered important to introduce the face identities as strangers at the first place, and to control for the social learning factors by keeping the social context constant for all subjects.

Second, study fMRI N° 1b included positive as well as negative person impressions induced through the pseudo-interactive social game feedback paired with either happy (smiling) or angry emotional faces during study fMRI N° 1a. All but one [198] previous experiments involving familiar face recognition only looked at positive impressions [154, 196, 197], thereby limiting the possibility to directly compare activity within one brain area – i.e. the amygdala – between people perceived as friendly *versus* opposing.

In addition, as already applicable for fMRI study N° 1a (see above), familiar face recognition in study fMRI N° 1b also relied on social characteristics that were personally relevant for the subjects, and not only indirectly related as in any other previous experiment (even the study by Todorov and colleagues [2007] used sentences introducing social context that described personality traits and/or behaviors of the depicted face identities and not the study participants themselves).

According to these considerations, the following main questions were addressed with fMRI study N° 1b:

- **What are the neural correlates of recognizing a previously stranger but now familiar face that has been encountered beforehand within either a positive (friend) or a negative (foe) personally relevant social context?**

- **Which interactions, positive (friend) or negative (foe), are more readily retrieved during familiar face recognition, and are such person impressions stored implicitly or explicitly in the human brain?**

V.1.3. Social *versus* Non-Social Distinction during Emotion *Regulation*

For the fMRI study N° 2a (see Chapter VI.4.), a basic social *versus* non-social stimulus content distinction was applied to investigate differences in emotion perception and particularly regulation. The main motivation for such an approach was the fact that no other fMRI study has assessed the differences in affect regulation induced by social *versus* non-social scenes up to date. Even though including both of the latter stimulus categories, previous experiments only differentiated for likeability (positive *versus* negative valence) and/or intensity (low *versus* high arousal) values [87, 92, 140, 141]. This was considered problematic because other research has shown qualitative differences between emotional reactions to social *versus* non-social images [157]. In addition, as difficulties in emotion regulation are often entailing mental disorders such as anxiety, and anxiety research distinguishes between social *versus* non-social forms of this condition (see Chapter IV.4.2.), a separation of the latter factors seemed important to advance in the research field of emotion regulation.

Even though not directly related to the social *versus* non-social distinction, fMRI study N° 2a included another important improvement in comparison to other emotion regulation studies: it simultaneously tested for neural correlates of re-appraisal as well as suppression. Despite the fact that emotion regulation theory [84, 85] emphasizes the difference between these two cognitive *versus* behavioral emotion regulation strategies, most previous studies included only one of them in their experimental design, which was mostly re-appraisal [87, 140, 141]. Only very few studies have investigated suppression [91], and only one study up to date has simultaneously included both behavioral as well as cognitive emotion regulations strategies [92].

In light of these points, fMRI study N° 2a had the following question as main point of focus:

- **How do re-appraisal and suppression affect the regulation of emotions as a function of social content as well as stimulus valence?**

V.2. Individual Differences

Chapters IV.2.2. and IV.4 provided basic theoretical knowledge and gave examples of previous experiments using personality traits and specific psychological conditions – like anxiety disorders – to assess individual differences in stimulus processing as a function of social motives and goals. These included 1) generally increased negative processing of surprised faces (see Chapter IV.2.2); 2) increased fear sensitivity underlying anxiety (see Chapter IV.4.2.); 3) increased punishment sensitivity as determining factor in neuroticism (see Chapter IV.4.3.); 4) altered reward sensitivity related to intra-/extraversion (see Chapter IV.4.4.); 5) increased rejection sensitivity (see Chapter IV.4.5.); and 5) adult attachment style including avoidant and anxious types (see Chapter IV.4.6.).

Another main aim of this doctoral thesis was to further specify the **mechanisms through which individual differences modulate social perception and familiar face recognition, as well as emotion regulation** induced by social (*versus* non-social) emotional scenes. Thereby, adult attachment style (AAS) was selected as the psychological concept of choice, because the latter includes different types of personality traits and provides developmental causes for their emergence (see Chapter IV.4.6.5. for more details), which can be easily translated into social motives and goals. Moreover, only one study by the starting point of this doctoral thesis has reported neural correlates of AAS in similar experimental settings [194].

V.2.1. Modulation of the Emotion x Social Context Interaction by Individual Differences

fMRI study N°1a (see Chapter IV.1.) not only assessed the modulatory role of social context on the perception of happy (smiling) and angry faces (see Chapter V.1.1.), but also tested how the latter emotion x social context (feedback) interaction was affected by individual differences in personality, and in particular AAS. Thereby, a distinction was made between avoidant and anxious types representing insecurity, and the secure type of AAS.

The main question of fMRI study N° 1a in the context of AAS therefore was:

- **How do individual differences in AAS modulate the perception of happy (friends) *versus* angry (foes) faces?**

V.2.2. Modulation of Social *versus* Non-Social Perception by Individual Differences

In the behavioral study (see Chapter IV.3.) that served as a basis for fMRI studies N° 2a and 2b, it was tested how AAS influenced the subjective judgment of social *versus* non-social images according to three different scales: 1) likeability (positive or negative stimulus valence); 2) intensity (low versus high arousal); and 3) controllability (the subjective presence or absence of control over any emotion elicited by the exposure to a social or non-social emotional scene).

Previous experiments on the role of AAS in social perception have mainly focused on vigilance to or attentive processing of words or emotional facial expressions [189-192]. These investigations revealed that highly anxiously attached subjects were more prone to detect emotions in facial expressions and had quicker access to attachment-related material in a lexical decision task. In addition, it was found that a combination of attachment avoidance and anxiety was associated with reduced attention for angry faces. Even though such results are interesting, they do not provide any clues about the underlying effects, i.e. whether anxious participants experienced emotional faces as more intense and/or less controllable, leading to better detection rates. One very recent study [193] has partially addressed this question by showing that anxiously attached participants experienced negative emotions of fear and sadness as more, but avoidantly attached participants positive emotions as less arousing while viewing emotional video-clips. The latter study however did not show any effects on valence and/or controllability, and no experiment included a distinction between social *versus* non-social, and therefore attachment-related *versus* attachment-unrelated scenarios.

The main question for the behavioral study therefore was:

- **How does AAS modulate the subjective attribution of likeability, intensity and controllability to positive and negative social *versus* non-social emotional scenes?**

V.2.3. Modulation of Social *versus* Non-Social Emotion Regulation by Individual Differences

For the fMRI Study N° 2b (see Chapter VI.5.), data from fMRI study N° 2a was re-analyzed to include measures of individual differences representing social motives and goals. Once more, the psychological concept of choice was AAS.

One previous fMRI experiment [194] assessed individual differences in the ability to suppress negative thoughts and thus partially covered this area of interest of the present doctoral thesis. The authors described that anxiously attached participants reacted more strongly to thoughts of loss with corresponding increased activity in anterior temporal pole related to sadness, whereas they showed impairments in the regulation – and particularly suppression – of negative emotions by less activity in prefrontal top-down control circuits. Avoidantly attached subjects were also found to display less deactivation as a function of suppression, namely in subcallosal cingulate and lateral prefrontal cortex, which was likewies attributed to less efficient emotion regulation skills. However, the experimental design of Gillath and colleagues (2005) did not include re-appraisal as emotion regulation strategy, and the stimuli were all relationship-related and thus social in addition to being selectively negative. fMRI study N° 2b was therefore designed to include cognitive and behavioral emotion regulation strategies – and a "natural" emotion experience condition –, social and non-social emotional scenes, as well as positively and negatively valenced social scenarios.

Accordingly, the main question of fMRI study N° 2b was:

- **How does AAS modulate the differential regulation (re-appraisal versus suppression) of social positive and negative emotions?**

V.2.4. Modulation of Self *versus* Other Perception by Individual Differences

fMRI study N° 3 (see Chapter VI.6.) investigated the modulatory role of AAS on self- versus other-perception (see Chapter IV.4.6.3.). Thereby, the authors made use of Bartholomew's [183, 184] theoretical concept of self- (dependence) and other- (avoidance) perception, suggesting that AV is linked with a positive self- but a negative other-model, whereas AX is associated with a positive other- but a negative self-model.

During the last 25 years after the first description of the four-category model underlying the self- versus other-perception in AAS [183], more insight has been gained regarding the dynamics and complexities of such insecure people's mental representations. Thereby, it has been shown that the positive self-model associated with AV is often linked with *"… defensive self-enhancement involving harsh self-criticism and perfectionist self-standards that can sometimes provoke self-doubts and demoralization …"* [185]. In addition, it was found that the positive other-model normally related to AX where a partner is seen as a *"… badly needed personal savior …"* [185] is rather unstable and often intruded by ambivalent or even negative associations. Therefore, it seems easier to draw sound conclusions on the other-model in AV and the self-model in AX than the two remaining categories, which have to be approached carefully.

Very little is known about the neural correlates of self- versus other-perception in relation with AAS, because no fMRI study has investigated such a relation up to date. fMRI Study N° 3 was therefore of an exploratory nature and had the main question:

- **How does AAS influence the neural correlates of self- *versus* other-perception?**

V.2.5. Selectivity of Results regarding Adult Attachment Style

Finally, a more general question could be addressed related to the specificity of any effects found by using AAS as a psychological model in the context of fMRI investigation of social emotional processes.

- **How specifically does AAS influence the neural correlates of social perception?**

CHAPTER VI

fMRI and Behavioral Studies

VI.1. fMRI Study N°1a

Published in: PLoS ONE 3 (8): e2868 (2008)

fMRI study N° 1a aimed at investigating the direct influence of external social context on emotion encoding and the modulation of the latter process by internal social motives and goals represented by individual differences in adult attachment style (AAS; either anxious [AX], avoidant [AV], or secure [SAS]). Thereby, the main questions were:

- **Are happy (smiling) and angry faces always encoded in the same way, or does personally relevant social context influence their neural correlates?**

- **How do individual differences in AAS modulate the perception of happy (smiling – friends) *versus* angry (foes) faces?**

Like in the original experiment by Kim and colleagues [41], the external social context of shown facial expressions was directly manipulated by additional written information. However, the experimental design was importantly improved by using 1) happy (smiling) and angry facial expressions as opposed to surprise; as well as 2) personally relevant social context by means of feedback in a pseudo-interactive game (i.e. winning *versus* losing) in contrary to positive or negative sentences not directly concerning the participants (see Chapter V.1.1. for more details).

The results revealed that the processing of emotional facial expressions was crucially influenced by the social context provided by the game feedback. Whereas no selective brain activations were detected for general emotion processing regardless of concomitant feedback (i.e. smiling *versus* angry faces), the four distinct emotion + feedback combinations produced selective activity in different limbic and cortical brain areas. This shows that a given emotional facial display does not always convey the same social message and thus leaves a differential signature in the brain as a function of external context.

Moreover, activity in reward- as well as threat-related brain areas to specific emotion + feedback combinations was found to vary by the degree of AV and AX of the study participants. Whereas high avoidant subjects displayed less reactivity to smiling faces during winning representing social reward in the ventral striatum and ventral tegmental area, high anxious subjects showed increased amygdala activation to angry faces during losing representing social punishment. These findings demonstrate that AAS differentially modulates social perception according to hyper- and de-activating strategies as hallmarks of AX and AV, respectively.

Individual attachment style modulates human amygdala and striatum activation during social appraisal

Pascal Vrtička (1,2), Frédéric Andersson (2), Didier Grandjean (1,3),
David Sander (1,3) & Patrik Vuilleumier (1,2)

(1) Swiss National Center for Affective Sciences, University of Geneva
(2) Laboratory for Neurology & Imaging of Cognition, Department of Neurosciences
& Clinic of Neurology, University Medical Center of Geneva
(3) Department of Psychology, FPSE, University of Geneva

ABSTRACT

Adult attachment style refers to individual personality traits that strongly influence emotional bonds and reactions to social partners. Behavioral research has shown that adult attachment style reflects profound differences in sensitivity to social signals of support or conflict, but the neural substrates underlying such differences remain unsettled. Using functional magnetic resonance imaging (fMRI), we examined how the three classic prototypes of attachment style (secure, avoidant, anxious) modulate brain responses to facial expressions conveying either positive or negative feedback about task performance (either supportive or hostile) in a social game context. Activation of striatum and ventral tegmental area was enhanced to positive feedback signaled by a smiling face, but this was reduced in participants with avoidant attachment, indicating relative impassiveness to social reward. Conversely, a left amygdala response was evoked by angry faces associated with negative feedback, and correlated positively with anxious attachment, suggesting an increased sensitivity to social punishment. Secure attachment showed mirror effects in striatum and amygdala, but no other specific correlate. These results reveal a critical role for brain systems implicated in reward and threat processing in the biological underpinnings of adult attachment style, and provide new support to psychological models that have postulated two separate affective dimensions to explain these individual differences, centered on the ventral striatum and amygdala circuits, respectively. These findings also demonstrate that brain responses to face expressions are not driven by facial features alone but determined by the personal significance of expressions in current social context. By linking fundamental psychosocial dimensions of adult attachment with brain function, our results do not only corroborate their biological bases but also help understand their impact on behavior.

CORRESPONDING AUTHOR:

Pascal Vrtička
Swiss Center for Affective Sciences
7, rue des Battoirs, 1205 Geneva, Switzerland
Tel: +41 – (0)22 379 9824 Fax: +41 – (0)22 379 9844
Email: pascal.vrticka@ unige.ch

INTRODUCTION

Since its description four decades ago, attachment theory [1,2] has become one of the most important frameworks for understanding affect regulation in social interactions [3,4]. Initially grounded on child-mother relationships [1,2], the functions of attachment were first related to the regulation of proximity-seeking behavior, with the goal to obtain protection and care from another person (an attachment figure). Because of a consistent pattern of engagement of attachment processes during a range of situations from infancy to adulthood, behavioral researchers have suggested that attachment models may become part of general interpersonal schemata for an individual, which will support social development and influence thoughts, feelings, and behavior throughout the lifespan [3]. Thus, attachment theory has been extended to adults to describe affective responses in the context of various relationships, particularly with romantic partners or close friends [5]. Moreover, subsequent work demonstrated that attachment style in adults may also predict behaviors and judgments regarding unfamiliar persons [6,7], as well as social and emotional processes in various tasks [3]. These findings are consistent with the idea that people use general schemata of interpersonal relationships at different levels of representations from parents and close partners through to less familiar persons [8]. Therefore, although the exact links with developmental aspects of the attachment system are still unclear [9], it is generally thought that attachment style in adults entails fundamental individual biases that can influence how people perceive and respond to social information in a wide range of relationship contexts.

Following earlier studies on children and parents, classic models have distinguished between three main prototypes of attachment style in adults: secure, anxious, and avoidant [5]; whereas subsequent work suggested the existence of four [10] or even five [11,12] distinct styles. More recent models proposed instead that these different styles might be mapped on two basic dimensions defined by orthogonal axes of anxiety and avoidance, with the secure style corresponding to both low anxiety and low avoidance [10]. Individuals with an anxious attachment style (AXS) tend to perceive others as unresponsive or inconsistent,

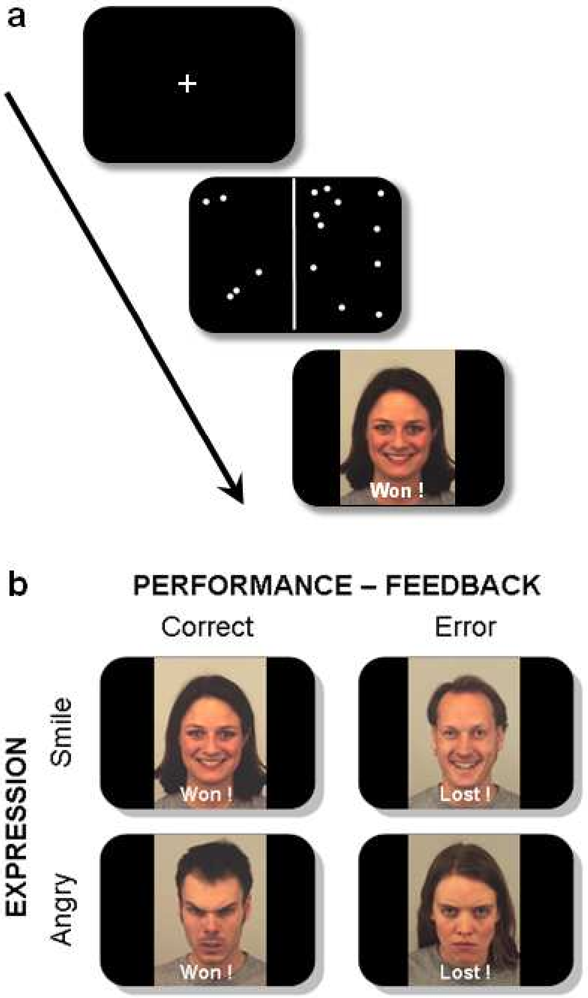


***Figure 1: a) Illustration of the paradigm.** Participants first saw a central fixation cross, followed by the dot-counting task (0.5 sec), where they had to indicate which side of the screen contained more dots (right vs left). Following each response, a visual feedback was shown (1.5 sec), composed of a word together with a face. b) Illustration of the four different feedback conditions: two socially “congruent” (Smiling Face on WON trial, Angry Face on LOST trials) and two socially “incongruent” (Smiling Face on LOST trial, Angry Face on WON trial) combinations were possible. Four different face identities (2 female and 2 male) were used in each of these 4 conditions. See Materials and Methods and SI for further details.*

worry about being rejected, and show heightened vigilance to signs of support or hostility; whereas individuals with an avoidant attachment style (AVS) prefer being distant and detached from others, report no need for close relationships, and tend to distrust affective signals from others. By contrast, individuals with a secure attachment style (SAS) report positive and trustful interactions with others. More generally, it has been shown that attachment style may shape the way in which individuals appraise social information in a variety of conditions [6,7,13], including during interactions with strangers [8,14]. For instance, attachment styles can determine whether feedback messages given by partners are perceived as supportive or not [15], predict the experience of conflict in interpersonal contexts [16,17], and influence the perception of emotional expressions in pictures of unknown faces [6,7]. Adult attachment style is also related to individual differences in cooperativeness, reward dependence, and novelty seeking [11,18]. Based on behavioral studies, researchers have proposed that AXS might reflect higher states of vigilance and sensitivity to socially significant cues, while AVS might involve either deactivation or inhibition of affective responses to interpersonal situations [3,6,17]. However, although attachment theory has generated a rich body of research in social and clinical psychology, the neural bases of these individual differences remain unknown.

Here we developed a new fMRI paradigm to determine the influence of the three classic types of adult attachment style (secure, anxious, or avoidant) on appraisal of social cues in the human brain. We tested whether individual differences in attachment style are linked to relative decreases or increases in the activity of brain regions associated with social and emotional processing, and whether such effects might depend on the personal significance of social signals, by presenting faces with expressions that could be perceived as either friendly (supportive) vs unfriendly (unsupportive or hostile). Previous neuroimaging studies concerning attachment have focused on particular relationships such as maternal and romantic love, without considering individual attachment style [19,20], and reported that pictures of loved individuals deactivate the amygdala and activate the striatum, two brain regions critical for affective processing and reinforcement [21,22]. Other studies found activation of amygdala [23] and medial prefrontal cortex [24] to sentences or scenarios with attachment-related meaning, but did not examine the differential effects of classic attachment styles on the perception of social cues with different affective meanings. Since adult attachment can shape emotional responses to socially relevant signals, our study specifically aimed at comparing the influence of distinct attachment styles on the processing of negative and positive social stimuli in attachment-related contexts. We hypothesized that individual differences in attachment styles should modulate activation patterns in brain circuits known to mediate social perception and behavior, particularly in emotional limbic regions such as the amygdala, ventral striatum, and ventromedial prefrontal cortex. Alternatively, since attachment style can also influence the formation of “mental models” of others [10,25,26], it might primarily modulate the recruitment of higher-level cortical regions associated with mentalizing and theory of mind such as superior temporal sulcus (STS) and anterior cingulate cortex [27].

To generate context-specific appraisal of social signals in our study, participants saw faces with smiling or angry expressions in a pseudo-game context, while they underwent event-related fMRI scanning (see Methods). Because attachment style is critically related to the way people evaluate signs of alliance and opposition during social interactions [17], we systematically manipulated the social significance of these facial expressions to elicit a perception of either supportive or unsupportive partners. We took advantage of the fact that smiling or angry expressions can have very different meanings based on current context. For instance, a smile may be perceived as praising an accomplishment or mocking a failure. Likewise, angry expressions may signal reproach or frustration. By inducing specific social meanings for these facial expressions presented in different scenarios, we could test the hypothesis that attachment style might influence affective appraisal of social facial signals, as suggested by previous behavioral studies [6,7,15].

In accord with these predictions, our results show for the first time that adult attachment style modulates neural responses to the perceived social meaning of facial expressions in brain regions critically associated with affective processing and learning, namely ventral striatum and amygdala. Furthermore, we show that the two dimensions of avoidance and anxiety produce distinct effects in these two regions, and thus appear sufficient to account for the effects of secure attachment, in agreement with previous theoretical proposals [10].

RESULTS

Participants were presented with smiling or angry faces accompanying a feedback message about their current performance in a difficult perceptual task (Fig. 1a). On each trial, they first saw a visual array in which they had to judge the number of dots. Feedback was then displayed, consisting of a word indicating actual performance ("WON" or "LOST") together with a face (Fig. 1a). Critically, the face could have either a smiling or angry expression, and could appear on either a WON or LOST trial (half each). This resulted in four feedback types, with two "congruent" and two "incongruent" conditions (Fig. 1b): Smiling Face on WON trial (SF-W) or LOST trial (SF-L), Angry Face on LOST trial (AF-L) or WON trial (AF-W). Participants were told that these faces were from two different groups receiving points based on their performance, such that they could be perceived as either allied partners (SF-W and AF-L; congruent conditions) or opponents (SF-L and AF-W; incongruent conditions) in a virtual game context (see Methods). After fMRI scanning, a series of questionnaires was given to assess attachment style [28] as well as other affective traits and debriefing measures (see Methods). We used a standardized scale, the Adult Attachment Questionnaire (AAQ; [28]), which provides three scores for each individual, corresponding to the relative strength of each of the three classic attachment styles: two indices for avoidant (AVS) and anxious (AXS) attachment, and one global score for secure attachment (see Methods for more details). Combinations of the first two scores also provide a reliable measure along two separate dimensions of avoidance and anxiety [29]. Individual differences in attachment style did not influence performance on the dot counting task (accuracy and reaction times). Debriefing questionnaires after scanning indicated that participants were motivated by the task and reported genuine affective reactions to facial expressions seen in different feedback context (see Methods).

Main effects of feedback and expression

First, we examined brain responses to each feedback type, regardless of the concomitant face expression. Success feedback (all WON > LOST trials) activated a widespread network in bilateral basal ganglia, left orbitofrontal cortex (OFC), anterior cingulate, and dorsolateral prefrontal areas (see Table 1), consistent with a general role of these regions in monitoring outcomes and rewards [22,30]. Conversely, error feedback (LOST > WON) activated retrosplenial cortex and right insula (see Table 1), consistent with previous studies on evaluation of negative events [22]. We also examined brain responses to each emotional face expression (smiling > angry or vice versa), regardless of concomitant success feedback. However, the latter contrasts did not reveal any region showing only main effects of expression at standard statistical thresholds ($p< .001$).

Social meaning of congruent positive feedback: Responses to perceived support

Next, we examined the different activation patterns evoked by the same expression in different feedback contexts, focusing first on the congruent conditions. When comparing responses to smiling vs angry faces on WON trials (SF-W > AF-W), a condition corresponding to the perception of social support, we found selective increases in left ventral striatum (Fig. 2a) and left OFC (Table 1). These activations were not due to success alone, because performance feedback was positive (WON) in both conditions. No such increases were found for smiling faces paired with the negative (LOST) feedback (Fig. 2b). Hence, these responses reflected the social value of reward, rather than reward or facial expression only. A repeated-measure ANOVA on the average parameter estimates of activity (betas) extracted from left ventral striatum confirmed a significant interaction between face expression and performance success ($F_{1,15}=5.99$, $p=.027$).

We then examined the relation between left ventral striatum activity and attachment indices, using parameter estimates of activity for this cluster in each trial type. This showed a strong negative correlation between the magnitude of response to smiling faces with success feedback (SF-W) and the degree of avoidant attachment (AVS, $r= -.787$, $p=.001$; Fig. 2d), but no such relation for other conditions, including success feedback with angry faces (AF-W). No significant relation was found for OFC. Other attachment indices did not correlate with neural activity in any of these areas.

These results were confirmed by a whole-brain multiple regression analysis of the response elicited by smiling faces with success feedback (SF-W), using the three attachment scores as separate parametric factors in a single SPM design, in order to test for any voxels throughout the brain where activation in this social reward condition (SF-W) varied as a function of each attachment style. To rule out that any correlation with SF-W would be partly confounded by an inverse correlation with AF-W, our SPM regression analysis was performed on the contrast of SF-W versus the other three conditions (SF-W > others), rather than on the contrast of SF-W > AF-W as used above (but results of these two analyses were in fact similar). In addition to the left ventral striatum, this whole-brain regression analysis revealed a highly significant negative relation ($p< .001$) between AVS and activity in anterior insula and left midbrain, overlapping with the ventral tegmental area (VTA, Figs. 2ce and Table 2). Thus, higher AVS scores predicted lower activation in several brain regions associated with dopaminergic function and reward, including both ventral striatum and VTA [22,30].

Conversely, the same regression analysis also revealed that high scores on the SAS dimension correlated positively with ventral striatum and insula activity (see Table 2). However, there was no significant correlation between SAS and VTA. Finally, no correlation was found for AXS scores and other personality factors related to anxiety (BIS/BAS, STAI-T) or more general affective traits (PANAS). This

Table 1 **Main Contrasts**

Brain Area	*BA*	*Voxel*	*x*	*y*	*z*	*Z*
Won>Lost						
OFC left	11	36	-36	39	-15	4.87
Caudate right		5	18	21	18	3.8
Nucleus accumbens right		6	12	12	-6	3.55
Putamen left		5	-24	9	0	3.48
dlPFC right	46	100	30	18	42	4.77
dlPFC right	46	41	33	51	24	3.92
dlPFC left	45/46	12	-42	48	24	4.07
dlPFC left	9	56	-30	15	48	3.99
Dorsal ACC right	25	40	15	-6	48	3.92
Hippocampus right		34	27	-21	-24	4.1
Occipital cortex left	17	6	-15	-108	9	3.7
Angular gyrus right	39	13	45	-63	33	3.66
Lost>Won						
Retrosplenial cortex left	26	18	-3	-36	15	3.3
Supramarginal gyrus left	40	5	-63	-48	30	3.15
Insula right		5	42	-3	-3	2.81
Smiling > Angry Faces with Won Feedback (SF-W > AF-W)						
OFC left	11	7	-24	48	-3	4.36
Parietal cortex right	7	25	15	-63	63	3.79
ACC right	24	7	9	33	18	3.72
Supramarginal gyrus right	40	9	51	-42	57	3.66
Ventral striatum left		5	-24	18	-12	3.46
Supramarginal gyrus left	40	5	-45	-48	57	3.41
Occipital cortex right	18	9	3	-90	12	3.36
Angry > Smiling Faces with Lost Feedback (AF-L > SF-L)						
Amygdala left (dorso-medial)		27	-12	-3	-15	4.62
Post hippocampus left		14	-33	-33	-6	3.37
Insula right		6	39	0	-18	3.04
Angry > Smiling Faces with Won Feedback (AF-W > SF-W)						
Supramarginal gyrus left	40	14	-48	-42	36	3.92
STS right	21	12	48	-39	0	3.86
Inferior frontal gyrus left	44	22	-45	9	24	3.63
Smiling > Angry Faces with Lost Feedback (SF-L > AF-L)						
STS left	21	15	-48	-39	-6	4.2
Angular gyrus right	39	7	60	-54	36	3.71
Inferior frontal gyrus left	44	16	-54	21	33	3.61
Parietal cortex right	40	23	54	-57	48	3.52
Occipital cortex left	17	39	-15	-105	6	3.48

Table 1: *Brain areas activated in the main contrasts between conditions, listed with peak coordinates and best estimates of anatomical location. Coordinates are given in MNI space. Activation sites were determined on the basis of the average anatomical MRI images of our 16 subjects. BA = Brodmann's area, OFC = Orbitofrontal cortex, dlPFC = Dorsolateral prefrontal cortex, ACC = Anterior cingulate cortex, STS = Superior temporal sulcus. p< .001 for all clusters except for right insula p<..002.*

complementary correlation profile between AVS and SAS suggests that activation of reward-related regions to situations representing social support is associated with secure attachment style, whereas a lack of activation is associated with avoidant attachment.

Social meaning of congruent negative feedback: Responses to perceived reproach

Brain responses to angry expressions also differed as a function of feedback and attachment style. When comparing angry vs smiling faces in LOST trials (AF-L > SF-L), a condition meant to evoke signs of reproach or social punishment, we found significant activation in left dorsal amygdala (Fig. 3a), as well as left hippocampus and right insula (see Table 1). Amygdala activation was not due to incorrect performance alone, because error feedback (LOST) was similar in both conditions. In addition, no such increase was found for angry vs smiling faces in WON trials (Fig. 3b). Again, these responses reflected the social meaning of feedback, rather than loss or facial expression per se. Accordingly, a repeated-measure ANOVA on the average parameter estimates of activity (betas) from this amygdala cluster confirmed a significant interaction between face expression and feedback type ($F_{1,15}$=8.19, p=.012). When testing for an association with attachment indices, we found a specific negative relation between the dorsal amygdala response to angry faces with error feedback (AF-L) and the degree of secure attachment (SAS, r= -.487, p=.033), but no such relation for the three other feedback conditions.

Again, we confirmed this correlation by a whole-brain multiple regression analysis of activation to angry faces with error feedback (AF-L > others), using attachment scores as separate parametric factors in a single SPM design (similar results were found using the contrast AF-L > SF-L). In keeping with the above, we found highly significant effects in the left amygdala (Fig. 3c). This region showed not only a strong positive correlation with AXS, but also a negative correlation with SAS, selectively for the AF-L condition in both cases (p< .001, Figs. 3de and Table 2). Similar correlations with AXS and SAS were also found in the left medial thalamus. No relation was found for the AVS dimension in this condition. Again, these correlations were specific to these two attachment indices, but not found for more general anxiety or affective measures (STAI-T, BIS/BAS, PANAS). This complementary correlation profile between SAS and AXS suggests that secure attachment is related to low anxiety as well as low avoidance, rather than to a single distinctive pattern of brain responses.

Incongruent feedback trials: Responses to perceived social conflict

For completeness, we also examined brain responses to the two socially “incongruent” feedback conditions, corresponding to the perception of opponent faces (smiling on LOST trials, SF-L > AF-L; or angry on WON trials, AF-W > SF-W). These conditions elicited selective activations in the left and right superior temporal sulcus (STS), respectively (see Table 1). However, these increases in STS did not correlate with attachment traits.

At a lower threshold, we also found a selective activation in the rostral ventral anterior cingulate cortex (vACC, BA 32, xyz= 9 54 3, z= 2.50, p <.006) for angry opponent faces (AF-W > SF-W), consistent with previous reports that ACC might be involved in conditions of social rejection and conflict [31-33]. Moreover, a whole-brain multiple regression analysis with attachment indices revealed a positive correlation

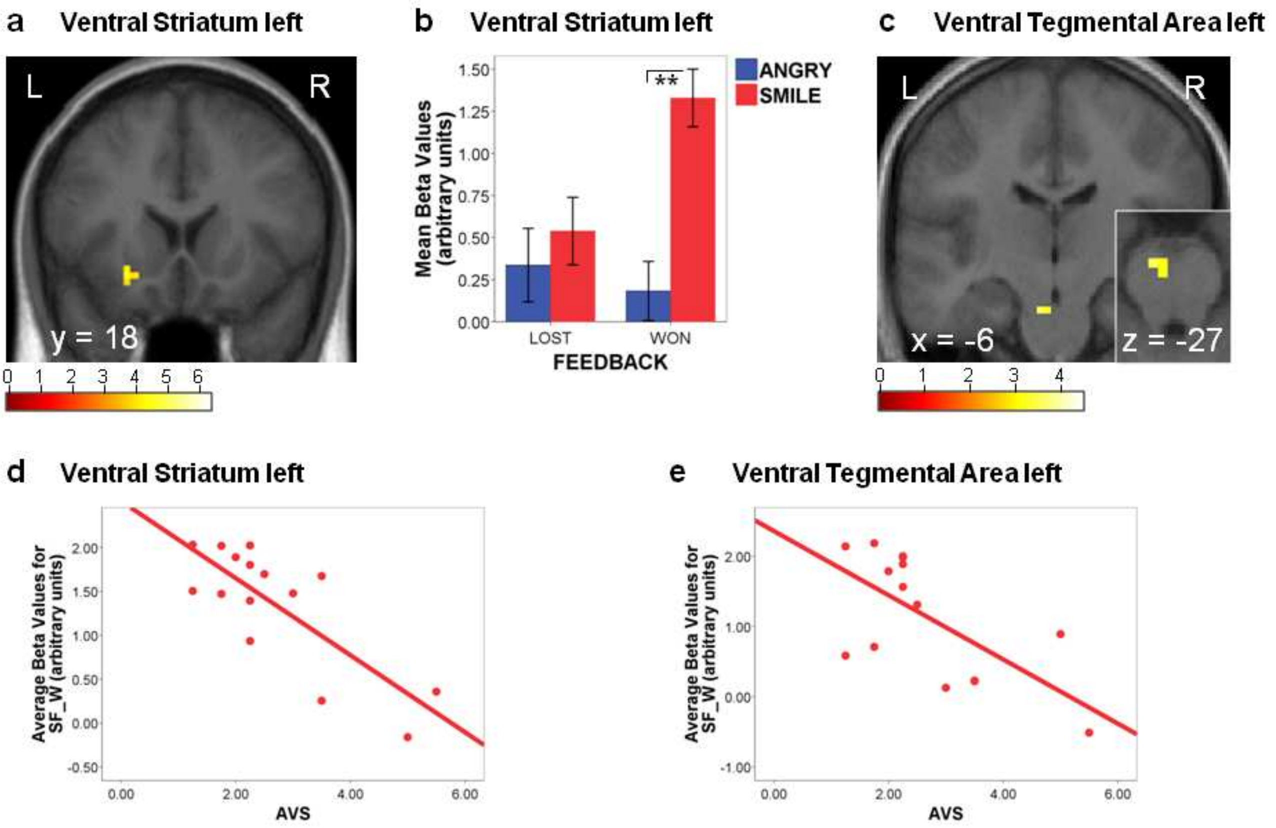


Figure 2:
a) Statistical parametric map for Smiling vs Angry expressions in success feedback context (contrast SF-W > AF-W), showing selective activation in left ventral striatum (xyz= -24 18 -12, z-score= 3.46, p< .001). **b)** Activation of the left ventral striatum cluster is plotted across all conditions (mean beta values ± inter-subject s.e.m.), showing significant increases to Smiling Faces only when paired with success feedback (t= 5.21, p< .001), not when paired with error feedback (SF-L > AF-L, p= .54). **c)** Statistical parametric map for the whole-brain regression analysis between AVS and activation to Smiling Faces in the success feedback condition (contrast SF-W > others), showing a selective effect in the left ventral tegmental area (VTA; xyz= -6 -18 -27; z-score= 2.80). The small inset panel shows a horizontal section through the midbrain at the level of VTA. **d)** Negative correlation between avoidant attachment scores (AVS) and activity in the ventral striatum cluster (beta values relative to baseline) for Smiling Faces in the success feedback condition (SF-W; Pearson r= -.787, p< .001). **e)** Activity in the VTA cluster for condition SF-W was also inversely correlated with AVS (Pearson r= -.706, p= .003). L= Left, R= Right. ** = p <.001

between activation in these medial prefrontal regions to SF-W and AVS (vACC: xyz= -12 30 -9, z-score= 3.07, p< .001; r= 0.530; MPFC: xyz= -9 42 36, z= 3.16, p< .001; r= 0.757). No effect was found for the SF-L condition.

DISCUSSION

Our study provides several new results. Firstly, using a pseudo-social interaction paradigm, we show that brain responses to facial expressions are strongly modulated by the perceived social meaning induced by the current context. Smiling faces enhanced activation in the ventral striatum and related regions only when associated with positive feedback, whereas angry faces increased activation in amygdala only when associated with negative feedback. This indicates that responses in both striatum and amygdala were influenced by the social relevance of rewarding and punishment signals expressed by faces, respectively. While many studies have shown activation in striatum and OFC to various types of rewards such as gains or food, a few others have reported activation in the same regions to smiling or attractive faces [34,35]. Here we found that such responses were not driven by facial features alone, but reflected the social meaning of a smiling expression, i.e., when perceived as rewarding current performance and as congruent with task-goals. Similarly, while several studies have reported activation of the amygdala to angry or negative facial expressions, here we show that this response may not be automatic and driven by specific facial features [36] but determined by the personal significance of perceived anger.

These findings provide support for the importance of appraisal of personal relevance in emotional processing [21,37]. Moreover, these data also demonstrate that participants were highly motivated by the task (as also confirmed during debriefing). Indeed we observed a reliable main effect of positive (WON) or negative (LOST) feedback in regions associated with reward and motivation processes, including basal ganglia, OFC, and dorsolateral prefrontal cortex for WON trials, as well as retrosplenial cortex and right insula for LOST trials [22]. By contrast, we did not find main effects of facial expressions (smiling or angry), but only interactions of expressions with feedback context that confirm that our task induced specific social appraisals as a function of the pseudo-game context. In other words, the brain response to a visually similar facial expression was crucially dependent on current task-goals and feedback congruency, since the social significance of smiles on SF-W trials (praising success) clearly differed from SF-L trials (mocking a failure), whereas anger also differed between AF-L (reproach or punishment) and AF-W trials (conflict or frustration), resulting in distinct patterns of brain responses.

Secondly, and most importantly, we found that individual differences in adult attachment style strongly modulated responses to facial expressions in brain

Table 2	**Correlations**					
Brain Area	*BA*	*Voxel*	*x*	*y*	*z*	*Z-Value*
SF-W > others x AVS (negative)						
Anterior insula left		12	-27	21	-15	3.23
Ventral striatum left		10	-24	18	-12	3.04
Ventral tegmental area left		3	-6	-18	-27	2.8
SF-W > others x SAS (positive)						
Anterior insula left		19	-27	21	-15	3.2
Ventral striatum left		12	-24	18	-12	3
AF-L > others x SAS (negative)						
Amygdala left		9	-24	-9	-21	3.35
Medial thalamus left		4	-9	-15	-30	3.21
AF-L > others x AXS (positive)						
Amygdala left		10	-24	-9	-21	3.3
Medial thalamus left		4	-9	-15	-30	2.91
AF-W > others x AVS (positive)						
Retrosplenial cortex left	30	82	-12	-51	15	4.21
Insula right		16	42	18	0	3.98
Dorsal ACC left	11	6	-9	42	36	3.16
Ventral ACC left	32	9	-12	30	-9	3.07

Table 2: *Brain areas activated in parametric correlation analyses using attachment scores from the Adult Attachment Questionnaire (secure = SAS, anxious = AXS, avoidant = AVS). Peak coordinates are given in MNI space and listed with best estimates of anatomical location. BA = Brodmann's area, STS = Superior temporal sulcus, VTA = Ventral tegmental area. All p< .001, except for VTA p<.003 and thalamus p<.002*

regions associated with affect and motivation; and that such modulation specifically concerned those conditions related to social appraisal. Our results therefore provide new support to the view that adult attachment style can shape how individuals perceive social information in various contexts, and converge with recent behavioral findings that it may modulate recognition judgments for emotional expressions in unfamiliar faces [6,7]. These results also reveal that distinct neural substrates may underlie the two major dimensions of the attachment construct in healthy adults (anxiety and avoidance, as defined by self-report measures used here).

In the congruent feedback condition of social support (SF-W), we found that higher scores on attachment avoidance (AVS) predicted lower activation in brain regions linked to dopaminergic function and reward, including both ventral striatum and VTA [22,30]. This correlation with AVS was highly specific for the SF-W condition representing a socially rewarding interaction (but not related to reward or positive affect of faces alone), supporting the hypothesis that individuals scoring higher on AVS may show reduced activation of affective processes in response to positive social signals [6,38]. These findings are in line with behavioral evidence that people with high AVS tend to prefer physical and emotional distance from others, and usually do not seek social support [3,39]. High AVS is also associated with greater self-reliance and a tendency to dismiss the benefits of group interactions [17]. Here we show that such tendencies to avoidant attachment may entail a relative down-regulation of reward-related activity in striatal circuits during socially reinforcing interactions, presumably underlying at least in part the relative impassiveness of individuals with high AVS to social rewards. Our results also provide a plausible substrate for behavioral observations that high AVS is negatively correlated with reward dependence [11], and add support to recent proposals that some forms of social avoidance may be associated with reduced positive experiences in social and non-social contexts [40].

On the other hand, we found that higher scores on anxious attachment (AXS) were correlated with selective increases in left amygdala responses to social signals of reproach or punishment (i.e. angry expressions combined with congruent negative feedback, AF-L). These data reveal that processing of socially aversive situations is specifically enhanced in brain systems associated with emotional arousal and fear [41] for people with higher anxious attachment. Because the amygdala is particularly implicated in processing self-relevant affective information [21,42], our findings support the notion that a key aspect of anxious attachment may involve enhanced vigilance towards emotionally-significant social cues [3,4,17,39]. These condition-specific responses also accord with the view that anxious attachment involves a "relation-specific anxiety" that is distinct from more general forms of anxiety or neuroticism [29]. In keeping with this, people with high AXS typically show increased monitoring and exaggerated appraisal of threats to the self, intensify negative emotional responses to emotional or social events, and unlike subjects with high AVS, tend to search more for external sources of support and comfort [3]. These results also converge with recent findings that greater amygdala responses to negative sentences may relate to attachment insecurity [23], although the latter study did not examine the distinct prototypes of attachment as here, but inferred more general attachment differences (secure or insecure) based on reaction times to the sentences (slow or fast).

Importantly, note that even though the AF-L condition represented negative social feedback, it was nevertheless congruent with the goals and expectations of participants on LOST trials, and thus did not correspond to a condition of social rejection or exclusion as implemented in other paradigms [31,32]. Here, angry faces were perceived as in-group partners or allies who disapproved failures in the task and hence expressed punishment – a condition meant to activate the need for support in challenging or distressing situations that is intrinsic to anxious attachment style [3]. Accordingly, this condition evoked selective activation in the amygdala, rather than in anterior cingulate cortex as reported in previous studies where social rejection implied group exclusion or conflict [31-33].

In our study, the third prototype of adult attachment style (secure) did not exhibit any unique correlate for neural responses to the perceived social meaning of facial expressions, but mirrored the pattern found for AVS and AXS, respectively. Thus, high scores on SAS correlated positively with activation of the ventral striatum to rewarding smiles (SF-W) and

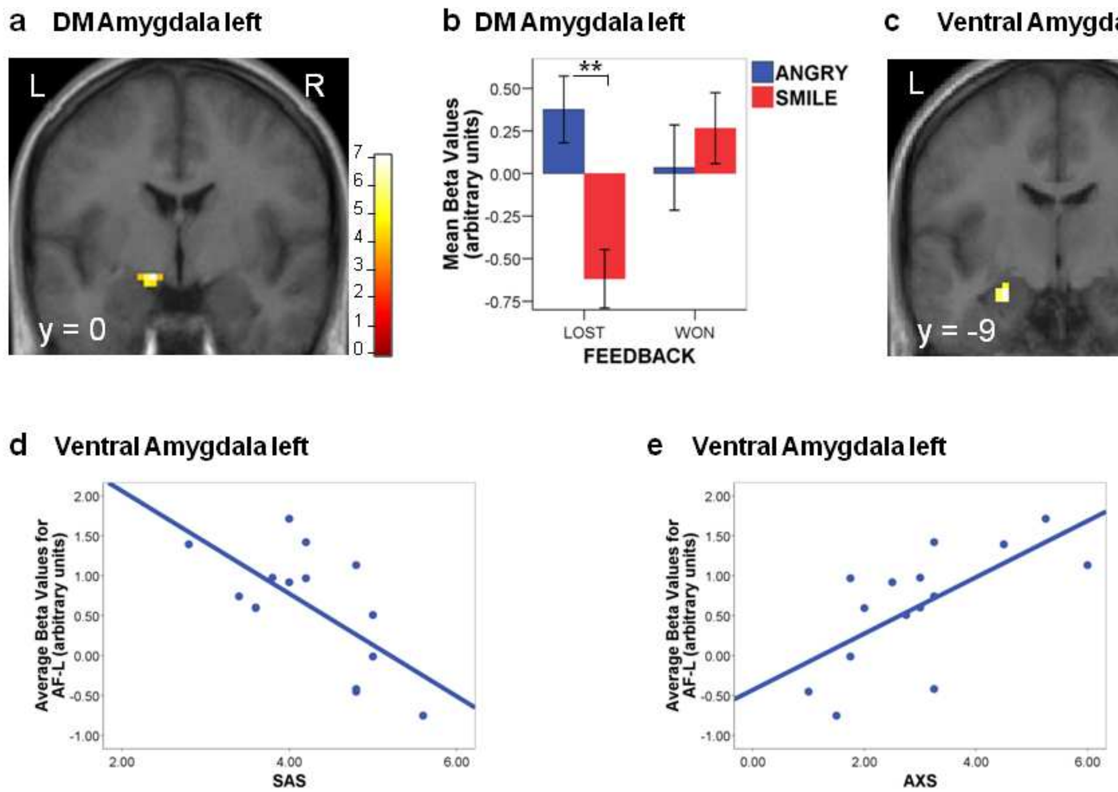


Figure 3:
a) *Statistical parametric map for Angry vs Smiling expressions in error feedback context (contrast AF-L > SF-L), showing activation in left dorso-medial amygdala (xyz= -12 -3 -15, z-score= 4.62, p< .001).* ***b)*** *Activation for this amygdala cluster is plotted across all feedback conditions (mean beta values ± inter-subject s.e.m), showing significant difference between Angry and Smiling faces with error feedback (t= 4.94, p< .001) but no significant difference with success feedback (p= .50).* ***c)*** *Statistical parametric map for the whole-brain multiple regression analysis between AXS and activation to Angry Faces in the lost feedback condition (contrast AF-L > others), showing a selective effect in left amygdala (xyz = -24 -9 -21; z-score= 3.30).* ***d)*** *Activity in this amygdala cluster (beta values relative to baseline) for Angry Faces in the lost feedback condition (AF-L) was negatively correlated with SAS (Pearson r= -.655, p= .008).* ***e)*** *Activity in the same amygdala cluster for condition AF-L was also positively correlated with AXS (Pearson r= .668, p= .006). L= Left, R= Right. ** = p <.001*

negatively with activation of the amygdala to reproach faces (AF-L). These data therefore accord with the theoretical view that secure attachment may correspond to a combination of low anxiety and low avoidance, and add new neurobiological evidence in support of bi-dimensional models postulating that these two major components may account for the different categories of adult attachment style [10]. Critically, our fMRI results reveal that these two dimensions (anxious and avoidant attachment) have distinct neural bases in two key brain systems implicated in affect and motivation, centered on the amygdala and striatum, respectively. Both the striatum and amygdala play important roles for learning and predicting motivational outcomes in specific situational contexts, and might therefore be well suited for the establishment of idiosyncratic affective responses to social cues based on past experience or developmental history.

Thus, although the exact correspondence between developmental aspects of attachment initially described in infancy [1,2] and attachment style in adults is still partly unclear [9], our results demonstrate that this social psychological construct taps specific affective processes, with distinct neural substrates, which can influence how people automatically perceive and respond to social signals in interaction contexts, beyond relationships with intimate partners or close personal acquaintances [3,8]. In line with our findings, adult attachment has been shown to affect the recognition of emotional expressions in morphs of unfamiliar faces [6,7], especially when such expressions are relevant to attachment concerns and interpersonal bonding [26,43], as in our pseudo-social game paradigm. The current imaging findings that brain regions activated by face expressions are differentially modulated by individual attachment style provide new insights on the neurobiological underpinnings of these effects. More generally, unveiling such links between fundamental social dimensions and brain function may not only validate traditional psychosocial conceptualization but also help understand their impact on human behavior.

Activation of STS and MPFC were found only in incongruent feedback conditions corresponding to social opposition or confrontation (AF-W and SF-L), but did not correlate with attachment traits. STS is implicated in theory of mind and perception of intentionality [27], suggesting that participants were more inclined to imagine particular mental states or intentions for faces seen with incongruent feedback information. However, activity in STS did not appear to subtend differences in “mental models” of others that are typically associated with different attachment styles [5,10,25]. On the other hand, incongruent feedback with angry faces on WIN trials (AF-W) also activated MPFC and vACC, previously implicated in responses to social inclusion-exclusion and emotional conflict [31-33]. Moreover, activity in MPFC and vACC correlated with AVS and overlapped with similar regions activated by social rejection[32] or emotion suppression [24] in other paradigms, suggesting that affective evaluation processes responding to conflict situations might be more active in avoidant subjects in keeping with their more negative appraisal of others [5,10,25]. This might further contribute to the reduced sensitivity to social reward observed in these subjects.

In sum, our study shows that the two dimensions of adult attachment have distinct neural substrates and produce specific effects on the appraisal of social facial signals. Ventral striatum and VTA were selectively activated by the rewarding feedback value of smiling faces accompanying a success, and thus representing social reward, but this response was blunted in individuals with high AVS scores. Amygdala was selectively activated by the reproach value of angry faces combined with errors, thus representing social punishment, and this response was enhanced in individuals with high AXS scores. In other words, both striatum and amygdala responses were specific to the perceived social meaning of face expressions in relation to current task goals, because no such activity was elicited by the same expressions with a different (incongruent) feedback. Moreover, high AVS also correlated with an increased response to potential social confrontation in ACC, consistent with negative relational schemata hold by avoidant individuals. In contrast, secure attachment was characterized by higher striatal response to rewarding faces and lower amygdala responses to reproach faces, but showed no unique activation pattern, supporting the idea that it may entail a combination of low avoidant and low anxious traits [10]. By revealing a critical involvement of emotional brain systems associated with social reward and threat in adult attachment style, our fMRI data provide the first direct neurobiological evidence in support of psychological models proposing two independent affective dimensions to explain these individual differences. More generally, our data also converge with bidimensional models of social disorders that suggest distinct contributions of negative and positive emotions in regulating social behavior and interpersonal communication in a wide range of social contexts [40,44]. Altogether, these results may ultimately help define appropriate intervention strategies in clinical disorders of attachment and social functioning, including autism, phobias, and other relational disturbances.

MATERIALS AND METHODS

Subjects

We recruited 16 healthy volunteers (8 males, mean age 23.6 ± 3.6, all right-handed) who had normal or corrected to normal vision, no history of neurological or psychiatric disease, and gave written informed consent according to the local ethical committee (Commission centrale d'éthique de la recherche sur l'être humain; le Comité départmental d'éthique de N.A.C.) regulation.

Stimuli and procedure

Visual dot-counting was presented as the primary task to participants. Each trial began with a white central fixation-cross on a black screen (for 3 to 7 sec, average 3.5 sec), followed by a brief visual display divided in two parts with a variable number of white dots on each side of the screen (presented for 500 ms). The number of dots on each side ranged from 10 to 15. Their quantity and position were randomly assigned on every trial for each side separately, in such a way that the display was never visually identical on both sides (see Figure 1a). Participants had to indicate which side of the screen contained more dots (right/left) by pressing one of two response-keys. The total number of dots and the difference between the two display sides were adjusted online based on the participant's performance on preceding trials, by reducing the difference after each correct trial (minimum 1 dot) or increasing the difference after each incorrect trial (maximum 5 dots), allowing us to maintain performance close to threshold and to obtain approximately equal numbers of correct and incorrect trials (mean correct = 57±1% across conditions). In addition, to further ensure this equal distribution, occasional displays with 15 dots on both sides were inserted whenever performance exceeded 60% correct of two consecutive trials (20 ± 5.6 out of 128 trials). None of the participants noticed these "trick" trials.

The dot display was followed by a black screen with a variable interval (jitter of 1000 to 1400 ms, average 1200 ms), during which participants gave their response; and then by a visual feedback screen (1500 ms) consisting of a face (with either a smiling or angry expression) paired with a verbal indication of actual performance on the counting task (either "WON" or "LOST"). The verbal feedback always corresponded to real performance success or failure on the preceding trial (except on the few "trick" trials with equal number of dots on both sides, where a negative "LOST" feedback was given to reduce an excess of correct over incorrect trials). By contrast, the facial emotional expression was pseudo-randomly assigned on every trial, with the constraint that smiling and angry faces appeared on an equal number of correct and incorrect trials each. This design resulted in 4 different combinations of verbal and facial feedback (see Figure 1b): Smiling Face on WON trial (SF-W) or LOST trial (SF-L), Angry Face on WON trial (AF-W) or LOST trial (AF-L). Face stimuli were colour photographs of 16 different individuals (8 males) from the Karolinska Directed Emotional Faces set (KDEF, Lundquist D., Flykt A., and Öhmann A., 1998). Each face identity was assigned to one condition only (2 males and 2 females in each of the 4 feedback types, counterbalanced across participants). Thus, for a given participant, a given face was always seen with the same expression (either smiling or angry) and the same feedback message (either positive "WON" or negative "LOST") throughout the task. Each face identity was repeated 8 times in the corresponding conditions, in random order, resulting in 128 trials in total per participants (with a total duration of approximately 15 min).

We induced a pseudo-social game context by telling participants a cover-story along the following lines: faces were those of other subjects who already participated and belonged to two different groups; the study goal was to compare perceptual abilities and cooperation among the groups; they had been randomly assigned to play for subjects in one of these groups; each correct response gave one point to this group while each incorrect response gave one point to the other group; these outcomes would be reminded to them during the game by displaying faces from the different groups with appropriate expressions. Subsequent debriefing after scanning indicated that participants accepted the cover-story and were highly motivated by the task. They consistently reported that they were “surprised”, “irritated”, or “annoyed” by incongruent feedback combinations, and typically described these faces as “untrustworthy”, “foe”, “envious”, etc. Together with brain data, these reports clearly suggest that feedback context strongly modulated the subjective interpretation of facial expressions.

Questionnaires

We used the Adult Attachment Questionnaire (AAQ), a validated French version [28] of the original Attachment Style Measure (ASM) [45], which includes a series of 13 statements rated along a 7-point scale (from “strongly disagree” to “strongly disagree”). This instrument yields three separate scores, one for each prototypical style including avoidant (AVS), anxious (AXS), and secure attachment (SAS), which have been shown to be reliably distinct from each other [28]. The AAQ thus provides quantitative indices for the relative strength of each of the three classic attachment categories, but also allows bi-dimensional measures for anxious and avoidant axes in attachment space. The 13 items of the ASM (or AAQ) are also included in the Relationship Scales Questionnaire (RSQ) [46], comprising 30 items in total (with some items reversed), and a recent review [29] suggests that the best method to analyze the 30-item RSQ is to rely on these 13 items alone to create a bi-dimensional attachment space made of the avoidance and anxiety measures. Moreover, we found strong correlations between avoidance and anxiety dimensions obtained by the AAQ and RSQ (Pearson r> .797, p< .0004). Other personality questionnaires included STAXI [47], STAI [48], BIS-BAS [49], and PANAS [50]. We also obtained other debriefing measures about credibility of the task and affect during the different feedback conditions using 5-point Lickert scales (e.g. degree of satisfaction or frustration on WON or LOST trials, respectively; subjective experience elicited by seeing angry or smiling faces; and subjective irritation elicited by incongruent expressions), as well as a likeability and memory test for the different face identities (these data showed no effect of attachment style and are not reported). Because the personality questionnaires from one subject were incomplete, only 15 subjects (n= 15) were included in our correlation analysis with these measures.

MRI acquisition and analysis

MRI data were acquired on a 1.5 T whole-body INTERA system (Philips Medical Systems), using a standard head-coil configuration. For each participant, structural images were obtained with a 3D-GRE T1-weighted sequence (FOV = 250 mm, TR/TE/Flip = 15 ms/5.0 ms/30°, matrix = 256 × 256, slice-thickness = 1.25 mm) and functional images with a GRE EPI sequence (TR/TE/Flip = 2500 ms/40 ms/80°, FOV = 250 mm, matrix = 128 × 128). Functional images covered the whole brain, consisting of 30 contiguous 4mm axial slices parallel to the inferior edge of the occipital and temporal lobes, and acquired continuously for a total of 232 images per participant.

Functional images were analyzed using the general linear model for event-related designs in SPM2 (Wellcome Department of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm). All images were realigned, corrected for slice timing, normalized to an EPI template (re-sampled voxel-size of 3 mm), spatially smoothed (8 mm FWHM Gaussian kernel). A high-pass frequency filter (cutoff 120 s) and corrections for auto-correlation between scans were applied to the time series.

Statistical analysis was performed using the general linear model implemented in SPM2, with a separate regressor for each event type convolved with a canonical hemodynamic response function. Six events were modelled, including the dot display on correct and incorrect trials, and the 4 critical feedback conditions (SF-W, SF-L, AF-W, AF-L). Movement parameters from realignment corrections were entered as additional covariates of no interest to account for residual movement artifacts after realignment. Statistical parametric maps were generated from linear contrasts between the different feedback conditions in each participant. A second-stage random-effect analysis was then performed using one-sample t tests on contrast images obtained in each subject for each comparison of interest. All contrasts were performed across the whole brain using standard threshold criteria [51] of significant activation at a voxel-level of p< .001 (uncorrected) and cluster size equal or greater than 5 voxels. Average parameter estimates of activity (betas) for each feedback condition were extracted from all voxels in regions of interest (ROIs), defined by the full-extend clusters showing significant activation at a voxel-level of p<.001 (uncorrected) in the SPM group analysis (random-effect contrasts).

Statistical correlations with attachment and personality traits were performed in two stages. Firstly, we tested for a relation of the average beta values from activated ROIs with standardized questionnaire scores (attachment security, anxiety, or avoidance: SAS, AXS, and AVS, respectively) using one-tailed Pearson product moment coefficient in SPSS 14.0 (SPSS, Chicago, Illinois, United States). Secondly, whenever this first stage showed a significant correlation or a strong trend for an activated region (e.g. striatum and amygdala), we performed a whole-brain multiple regression analysis on the contrast image of interest using the relevant questionnaire scores (e.g. AVS) as a linear parametric factor in SPM2, allowing us to test for any voxels throughout the brain where activation in the given contrast varied as a function of this behavioral measure [52]. For this second-stage correlation analysis, significant effects were identified using a threshold of p< .001 at the voxel-level (uncorrected) and cluster size equal or greater than 5 voxels.

Eye Tracker acquisition and analysis

To compare visual inspection between different conditions, eye movements were monitored continuously during scanning with an MRI-compatible infra-red eyetracker LRO L6 (Applied Science Laboratories, Bedford, MA, USA). Eye position coordinates (x and y) were recorded at 60Hz and saved for offline analysis. Data from two subjects had to be discarded for technical reasons (inaccurate calibration or missing data). Trial-by-trial epoching and processing of the data was performed with Eyenal 6000 software (ASL, Bedford, USA). For each trial in each condition, the number and duration of fixations were calculated over two areas of interest (AOI) on the feedback screen, corresponding to the face and the word message (WON or LOST), respectively. Fixation data were averaged for each AOI, in each of the four conditions (SF-W, AF-L, AF-W, SF-L) and each participant, and then submitted to a 2x2x2 repeated-measure ANOVA using SPSS 14.0 (SPSS, Chicago, Illinois, USA). Within-subject factors were face expression (smiling or angry), success feedback (won or lost), and screen area (face or word). These analyses revealed only a significant main effect of AOI (face > word; $F_{1,12} = 74.3$; $p < .001$), indicating that subjects spent more time looking at the faces than words. Critically, there was no effect of expression, success, or any interaction. ANOVAs were also performed with a between-subject factor of “attachment style” (SAS, AXS, or AVS, as determined individually by AAQ scores), to examine whether these individual traits would modulate visual inspection times, but these analyses showed no main effect of attachment nor interaction with conditions (all Fs < 1.1). These results indicate that modulation of brain responses between different attachment styles are not simply due to changes in visual attention to faces in feedback displays.

ACKNOWLEDGMENT

This research was supported by the National Center of Competence in Research (NCCR) Affective Sciences financed by the Swiss National Science Foundation (n° 51NF40-104897) and hosted by the University of Geneva, and grants of the Swiss National Science Foundation to DS and PV.

REFERENCES

1. Bowlby J (1969) Attachment and loss. New York,: Basic Books. v. p.
2. Ainsworth MDS (1978) Bowlby-Ainsworth Attachment Theory. Behavioral and Brain Sciences 1: 436-438.
3. Mikulincer M, Shaver PR (2007) Attachment in Adulthood: Structure, Dynamics, and Change. New York: The Guilford Press.
4. Van Emmichoven IAZ, Van Ijzendoorn MH, De Ruiter C, Brosschot JF (2003) Selective processing of threatening information: Effects of attachment representation and anxiety disorder on attention and memory. Development and Psychopathology 15: 219-237.
5. Hazan C, Shaver P (1987) Romantic Love Conceptualized as an Attachment Process. Journal of Personality and Social Psychology 52: 511-524.
6. Fraley RC, Niedenthal PM, Marks M, Brumbaugh C, Vicary A (2006) Adult attachment and the perception of emotional expressions: Probing the hyperactivating strategies underlying anxious attachment. Journal of Personality 74: 1163-1190.
7. Niedenthal PM, Brauer M, Robin L, Innes-Ker AH (2002) Adult attachment and the perception of facial expression of emotion. Journal of Personality and Social Psychology 82: 419-433.
8. Baldwin MW (1992) Relational Schemas and the Processing of Social Information. Psychological Bulletin 112: 461-484.
9. Riggs SA, Paulson A, Tunnell E, Sahl G, Atkison H, et al. (2007) Attachment, personality, and psychopathology among adult inpatients: Self-reported romantic attachment style versus Adult Attachment Interview states of mind. Development and Psychopathology 19: 263-291.
10. Bartholomew K, Horowitz LM (1991) Attachment Styles among Young-Adults - a Test of a 4-Category Model. Journal of Personality and Social Psychology 61: 226-244.
11. Chotai J, Jonasson M, Hagglof B, Adolfsson R (2005) Adolescent attachment styles and their relation to the temperament and character traits of personality in a general population. European Psychiatry 20: 251-259.
12. Feeney JA, Ryan SM (1994) Attachment Style and Affect Regulation - Relationships with Health Behavior and Family Experiences of Illness in a Student Sample. Health Psychology 13: 334-345.
13. Simpson JA, Collins WA, Tran S, Haydon KC (2007) Attachment and the experience and expression of emotions in romantic relationships: A developmental perspective. Journal of Personality and Social Psychology 92: 355-367.
14. Overall N (2003) Mapping the intimate relationship mind: Comparisons between three models of attachment representations. Australian Journal of Psychology 55: 57-57.
15. Collins NL, Feeney BC (2004) Working models of attachment shape perceptions of social support: Evidence from experimental and observational studies. Journal of Personality and Social Psychology 87: 363-383.
16. Simpson JA, Rholes WS, Phillips D (1996) Conflict in close relationships: An attachment perspective. Journal of Personality and Social Psychology 71: 899-914.
17. Rom E, Mikulincer M (2003) Attachment theory and group processes: The association between attachment style and group-related representations, goals, memories, and functioning. Journal of Personality and Social Psychology 84: 1220-1235.
18. Fuendeling JM (1998) Affect regulation as a stylistic process within adult attachment. Journal of Social and Personal Relationships 15: 291-322.
19. Aron A, Fisher H, Mashek DJ, Strong G, Li HF, et al. (2005) Reward, motivation, and emotion systems associated with early-stage intense romantic love. Journal of Neurophysiology 94: 327-337.
20. Bartels A, Zeki S (2004) The neural correlates of maternal and romantic love. Neuroimage 21: 1155-1166.
21. Sander D, Grafman J, Zalla T (2003) The human amygdala: an evolved system for relevance detection. Rev Neurosci 14: 303-316.
22. Dalgleish T (2004) The emotional brain. Nature Reviews Neuroscience 5: 582-589.
23. Lemche E, Giampietro VP, Surguladze SA, Amaro EJ, Andrew CM, et al. (2006) Human attachment security is mediated by the amygdala: Evidence from combined fMRI and psychophysiological measures. Human Brain Mapping 27: 623-635.
24. Gillath O, Bunge SA, Shaver PR, Wendelken C, Mikulincer M (2005) Attachment-style differences in the ability to suppress negative thoughts: Exploring the neural correlates. Neuroimage 28: 835-847.
25. Bowlby J (1982) Attachment and loss. New York: Basic Books. v. p.
26. Collins NL, Read S (1994) Cognitive representations of attachment: the structure and function of attachment models. In: Bartholomew K, Perlman D, editors. Attachment processes in adulthood: Advances in personal relationships. London: Kingsley.
27. Frith U, Frith CD (2003) Development and neurophysiology of mentalizing. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 358: 459-473.
28. Bouthillier D, Tremblay N, Hamelin F, Julien D, Scherzer P (1996) French translation and validation of an adult attachment styles measure. Canadian Journal of Behavioural Science-Revue Canadienne Des Sciences Du Comportement 28: 74-77.
29. Kurdek LA (2002) On being insecure about the assessment of attachment styles. Journal of Social and Personal Relationships 19: 811-834.
30. O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. Nature Neuroscience 4: 95-102.
31. Eisenberger NI, Lieberman MD, Williams KD (2003) Does rejection hurt? An fMRI study of social exclusion. Science 302: 290-292.
32. Somerville LH, Heatherton TF, Kelley WM (2006) Anterior cingulate cortex responds differentially to expectancy violation and social rejection. Nature Neuroscience 9: 1007-1008.
33. Kross E, Egner T, Ochsner K, Hirsch J, Downey G (2007) Neural dynamics of rejection sensitivity. Journal of Cognitive Neuroscience 19: 945-956.
34. O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, et al. (2003) Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. Neuropsychologia 41: 147-155.
35. Bray S, O'Doherty J (2007) Neural coding of reward-prediction error signals during classical conditioning with attractive faces. Journal of Neurophysiology 97: 3036-3045.
36. Whalen PJ, Kagan J, Cook RG, Davis FC, Kim H, et al. (2004) Human amygdala responsivity to masked fearful eye whites. Science 306: 2061-2061.
37. Kim H, Somerville LH, Johnstone T, Polis S, Alexander AL, et al. (2004) Contextual modulation of amygdala responsivity to surprised faces. Journal of Cognitive Neuroscience 16: 1730-1745.

38. Simpson JA, Rholes WS, Orina MM, Grich J (2002) Working models of attachment, support giving, and support seeking in a stressful situation. Personality and Social Psychology Bulletin 28: 598-608.
39. Dozier M, Kobak RR (1992) Psychophysiology in Attachment Interviews - Converging Evidence for Deactivating Strategies. Child Development 63: 1473-1480.
40. Kashdan TB (2007) Social anxiety spectrum and diminished positive experiences: Theoretical synthesis and meta-analysis. Clinical Psychology Review 27: 348-365.
41. Davis M, Whalen PJ (2001) The amygdala: vigilance and emotion. Molecular Psychiatry 6: 13-34.
42. Fitzgerald DA, Angstadt M, Jelsone LM, Nathan PJ, Phan KL (2006) Beyond threat: Amygdala reactivity across multiple expressions of facial affect. Neuroimage 30: 1441-1448.
43. Fraley RC, Davis KE, Shaver PR (1998) Dismissing-avoidance and the defensive organization of emotion, cognition, and behaviour. In: Simpson JA, Rholes WS, editors. Attachment theory and close relationships. New York: Guilford Press. pp. 249-279.
44. Rapee RM, Spence SH (2004) The etiology of social phobia: Empirical evidence and an initial model. Clinical Psychology Review 24: 737-767.
45. Simpson JA (1990) Influence of Attachment Styles on Romantic Relationships. Journal of Personality and Social Psychology 59: 971-980.
46. Griffin D, Bartholomew K (1994) Models of the Self and Other - Fundamental Dimensions Underlying Measures of Adult Attachment. Journal of Personality and Social Psychology 67: 430-445.
47. Spielberger CD, Sydeman, S. J., Owen, A. E., & Marsh, B. J. (1999) Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI). In M D Maruish (Ed), The use of psychological testing for treatment planning and outcomes assessment (2nd ed) (pp 993-1021) Mahwah: Laurence Erlbaum.
48. Spielberger CD (1983) Manual for the State-Trait Anxiety Inventory Palo-Alto: Consulting Psychologists Press.
49. Carver CS, White TL (1994) Behavioral-Inhibition, Behavioral Activation, and Affective Responses to Impending Reward and Punishment - the Bis Bas Scales. Journal of Personality and Social Psychology 67: 319-333.
50. Watson D, Clark LA, Tellegen A (1988) Development and Validation of Brief Measures of Positive and Negative Affect - the Panas Scales. Journal of Personality and Social Psychology 54: 1063-1070.
51. Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, et al. (1996) A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapping 4: 58-73.
52. Phan KL, Taylor SF, Welsh RC, Ho SH, Britton JC, et al. (2004) Neural correlates of individual ratings of emotional salience: a trial-related fMRI study. Neuroimage 21: 768-780.

VI.2. fMRI Study N°1b

Published in: Social Neuroscience (4) 5: p. 384 - 401(2009)

fMRI study N° 1b was designed to explore the neural correlates of person perception using a familiar face recognition paradigm. Thereby, the main questions were:

- **What are the neural correlates of recognizing a previously stranger but now familiar face that has been encountered beforehand within either a positive (friend) or a negative (foe) personally relevant social context?**
- **Which interactions, positive (friend) or negative (foe), are more readily retrieved during familiar face recognition, and are such person impressions stored implicitly or explicitly in the human brain?**

The most important improvements compared to similar previous experiments included the investigation of both positive (friends) and negative (foes) person impressions, as well as the use of faces that were unknown to the participants before the task and only made familiar through a pseudo-social game context (see Chapter V.1.2. for more details).

The results showed that a number of brain areas including fusiform gyrus (FG), precuneus, posterior cingulate (PCC), temporal pole, pSTS, insula, and amygdala displayed increased activity to previously encountered as compared to new faces. However, social context only led to differential activations in FG, PCC and amygdala. Thereby, FG showed selective increased BOLD signal change to smiling faces previously seen in a negative (and thus incongruent) context, PCC was more activated to smiling faces in general, and in bilateral amygdala, there was a significant activation difference between both incongruent context conditions (smiling face while losing and angry face while winning) and new faces. Additional analyses revealed a similar incongruent *versus* congruent effect in ventral ACC, and a selective activation difference between angry faces seen while winning *versus* losing in left caudate nucleus.

These activation patterns suggest that several cortical and limbic brain areas keep track of negative person impressions through processes related to vigilance (amygdala), social exclusion (ACC), or social hierarchy (caudate nucleus). Importantly, such activations were found despite the lack of explicit knowledge of the social context faces were encountered at the first place, and even though faces were shown with a neutral expression during recognition. Our results thus show that automatic, unconscious emotional processes play a crucial role in familiar face recognition, providing us with important clues how to guide our social behavior during interactions.

Memory for friends or foes:

The social context of past encounters with faces modulates their subsequent neural traces in the brain

Pascal Vrtička (1,2), Frédéric Andersson (2), David Sander (1,3), & Patrik Vuilleumier (1,2)

(1) Swiss Center for Affective Sciences, University of Geneva

(2) Laboratory for Neurology & Imaging of Cognition, Department of Neurosciences
& Clinic of Neurology, University Medical Center of Geneva

(3) Department of Psychology, FPSE, University of Geneva

ABSTRACT

Every day we encounter new people, interact with them, and form person impressions based on quick and automatic inferences from minimal contextual information. Previous studies have identified an extensive network of brain areas involved in familiar face recognition, but there is only little evidence up to date concerning the neural bases of negative *versus* positive person impressions.

In the present study, participants were repeatedly exposed to 16 unfamiliar face identities within a pseudo-interactive game context to generate a perception of either “friends” or “foes”. Functional magnetic resonance imaging (fMRI) was then performed during an old/new memory task to assess any difference in brain responses to these now familiar face identities, relative to unfamiliar faces. Importantly, whereas facial expressions were always emotional (either smiling or angry) during the encoding phase, they were always neutral during the memory task.

Our results reveal that several brain regions involved in familiar face recognition, including fusiform cortex, posterior cingulate gyrus, and amygdala, plus additional areas involved in motivational control such as caudate and anterior cingulate cortex, were differentially modulated as a function of a previous encounter, and generally more activated when faces were perceived as “foes” rather than “friends”. These findings underscore that a key dimension of social judgments, based on past impressions of who may be supportive or hostile, may lead to long-lasting effects on memory for faces and thus influence affective reactions to people during a subsequent encounter even in a different (neutral) context.

CORRESPONDING AUTHOR:

Pascal Vrtička
Swiss Center for Affective Sciences
7, rue des Battoirs, 1205 Geneva, Switzerland
Tel: +41 – (0)22 379 9824 Fax: +41 – (0)22 379 9844
Email: pascal.vrticka@unige.ch

INTRODUCTION

Our everyday life is marked by innumerable social interactions with a wide range of different people, including family members, friends, professional colleagues, as well as complete strangers. To determine how to appropriately interact with these various individuals, the ability to form and remember, even implicitly, general impressions about a given person is a crucial social skill of human beings.

Several studies have shown that only minimal information is needed to form person impressions from just seeing faces (Todorov & Uleman, 2002, 2003, 2004), or to make trait judgements such as trustworthiness, competence, and aggressiveness (Willis & Todorov, 2006). This social evaluation process may not only occur unintentionally, but also very rapidly, since very short exposure times to faces are sufficient for some of these judgments (Bar, Neta, & Linz, 2006; Willis & Todorov, 2006). Besides trait inferences from seeing faces, other studies have also demonstrated that people can make quick unreflective inferences concerning a person based on minimal behavioural information, even when the behaviour can be explained by situational context rather than personal dispositions (Gilbert & Malone, 1995; Trope & Alfieri, 1997). Moreover, it has been shown that people make such rapid inferences about another person even when they do not intend to do so (Uleman, Blader, & Todorov, 2005), and when their cognitive resources are constrained by another concomitant task (Bargh, 1994), suggesting that this impression formation process is fairly automatic. Importantly, once person impressions are formed during social interactions, storing this information in memory is crucial to make use of these inferences and other person-related knowledge when a person is encountered again (see Mealey, Daood, & Krage, 1996; Oda, 1997; Takahashi, 2005). In particular, it has been shown that memory for person identities can be modulated at subsequent retrieval when these persons were first seen with some emotion expressed in the face (D'Argembeau & Van der Linden, 2007) or the voice (Armony, Chochol, Fecteau, & Belin, 2007), as compared with when their expression is neutral.

The neural mechanisms underlying these social processes are still poorly known, but there is increasing evidence that several brain regions are critically important for social appraisals. Among these, the amygdala has been shown to be involved in tracking the trustworthiness of faces by responding to some categories of faces that have properties commonly perceived as signalling untrustworthiness (Adolphs, Tranel, & Damasio, 1998; Armony, Chochol, Fecteau, & Belin, 2007; Engell, Haxby, & Todorov, 2007; Winston, Strange, O'Doherty, & Dolan, 2002). Other regions involved in face processing such as the fusiform gyrus (FG) and superior temporal sulcus (STS) have been found to be modulated by socially significant signals, even during tasks that do not require any explicit social or affective inferences (Frith & Frith, 2003; Singer, Kiebel, Winston, Dolan, & Frith, 2004; Vuilleumier, Armony, Driver, & Dolan, 2001; Winston, Strange, O'Doherty, & Dolan, 2002). In addition, prefrontal regions including the anterior cingulate cortex (ACC) may also be differentially activated as a function of perceived social rejection during exposure to faces (Somerville, Heatherton, & Kelley, 2006) or pseudo-interactive social games (Eisenberger, Lieberman, & Williams, 2003).

However, it still remains largely unknown how information from a first encounter with a person determines the outcome of future interactions with the same person, and how memory traces from past impressions may affect brain responses during subsequent encounters in a different context. Only a few recent studies have begun to investigate how person impressions based on facial and/or behavioural information may translate into more enduring memory representations, and later influence social processing during new encounters with familiar persons or faces.

A first approach widely used in early studies investigating the effects of social knowledge and memory of face perception was to compare brain responses to personally familiar or intimate persons, relative to famous and unknown faces (Bartels & Zeki, 2004; Shah et al., 2001). Results from these studies have provided the basis for an influential model of face recognition and social cognition (Gobbini & Haxby, 2007). This model proposed that recognition of familiar faces does not only involve brain areas coding for invariant visual features of face identity (such as FG), but also brain regions associated with person knowledge (including STS, posterior cingulate cortex [PCC], and anterior temporal areas) as well as emotional processing (such as insula and amygdala). This model therefore implies that, besides the retrieval of purely visual traces stored from previous encounters, successful recognition of familiar faces involves the reactivation of several other essential components including information about the personality, attitude, and outlook of a person, as well as the appropriate emotional response associated with this individual (Gobbini & Haxby, 2007).

Even though the model developed by Gobbini and Haxby (Gobbini & Haxby, 2007) provides a solid framework for the role of specific brain regions in these different aspects of familiar face recognition, it does not delineate the neural components involved in the formation and maintenance of different types of person impressions, such as basic distinctions between positive or negative impressions. Importantly, because the model was initially based on two studies comparing personally familiar faces *versus* famous or unknown faces (Gobbini, Leibenluft, Santiago, & Haxby, 2004) or pictures of one’s own children *versus* familiar unrelated and unfamiliar children (Leibenluft, Gobbini, Harrison, & Haxby, 2004), neural activations imputed to knowledge and affect associated with familiar faces might have been biased (or actually limited) to positive traits. Another imaging study that manipulated the acquired moral status of previously unfamiliar faces, by using a pseudo-interactive economic game (Singer, Kiebel, Winston, Dolan, & Frith, 2004), also reported only activations evoked by pictures of cooperators versus defectors (hence related to inherently positive and rewarding values), but did not report the inverse comparison, concerning any response associated with negative impressions. Only a single recent study tested for the neural correlates of negative person traits by comparing activations to previously unknown faces after they were paired with scenarios describing neutral, nice, aggressive, and disgusting behaviours (Todorov, Gobbini, Evans, & Haxby, 2007). This study reported that the left STS was more activated for negative versus positive traits, which was attributed to a stronger engagement of processes underlying the analysis of social intentions (Frith & Frith, 1999) in the case of negative affective knowledge about a person. In addition, bilateral anterior insula was also found to show stronger responses to faces associated with aggressive and disgusting behaviours, consistent with the role of this region in processing disgust-related stimuli (Calder, Keane, Manes, Antoun, & Young, 2000).

In the present fMRI study, we therefore designed a new paradigm that allowed us to investigate the neural correlates of processing both positive and negative person traits in familiar face recognition, as formed implicitly following encounters in different social contexts. For this purpose, we engaged our participants in a pseudo-social interaction game task with a set of unknown faces that were presented as virtual partners. Distinct social attitudes were generated for these faces by systematically manipulating the social significance of their facial expression in relation to different task contexts, so as to elicit a perception of either social alliance (i.e. friendly partners) or social hostility (i.e. opponent partners) during this pseudo-game phase (see Methods). Critically, the same face identities were later shown again during fMRI, now with a neutral expression but at present familiar to the participants and associated with different affective values. Importantly, because all face identities were presented with a neutral facial expression and without any social context during this second phase, any differences in brain responses to these faces should reflect acquired person knowledge based on the past encounters and previously formed person impressions.

We predicted that brain activations related to the familiarity of previously seen faces should show a distinct profile as a function of their social-emotional value, including for regions involved in face-processing, such as FG or STS, but also for regions involved in affective processing such as the amygdala, ACC, or orbitofrontal areas, which have previously been found to track (un)trustworthiness of faces (Winston, Strange, O'Doherty, & Dolan, 2002), signs of social rejection (Somerville, Heatherton, & Kelley, 2006), or social attachment (Delgado, Frank, & Phelps, 2005; Singer, Kiebel, Winston, Dolan, & Frith, 2004), respectively.

METHODS

Subjects

We recruited 16 healthy paid volunteers (8 males, mean age 23.6 ± 3.6, all right-handed) from the student population of Geneva University. All had a normal or corrected to normal vision, no history of neurological or psychiatric disease, and gave informed consent according to the local ethical committee regulation.

Stimuli and procedure

The experiment consisted of two phases. In the first phase, a series of faces were presented with different emotional expressions in different task contexts to elicit distinct social appraisals associated with each face identity (encoding phase). Next, in the critical test phase, the same faces were presented again, together with new faces, for an explicit recognition task (memory phase).

ENCODING PHASE:

Participants were first exposed to a pseudo-interactive game context in which they were presented with faces from virtual partners as feedback on their performance. The game was a visual dot-counting task, presented to participants as a test of perceptual speed.

Each trial began with a white central fixation-cross on a black screen (for 3 to 7 sec, average 3.5 sec), followed by a brief visual display divided in two parts with a variable number of white dots on each side of the screen (presented for 500 ms; see Figure 1a). The number of dots on each side ranged from 10 to 15. Their quantity and position were randomly assigned on every trial for each side separately, in such a way that the display was never visually identical on both sides. Participants had to indicate which side of the screen contained more dots (right/left) by pressing one of two response-keys. The total number of dots and the difference between the two display sides were adjusted online based on the participant's performance on the preceding trial, by reducing the difference after each correct trial (minimum 1 dot) or increasing the difference after each incorrect trial (maximum 5 dots), allowing us to maintain performance close to threshold and to obtain approximately equal numbers of correct and incorrect trials (mean correct = 57±1% across conditions). In addition, to further ensure this equal distribution (crucial for a subsequent unbiased comparison of conditions), we inserted occasional "trick" displays with 15 dots on both sides whenever performance exceeded 60% correct of two consecutive trials (20 ± 5.6 out of 128 trials). None of the participants noticed these "impossible" trials. The dot display was then replaced by a black screen with a variable interval (jitter of 1000 to 1400 ms, average 1200 ms), during which participants gave their response.

Critically, each response was followed by a feedback screen (1500 ms), consisting of a face that could have either a smiling or angry expression, together with a verbal indication of the actual outcome on the counting task (either "WON" or "LOST"; see Figure 1a). The latter verbal feedback always corresponded to the real performance (correct or incorrect) on the preceding trial (except on the few "trick" trials with equal number of dots on both sides, where a negative "LOST" feedback was given to reduce any excess of correct over incorrect trials). By contrast, the emotional expression of faces was pseudo-randomly assigned on every trial, with the constraints that smiling and angry faces were presented on an equal number of correct and incorrect trials each; and that a given face identity was always seen with the same expression (smiling or angry) and the same feedback message (WON or LOST) throughout the task for any given participant (with different face identities in the different conditions counterbalanced across participants). This design resulted in 4 different combinations of verbal and facial feedback (see Figure 1b), of which two were *congruent*: smiling face on WON trial (SF-W) or angry face on LOST trial (AF-L); and two were *incongruent*: angry face on WON trial (AF-W) or smiling face on LOST trial (SF-L).

We induced a pseudo-social interactive context by telling participants a cover-story along the following lines: faces were those of other subjects included in the study and belonged to two different teams; the participant was randomly assigned to play for the benefit of one team (allied), and hence against the other team (opponent); the goal of the study was to compare perceptual abilities and cooperation among the groups, and performance of the participant determined gains for the allied team but losses for the other opponent team when his/her responses were correct (and vice versa, incorrect performance of the participant led to losses for the allied team but gains for the other team correct). Hence, each correct response gave one point to the allied group, while each incorrect response gave one point to the other group. These outcomes would be reminded during the game by presenting faces from individuals in the different groups with appropriate expressions (either smiling or angry) together with feedback about the participant's performance (won or lost). Subsequent debriefing after scanning indicated that participants accepted the cover-story, were highly motivated by the task, and experienced an expected affective response including "irritation", or "satisfaction" as a function of the incongruent or congruent social meaning of the facial expression paired with feedback (see Vrtička et al., 2008). Importantly, different faces were presented for the different groups of allies and opponents, counterbalanced across participants.

All face stimuli were colour photographs taken from the Karolinska Directed Emotional Faces set (KDEF, Lundquist, Flykt, and Öhmann, 1998). We selected 16 different individuals (8 males) to be shown in the first phase (learning), while 16 others were shown only the second phase (memory recognition task, see below). Each face identity was assigned to one condition only (2 male and 2 female in each of the 4 feedback type for faces seen in phase 1; plus 8 male and 8 female for the new unfamiliar faces subsequently seen in phase 2), with all identities counterbalanced across participants. During the pseudo-game task (phase 1), each of the 16 faces was repeated 8 times in the corresponding condition, in random order, resulting in 128 trials in total per participant (total duration ~15 min).

Imaging results from this phase have been reported separately in relation to the influence of individual personality traits on social behaviour and brain response (Vrtička, Andersson, Grandjean, Sander, & Vuilleumier, 2008).

MEMORY PHASE:

After a short break of ~5 min without any intervening task, participants were given a surprise recognition memory test, in which they had to distinguish between

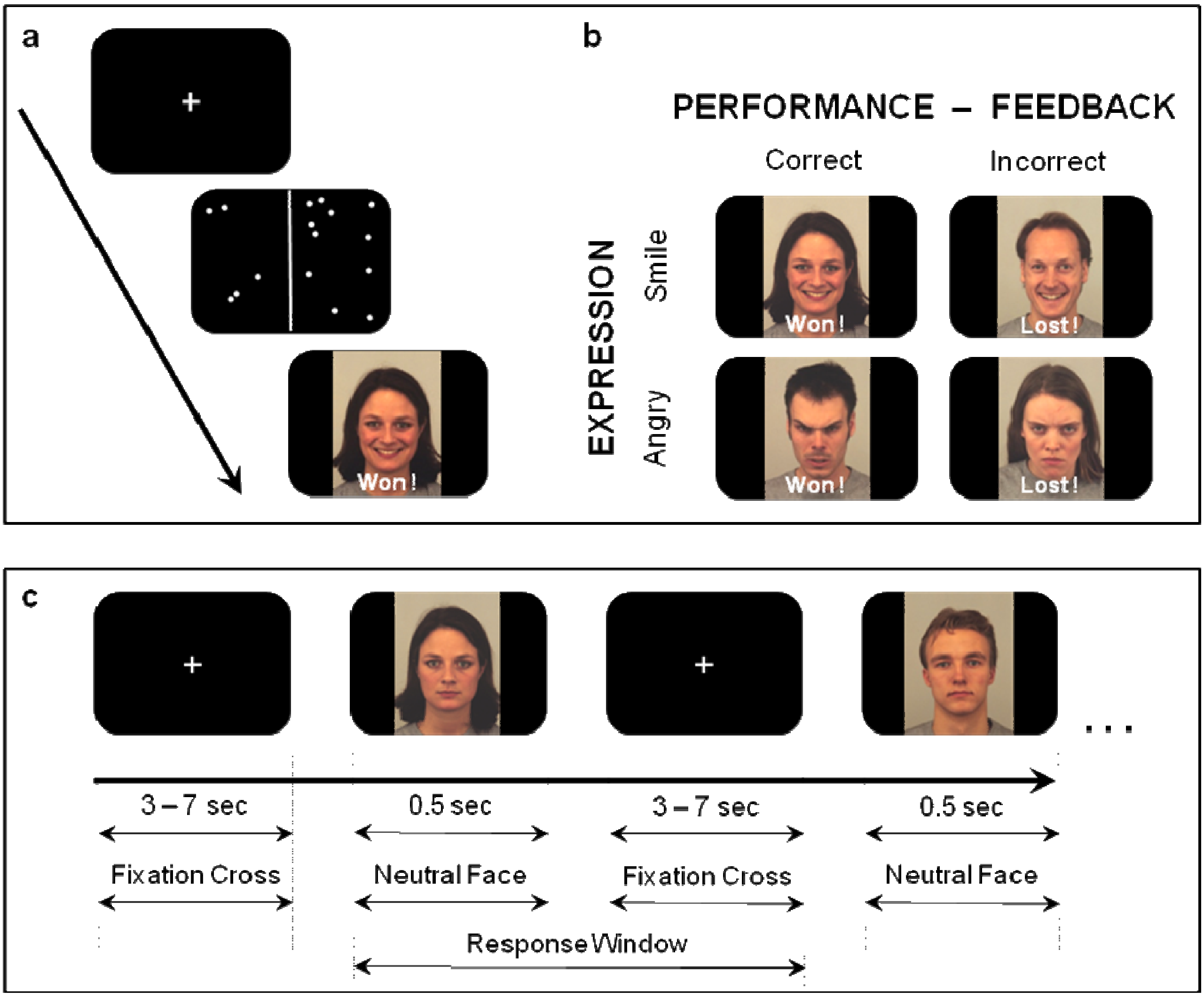


Figure 1: Illustration of the paradigm and the four different feedback conditions
a) *Experimental paradigm during the initial encoding phase. On each trial, participants first saw a central fixation cross, followed by the dot-counting task (0.5 sec), in which they had to indicate the side of the screen that contained more dots (right vs left). Following each response, a visual feedback was shown (1.5 sec), composed of a word together with a face.* ***b)*** *Illustration of the four different feedback conditions during encoding. Two socially "congruent" (Smiling Face on WON trial, Angry Face on LOST trials) and two socially "incongruent" (Smiling Face on LOST trial, Angry Face on WON trial) combinations were possible. Four different face identities (2 female and 2 male) were used in each of these 4 conditions.* ***c)*** *Experimental paradigm during the recognition phase. On each trial, participants first saw a central fixation cross, followed by a face with a neutral expression. They had to indicate whether each face was familiar ("old") or unfamiliar ("new"). See Materials and Methods for further details.*

previously seen faces and new faces ("yes" or "no"). Each trial began with a white central fixation-cross on a black screen (for 3 to 7 sec, average 3.5 sec), followed by a brief visual display showing a single face with a neutral expression (presented for 500 ms). After its offset, participants had to press one of two response-keys to indicate whether the current facial identity had already been presented during the previous game phase ("old" - familiar face condition) or whether it was a novel face shown only during this second memory test phase ("new" - unfamiliar face condition; Figure 1c).

The familiar faces were from the 16 identities already displayed during the encoding phase (phase 1), with the important difference that now all faces were seen with a neutral expression (contrary to the first phase, where they were shown with angry or smiling expressions). Thus, although these faces were initially seen in 4 distinct conditions during the previous pseudo-game, all now were visually similar and required the same response (i.e. "old" judgement), so that only their previous "social history" could make them different from each other or from novel faces. The unfamiliar faces were 16 new identities (also with a neutral facial expression).

All faces (familiar and unfamiliar) were intermingled in a pseudo-randomized order, with the only constraint that there were never more than two successive presentations of new or old faces. Each face identity was repeated 6 times (in different successive order) during this phase, resulting in 192 trials in total per participant (total duration ~16 min). Participants were instructed about these repetitions and instructed to respond "old" only to faces seen during the previous game phase. To avoid fatigue, the memory task was split in two scanning runs (each 8 min).

EYE-TRACKING

Eye movements were monitored continuously during both phases with an MRI-compatible infra-red eyetracker LRO L6 (Applied Science Laboratories, Bedford, MA, USA). Eye position coordinates (x and y) were recorded at 60Hz and saved for offline analysis, allowing us to test for any difference in visual inspection of faces between different conditions, during either encoding or recognition.

POST-SCANNING MEMORY ASSESSMENT:
After scanning, all participants were presented with a "likeability" and "sorting" rating task for "old" faces. For likeability ratings, subjects were shown pictures of these 16 faces (with a neutral expression) and asked to report how likeable each face appeared for them (on a five point scale from -2 to 2). For the sorting task, subjects were again shown these 16 "old" faces (with a neutral expression), and now told explicitly about the four feedback conditions of the encoding phase; they were then asked to classify each face in the corresponding feedback condition in which it was initially presented (4 faces in each of the 4 categories, forced-choice). This allowed us to assess any explicit knowledge for the social context previously associated with each face.

MRI acquisition and analysis

MRI data were acquired on a 1.5 T whole-body INTERA system (Philips Medical Systems), using standard head-coil configuration. For each participant, a structural image was obtained with a 3D-GRE T1-weighted sequence (FOV = 250 mm, TR/TE/Flip = 15 ms/5.0 ms/30°, matrix = 256 × 256, slice-thickness = 1.25 mm) and functional images with a GRE EPI sequence (TR/TE/Flip = 2500 ms/40 ms/80°, FOV = 250 mm, matrix = 128 × 128). Functional images covered the whole brain, composed of 30 contiguous 4mm axial slices parallel to the inferior edge of the occipital and temporal lobes, and acquired continuously for a total of 232 images per participant while they performed the recognition memory task (2 successive runs of 116 images each).

Functional data were analyzed using the general linear model for event-related designs in SPM2 (Wellcome Dept. of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm). All images were realigned, corrected for slice timing, normalized to an EPI template (re-sampled voxel-size of 3 mm), spatially smoothed (8 mm FWHM Gaussian kernel). A high-pass frequency filter (cutoff 120 s) and corrections for auto-correlation between scans were applied to the time series.

Statistical analysis was performed using the general linear model implemented in SPM2, with a separate regressor for each event type convolved with a canonical hemodynamic response function. Ten events from the recognition memory task were modelled, including correct and incorrect trials for the old faces previously seen in each of the 4 possible feedback conditions (SF-W, SF-L, AF-W, AF-L, either correctly recognized or missed); plus the correct and incorrect trials for the new faces (correctly rejected and false alarms, respectively). Due to our splitting of data between correct and incorrect recognition, this distribution resulted in an average of ~15 events for the correct recognition of old faces in each condition. Movement parameters from realignment corrections were entered as additional covariates of no interest to account for residual movement artifacts after realignment. Statistical parametric maps were generated from linear contrasts between the different conditions in each participant. A second-stage random-effect analysis was then performed using one-sample t-tests on contrast images obtained in each subject for each comparison of interest. All contrasts were performed across the whole brain using standard threshold criteria (Worsley et al., 1996) of significant activation at a voxel-level of p< .001 uncorrected and cluster size greater than 5 voxels (135 mm^3). Average parameter estimates of activity (betas) for each feedback condition were extracted from all voxels in regions of interest (ROIs), defined by the full-extent clusters showing significant activation at a voxel-level of p < .001 (uncorrected) in the SPM group analysis (random-effect contrasts).

RESULTS

BEHAVIORAL DATA:
Recognition memory during fMRI scanning required subjects to make "old" vs. "new" judgments on neutral faces corresponding to either familiar identities that were seen in different social context during the encoding phase; or unfamiliar identities that were seen only during the memory phase. Overall memory performance was well above chance [total mean correct: 71.54 ± 1.65%], for both familiar and unfamiliar faces (Figure 2a). Correct rejection rates for unfamiliar faces were higher than correct recognition rates for familiar faces [83.6 ± 2.1% and 59.5 ± 2.3% respectively; t(15)= 7.852, p< .001, paired t-test]. However, there was no significant difference in memory performance between the four familiar face categories (Figure 2b), as determined first by a 2x2 repeated measures ANOVA with expression (EXP; smiling or angry) and feedback (FDB; won or lost) as factors [Fs(15)< .711, ps> .412], and further confirmed by planned paired t-tests [ts(15)< 1.169, ps> .261].

To examine if there were significant differences in the memory rate for some of faces, we performed an additional item-based analysis that showed consistent effects for only 4 out of the 16 different face identities, with one male and one female being remembered significantly better than average, but one male and one female being remembered significantly worse than average (more "hits" and more "misses" relative to the mean for all faces, one-tailed t-test, p< .05), However, more importantly, there were no significant differences as a function of the feedback category (see Figure 2b), because face identities were assigned to the four different feedback conditions in a counter-balanced order, so as to cancel out any visual memory effects related to a few specific faces overall.

After scanning, subjects also performed an explicit "sorting task" in which they had to classify old faces into the corresponding feedback categories of the encoding phase. There was no significant difference in accuracy rates between the four familiar face categories, as first determined by a 2x2 repeated measures ANOVA with EXP (smiling or angry) and FDB (won or lost) as factors [Fs(15)< .952, ps> .345], and further confirmed by planned paired t-tests [all ts(15)< 1.221, ps> .241]. These results indicate that participants did not explicitly recall the exact feedback category in which old faces were seen during the encoding phase. Finally, subjects were also asked to make likeability ratings on old faces after scanning (using a 5-point scale), but again there were no significant differences between the previous context conditions, as determined by a 2x2 repeated measures ANOVA with EXP (smiling or angry) and FDB (won or lost) as factors [Fs(15)< 1.297, ps> .273], and by paired t-tests [all ts(15)< 1.499, ps> .155].

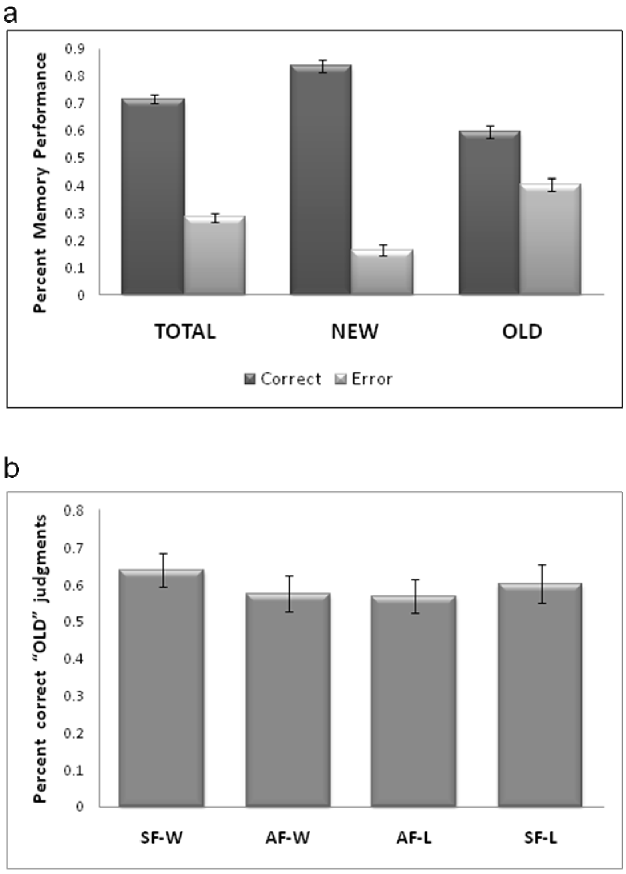


Figure 2: Behavioral memory performance during scanning.
a) *Memory performance during the face recognition phase. TOTAL correct: 71.54 ± 1.65%; NEW correct: 83.58 ± 2.15%; OLD correct: 59.50 ± 2.34%.* ***b)*** *Memory performance across the different conditions of previous feedback context (SF-W: smiling face won, AF-W: angry face won, AF-L: angry face lost, SF-L: smiling face lost) showed no significant differences between the four face categories.*

EYE-TRACKING DATA:

During encoding, there were no significant differences between the number and duration of fixations over the face in the four different social contexts [2 x 2 repeated measure ANOVA: all Fs< 2.56, ps> .14; paired comparisons: all ts< 1.93, ps> .08]. During recognition, there was a significant main effect of feedback on fixation duration [lost > won: F= 7.42, p= .02], but no significant main effect of expression and no interaction [all Fs< 3.86, ps> .075]. To ensure that increased fixation to faces did not modulate BOLD signal during recognition, we correlated the fixation times with beta values extracted from several ROIs identified in subsequent fMRI analysis (see below), but this revealed no significant effect, except for a positive correlation between responses to AF-L faces in left FG and fixation times [r= .607, p= .036]. However, this correlation cannot account for the fMRI results obtained in other conditions. Overall, these eyetracking data therefore indicate that familiarity effects reported below cannot be explained by the visual inspection time for different faces, either during encoding or recognition.

fMRI RESULTS:

We first identified brain regions involved in successful recognition of familiar faces by computing the contrast between all "old" faces *versus* "new" faces, using correct answers only. This contrast allowed us to test for both implicit and explicit memory processes together (unlike a comparison of hits *versus* misses for "old" faces that would primarily test for explicit memory and confidence). Our analysis revealed a network of activated regions (see Table 1) including left fusiform gyrus (FG), right posterior STS, right temporal pole [TP], as well as several limbic areas in left insula [INS], bilateral amygdala [AMY], and posterior cingulate cortex [PCC] (Figure 3). These areas were then used to define regions of interest (ROIs) for further analysis (see below).

We also computed the inverse contrast (new faces *versus* old faces), again for correct answers only, but this showed no significant activations at the same threshold.

Next, we specifically tested whether increases in activity for successful recognition of familiar faces would be modulated by the different context conditions in which faces were previously seen during the encoding phase. Our main goal was to determine any difference in brain responses to faces that were seen as "friends" or "allies" in the past (congruent feedback context, SF-W and AF-L), as opposed to faces that were previously seen as "foes" (incongruent feedback context, SF-L and AF-W). This comparison was performed in two steps, as follows.

In a first step, we extracted parameter estimates of activity (beta values) from ROIs that showed general increases to familiar faces, as determined by the main contrast of all "old" *versus* "new" faces (correct answers only, see above and Table 1). We then split data from the familiar "old" face category into the four conditions of previous encounter: SF-W, AF-L, HF-L, and AF-W, respectively. Repeated-measure ANOVAs were performed on these parameter estimates to test whether activity in the ROIs was modulated by previous emotional expression (smiling or angry), previous feedback type (winning or losing), or any interaction between these two factors (which specifically determined the perceived social context of faces due to expression congruency/incongruency), in addition to direct t-test comparisons to confirm significant increases for each of the "old" face conditions relative to "new" faces.

These analyses revealed that only three of these ROIs were modulated by the social significance of previous encounters. The left FG showed a significant main effect of previous feedback type [lost > won; F(1,15)= 5.97, p= .027] and, more importantly, a significant interaction between feedback type and expression [F(1,15)= 5.05, p= .04]. The latter was due to the fact that faces of "foes" previously seen with a smile on lost trials (SF-L) activated the FG more than all other feedback conditions (Figure 4a). In addition, the PCC showed a marginally significant effect of previous expression [smiling > angry; F(1,15)= 4.34, p= .055], but no effect of feedback [F(1,15)= .87, p= .36] and no interaction [F(1,15)= .34, p= .57] (Figure 4b). Finally, bilateral amygdala showed no significant effect in ANOVAs, but a pattern of selective increases for the two conditions of "foes" (SF-L and AF-W) relative to new (unfamiliar) faces [*left AMY*: SF-L > new, t(15)= 3.42, p= .004; AF-W > new, t(15)= 3.22, p= .006; *right AMY*: SF-L > new, t(15)= 2.58, p= .021; AF-W > new, t(15)= 2.74, p= .015; all paired t-tests] (see Figure 5ab). By contrast, the amygdala response to the two conditions of "allied" faces (SF-W and AF-L) did not significantly differ from responses to new faces (all paired t-tests < 1.92, ps> .075, on both sides]. All other

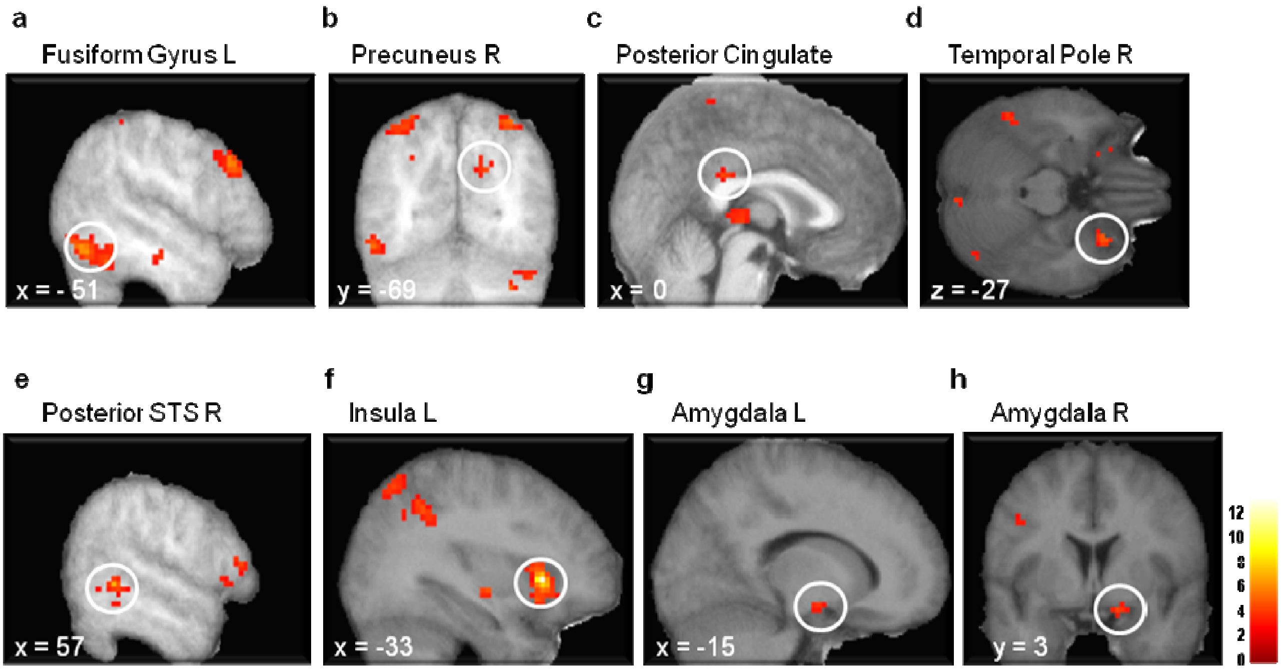


Figure 3: Main effect of face familiarity.
Statistical parametric map for the contrast of "old faces" versus "new faces" (correct answers only), showing an extensive network of brain areas responding to face familiarity. ***a)*** *Left fusiform gyrus (xyz= -51 -63 -18, z-value= 4.26, p< .001),* ***b)*** *Right precuneus (xyz= 18 -69 30, z= 3.40, p< .001),* ***c)*** *Posterior cingulate (xyz= 0 -36 27, z= 3.38, p< .001),* ***d)*** *Right temporal pole (xyz= 33 18 -27, z=3.62, p< .001),* ***e)*** *Right superior temporal sulcus (xyz= 57 -45 -3, z= 4.48, p< .001),* ***f)*** *Left insula (xyz= -33 21 0, z= 6.11, p< .001),* ***g)*** *Left amygdala (xyz= -15 -6 -15, z= 3.05, p< .001),* ***h)*** *Right amygdala (xyz= 15 3 -18, z= 3.33, p< .001). All p-values uncorrected, whole brain analysis.*

ROIs (including right STS, right temporal pole, and left insula) showed a general increase for the old/familiar face conditions as compared with the new/unfamiliar faces, but no differential effect of previous social context.

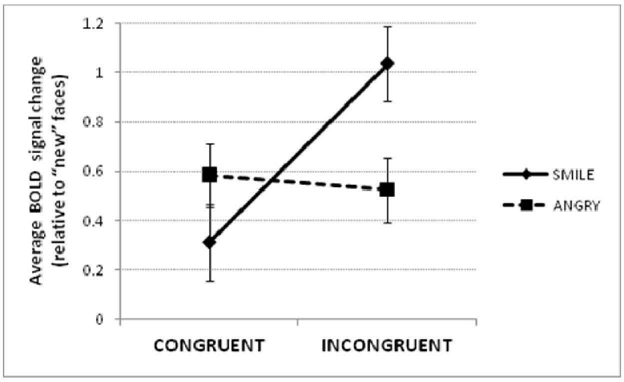
In a second step, we computed specific SPM contrasts on whole-brain data to directly compare the different conditions of old faces with each other. We tested for any main effect of previous emotional expression (EXP; smiling *versus* angry), previous feedback type (FDB; winning *versus* losing), and previous social significance of the face context (congruent *versus* incongruent expression, yielding the perception of "allied" or "friend" faces *versus* "foes", respectively). This analysis revealed two additional regions that were modulated by past impressions, including the left caudate nucleus (Figure 6a) when comparing faces seen on "winning" *versus* "losing" trials, and the anterior cingulate cortex (ACC; Figure 6c) when comparing faces with incongruent ("foes") versus congruent ("friends") expressions.

For the left caudate nucleus (Figure 6b), subsequent pairwise comparisons showed that this increased activation to faces previously seen on WON trials was essentially due to a significant difference for angry faces associated with the incongruent feedback condition (i.e. "foes") [AF-W *versus* AF-L: t(15)= 3.35, p= .004, paired t-test]; but there was no difference between WON *versus* LOST trials for smiling faces [SF-W versus SF-L: t(15)= .77, p= .45]. In contrast, for the rostral ACC (Figure 6d), pairwise comparisons showed that there were significant increases for both angry and smiling faces when previously associated with an incongruent feedback context (AF-W and SF-L, "foes"), with a greater activation to angry faces from WON relative to LOST trials [AF-W *versus* AF-L: t(15)= 2.19, p= .045; paired t-test], but conversely, greater activation to smiling faces from LOST relative to WON trials [SF-L *versus* SF-W: t(15)= 3.65, p= .002].

DISCUSSION

The current study investtigated the neural substrates of social memory for positive and negative person traits, which were learned by exposure to faces associated with different social contextual cues. Face identities seen by the participants were those of strangers who became familiar through a pseudo-interactive social game, allowing us to avoid any knowledge-related or intrinsic emotional differences prior to exposure. After this first encoding phase, face recognition was probed while brain activity was measured by fMRI, and neural responses were compared as a function of the different conditions in which the faces were seen during initial exposure. Importantly, all faces were presented with a neutral expression during the recognition phase, whereas they were repeatedly seen with a distinct emotional expression and paired with a distinct feedback context during initial exposure. Thus, by orthogonally manipulating emotional expression (smiling vs angry) and feedback context (positive vs negative), we could induce a different social meaning for each face identity even when they shared the same expression (e.g. a smile praising success or mocking a failure). This allowed us to distinguish between subsequent effects in memory due to the past emotional expression alone (which might result from relatively automatic perceptual analysis), and subsequent effects due to systematic association with positive or negative context, or any interactions between these two factors that actually determined the perceived social significance of alliance or hostility for the viewer.

a Fusiform Gyrus left



b Posterior Cingulate

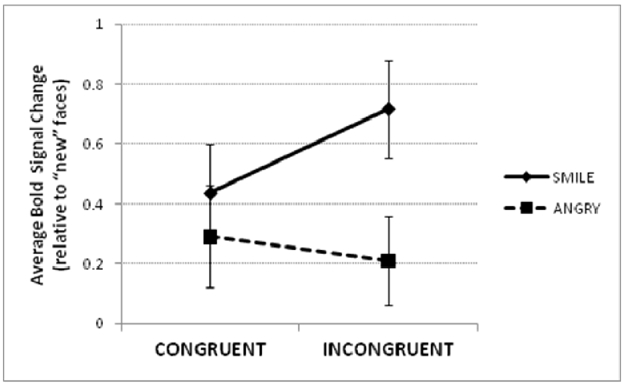


Figure 4: Activation of left fusiform gyrus and posterior cingulate across the familiar face conditions.
a) *Parameter estimates (beta values) extracted from the left fusiform gyrus (contrast "old faces versus new faces"), averaged across voxels and participants, showing a significant interaction between expression and feedback type, with the greatest response to faces previously seen with a smiling expression on LOST trials.* ***b)*** *Parameter estimates (beta values) extracted from the posterior cingulate (same contrast), showing greater responses to faces previously seen with a smiling than angry expression. The average parameter estimates are shown relative to activation for the "new"/unfamiliar faces (by subtraction). All values are displayed with +/- 1 SEM.*

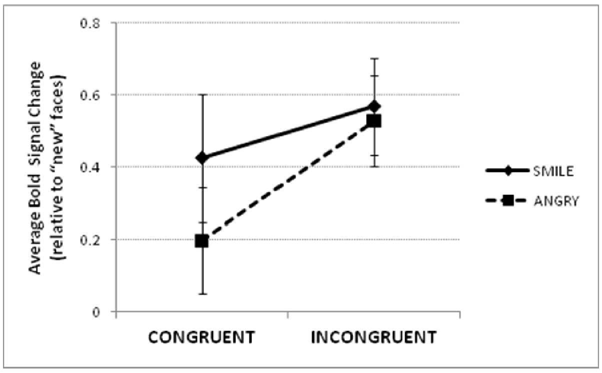
Our results reveal that familiarity of the faces induced by previous encounters in the game task produced increased BOLD responses in a widespread network of brain regions previously reported to be involved in familiar face recognition (Gobbini & Haxby, 2007), including the left fusiform gyrus, right posterior superior temporal sulcus, right precuneus, posterior cingulate, right temporal pole, left insula, and bilateral amygdala. More critically, we found that brain regions within both the "core system" of face perception and "extended system" of person recognition (Gobbini & Haxby, 2007) were generally modulated by familiarity alone, with no or little differences as a function of the past social significance of individual faces (i.e., the four different feedback conditions). Only the left fusiform gyrus showed relatively greater increases for faces perceived as hostile in the past (SF-L condition), while posterior cingulate cortex showed relatively greater increases for faces seen with a smile in the past (SF-W and SF-L). In contrast, subsequent effects of social context were found in brain regions associated with motivational processes and behavioural monitoring, such as the amygdala, caudate nucleus, and anterior cingulate, which showed strongest or even selective responses to faces previously perceived as "foes" (i.e., smiling when the participant failed and frowning when the participant succeeded in the task, SF-L and AF-W conditions, respectively). Although ACC and caudate have traditionally not been integrated in human brain networks for person recognition (Adolphs, 2003; Gobbini & Haxby, 2007), our new data converge with other recent findings in humans (de Quervain et al., 2004; Somerville, Heatherton, & Kelley, 2006) and monkeys (Rushworth, Behrens, Rudebeck, & Walton, 2007) to suggest that these regions may play a crucial role in social cognition and perhaps more generally in affect regulation.

Note that these modulations were not due to previous exposure to different expressions or different feedback types alone; rather, such differences were specifically induced by the interaction between facial expression and feedback context which conveyed a distinctive social meaning to previously seen faces. Likewise, different brain responses to faces encountered in different social contexts could not be explained by differences in eye movements, as online eye-tracking measures did not show any difference between the four feedback conditions – a factor not systematically taken into account in previous studies of familiar face recognition (Gobbini & Haxby, 2007). These new imaging data are consistent with the notion that person impressions can readily be formed even during brief social encounters with people (Bar, Neta, & Linz, 2006; Todorov, Gobbini, Evans, & Haxby, 2007; Willis & Todorov, 2006), and then efficiently transfer to a new view of the same person (despite changes in appearance and expression). Furthermore, these findings also confirm an important role of situational context for inducing such enduring person impressions (Gilbert & Malone, 1995).

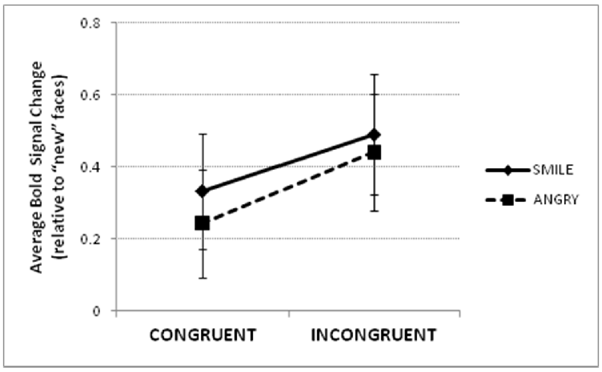
Below we will discuss activation patterns for the most relevant brain regions.

Modulations by familiarity in the "core" and "extended" face processing systems

Among regions showing greater activation to previously seen faces than to new faces, many were essentially sensitive to their past familiarity irrespective of specific social context and induced person impression, such as right posterior superior temporal sulcus (STS), right precuneus, right temporal pole, and left insula. These regions are likely to be implicated in various aspects of person knowledge (Gobbini & Haxby, 2007). In particular, STS has previously been found to adapt to specific combinations of face identity and expression (Winston, Henson, Fine-Goulden, & Dolan, 2004), and to activate when processing cues that imply intentional action (Gallagher & Frith, 2003) or mentalizing about intentions of others (U. Frith & Frith, 2003). Moreover, STS activity has already been related to implicit social judgments during face processing tasks (Gobbini, Leibenluft, Santiago, & Haxby, 2004; Leibenluft, Gobbini, Harrison, & Haxby, 2004; Singer, Kiebel, Winston, Dolan, & Frith, 2004; Todorov, Gobbini, Evans, & Haxby, 2007). Our findings therefore suggest that prior experience with faces in a social context as implemented here during initial exposure may then enhance the perception of intentionality and mental states for these faces, as compared to new faces without any previous personal history. However, such activity was not significantly modulated by the content of mental state attribution, since no statistical difference was observed as a function of the previous context.



a Amygdala left



b Amygdala right

Figure 5: Activation of bilateral amygdala across the familiar face conditions.
a) *Parameter estimates (beta values) extracted from the left amygdala (contrast "old faces versus new faces"), averaged across voxels and participants, showing a greater response to faces previously seen in the two incongruent social context (smiling on LOST trials or angry on WIN trials). b) Parameter estimates (beta values) extracted from the right amygdala showed the same pattern as the left amygdala. The average parameter estimates are shown relative to activation for the "new"/unfamiliar faces (by subtraction). All values are displayed with +/- 1 SEM.*

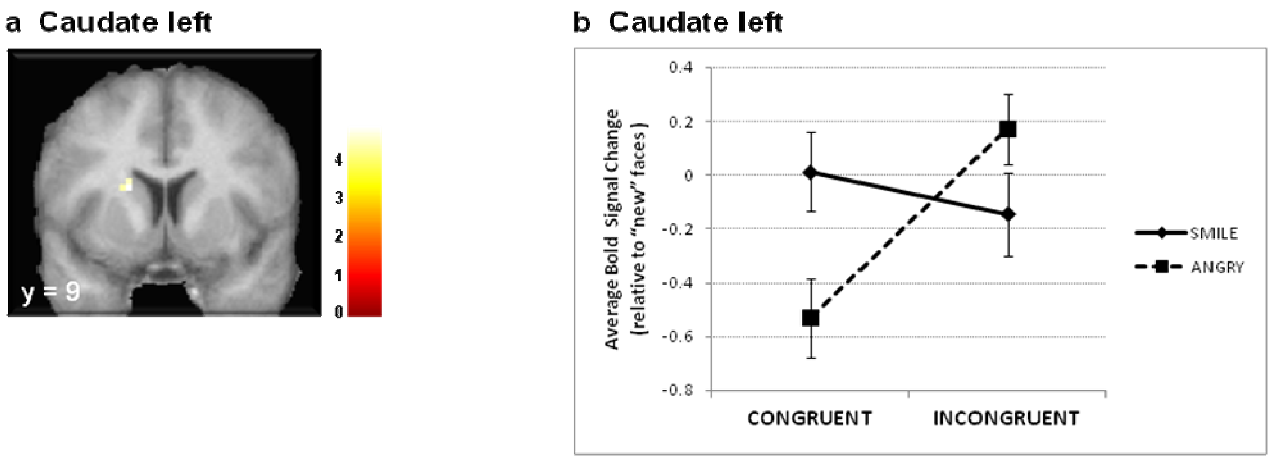
Similarly, the right precuneus and right temporal pole were activated by face familiarity but not by their previous expression, feedback, or contextual associations. Precuneus does not only activate to familiar faces in general (Gobbini, Leibenluft, Santiago, & Haxby, 2004; Leibenluft, Gobbini, Harrison, & Haxby, 2004), but also during mental imagery (Fletcher et al., 1995; Ishai, Haxby, & Ungerleider, 2000), and long-term and episodic memory retrieval (Burgess, Maguire, Spiers, & O'Keefe, 2001; Fletcher et al., 1995). Likewise, anterior temporal regions may respond to various familiar stimuli, including faces (Gorno-Tempini et al., 2000), names (Grabowski et al., 2001), or landscapes (Nakamura et al., 2000). Moreover, lesions in precuneus and/or temporal poles may impair recognition and access to semantic information about people (see e.g. Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996; Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004; Vuilleumier, Mohr, Valenza, Wetzel, & Landis, 2003). These findings suggest that both the precuneus and temporal poles might contribute to familiar face recognition by subserving general processes associated with mental imagery and retrieval of episodic representations in memory, but do not code for more specific social or affective representations influenced by previous context in which faces are encountered.

The left insula was another region modulated by familiarity irrespective of past person impression. Insula responses to familiar faces have already been observed not only for one's own versus unrelated children (Leibenluft, Gobbini, Harrison, & Haxby, 2004), but also for faces previously experienced as cooperators in an economic exchange (Singer, Kiebel, Winston, Dolan, & Frith, 2004) or perceived as aggressive and disgusting based on fictive scenarios (Todorov, Gobbini, Evans, & Haxby, 2007). These findings suggest that insula responses may not differentiate between specific affective values (i.e., positive or negative) associated with familiar faces, but represent a more general affective response related to increased autonomic arousal (Critchley, Melmed, Featherstone, Mathias, & Dolan, 2002) which might be elicited by the recognition of familiarity and/or the retrieval of information about known persons and past encounters with them.

Modulations by person impression in the face processing system

In contrast to the above regions, left fusiform cortex did not only show greater responses to successfully recognized familiar faces, as compared with new faces, but was further enhanced when faces were previously paired with negative feedback (lose *versus* win), and more specifically when this was combined with a smiling expression that induce negative social impressions of hostility ("foes", SF-L condition, cf. Figure 4a). These results suggest that face processing in this condition might have been modulated by top-down influences from other brain areas associated with social or affective memory, based on negative person impressions elicited by past experience. Similar modulations of face processing in visual cortex by familiarity or affective value have already been reported in other studies in humans (Gobbini & Haxby, 2007; Singer, Kiebel, Winston, Dolan, & Frith, 2004) and single-cell recordings in monkeys (Sugase, Yamane, Ueno, & Kawano, 1999). In our experiment, this modulation of fusiform activation appeared particularly driven by the combination of expression and feedback context, rather than by any of these factors alone, suggesting that it could not simply be explained by different visual information perceived during encoding. Such top-down modulation of fusiform activity might possibly originate from amygdala, since this region is critically involved in affective and social processing, as well as in learning (Adolphs, 2003; Dolan, 2002), and thought to modulate fusiform activation to emotional faces (Vuilleumier, Armony, Driver, & Dolan, 2001). In our study, the amygdala and anterior cingulate cortex were also activated by faces associated with negative social context ("foes", see Figures 5-6 and below), and both of these regions might therefore contribute to modulatory signals affecting fusiform activity (directly or indirectly).

Finally, the posterior cingulate cortex (PCC) was another region within the person recognition network that showed general increases to all familiar faces (Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005; Shah et al., 2001), relative to new faces, as well as a modulation by the past social context. PCC activity was significantly enhanced when faces were previously seen with a smiling (*versus* angry) expression. Other studies reported activation of



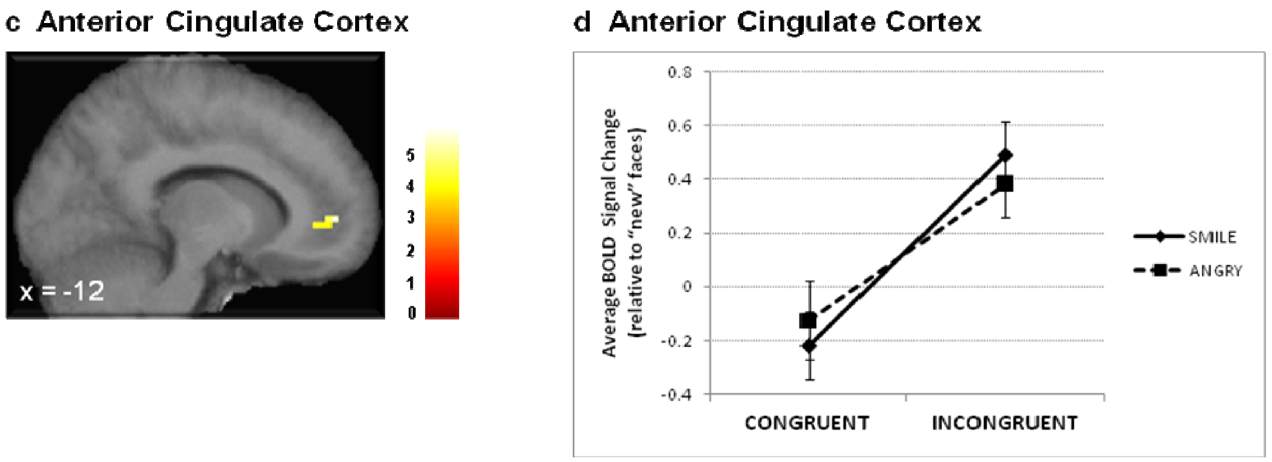


Figure 6: Activation to familiar faces previously seen in different feedback contexts.
a) *Statistical parametric map for the contrast between faces seen on "won versus lost" trials, showing increases in the left caudate nucleus (xyz = -18 9 15; T-Value= 4.93; p< .001; 8 voxels).* ***b)*** *Parameter estimates (beta values) extracted from the left caudate cluster, averaged across voxels and participants.* ***c)*** *Statistical parametric map for the contrast between faces seen with "incongruent versus congruent" expressions, showing increases in the anterior cingulate cortex (xyz = -12 51 0; T-Value= 3.86; p< .001; 23 voxels),* ***d)*** *Parameter estimates (beta values) extracted from the anterior cingulate cluster, averaged across voxels and participants. The average parameter estimates are shown relative to activation for the "new"/unfamiliar faces (by subtraction). All values are displayed with +/- 1 SEM.*

PCC by autobiographical memory retrieval concerning familiar people (Maddock, Garrett, & Buonocore, 2001; Shah et al., 2001), judging emotional valence or content for words and other stimuli (Maddock, Garrett, & Buonocore, 2003), and judging the morality of one's own or someone else's behaviour (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Our results support the notion that PCC might not only be involved in appraising the affective significance of a stimulus (for oneself or others), but also in modulating memory retrieval in relation to affectively relevant stimuli. However, unlike previous data suggesting that PCC activates independent of emotional valence (e.g. for both pleasant and unpleasant arousing words; (Maddock, Garrett, & Buonocore, 2001), our finding of greater activation by previously smiling faces than previously angry faces reveals a bias for positive emotional values associated with people, which might account for reliable responses to personally familiar faces in previous imaging studies (cf. Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005; Shah et al., 2001).

Modulations in emotion and prefrontal systems

A major finding of our study was that bilateral amygdala, left caudate nucleus, and anterior cingulate cortex showed increases to faces that elicited negative impressions of hostility during encoding (SF-L and AF-W conditions), reflecting that a distinctive social value (i.e. "foes") was retained from prior context and modulated activity in these regions during a subsequent encounter with the same identities.

The amygdala was found to activate bilaterally for familiar faces, but more specifically in the two "foe" conditions (see Figure 5ab). This finding might help reconcile apparent discrepancies in the literature on familiar person recognition. On the one hand, several studies reported weaker amygdala response to known versus unknown faces (Dubois et al., 1999), personally familiar versus famous faces and strangers (Gobbini, Leibenluft, Santiago, & Haxby, 2004), or a lover versus a stranger face (Bartels & Zeki, 2004), which was interpreted as lower vigilance or threat when encountering a known person or reduced need to assess social integrity. This interpretation also accords with greater amygdala activation by the perception of untrustworthy faces

OLD CORRECT > NEW CORRECT						
Voxel	*T*	*Z*		*x,y,z {mm}*	*REGION*	*BA*
88	6.87	4.55		-54 24 33	Dorsal PFC left	44
48	4.22	3.38		-42 45 0	Lateral PFC left	45
11	4.13	3.32		51 33 30	Dorsal PFC right	45
21	4.02	3.26		42 9 36	Lateral PFC right	44
12	4.14	3.33		15 3 -18	Amygdala right	
6	3.67	3.05		-15 -6 -15	Amygdala left	
171	13.36	6.11		-33 21 0	Anterior Insula left	
54	6.68	4.48		57 -45 -3	Posterior STS right	21
8	3.77	3.11		-51 -18 -21	Mid STG left	20
29	4.93	3.75		48 -27 -15	Mid STG right	20
62	4.67	3.62		33 18 -27	Temporal Pole right	38
19	4.26	3.4		18 -69 30	Precuneus right	18/23
14	4.24	3.38		0 -36 27	Posterior Cingulate	23
169	6.1	4.26		-51 -63 -18	Lateral Fusiform Gyrus left	37
21	4.77	3.66		-27 -66 33	Occipital Cortex left	19
33	4.57	3.56		45 -75 -33	Occipital Cortex right	19
35	5.31	3.92		33 -69 57	Parietal Superior Cortex right	7
13	4.24	3.39		9 -42 69	Paracentral Cortex right	4/5
31	5.11	3.83		-42 -36 42	Parietal Inferior Cortex left	40
134	4.56	3.56		-48 -42 57	Parietal Infeiror Cortex left	40
15	3.9	3.19		48 -39 42	Supramarginal Gyrus right	40
14	4.31	3.43		36 -51 39	Angular Gyrus right	40
8	5.74	4.11		6 -84 -30	Cerebellum right	
5	4.89	3.73		33 -69 -39	Cerebellum right	
92	5.32	3.93		-6 -18 -3	Thalamus left	

Table 1: SPM results showing brain areas activated by familiar faces during the recognition test phase (main contrast “old” versus “new” faces, for correct answers only). Coordinates are given in MNI space. Activation sites were verified using the average anatomical MRI image of our 16 subjects. BA = Brodmann’s area, PFC = prefrontal cortex, STS = posterior superior temporal sulcus, STG = superior temporal gyrus.

(Winston, Strange, O'Doherty, & Dolan, 2002) or faces from racial outgroups (Phelps, Cannistraci, & Cunningham, 2003), in keeping with a more general role for appraising potential threats (Breiter et al., 1996; Morris et al., 1996). On the other hand, some studies reported increased amygdala activation to pictures of one’s own children *versus* unknown or familiar but unrelated children (Leibenluft, Gobbini, Harrison, & Haxby, 2004), as well as to cooperators relative to defectors in an economic game (Singer, Kiebel, Winston, Dolan, & Frith, 2004), which was attributed to either increased vigilance, protectiveness, or reward expectation for close social relationships. Since in our task, amygdala response was stronger to familiar faces than new faces, but also stronger to familiar faces associated with incongruent social feedback than to those associated with congruent feedback, our results add support to the notion that the human amygdala might primarily encode self-relevant information (Sander, Grafman, & Zalla, 2003) and act to increase vigilance in conditions of ambiguity (Kim, Somerville, Johnstone, Alexander, & Whalen, 2003; Kim et al., 2004) or uncertainty (Hsu, Bhatt, Adolphs, Tranel, & Camerer, 2005). This notion would fit well with our data, because amygdala activity did not appear to track the valence of past expression or past feedback per se, but rather the personal relevance of past experiences with faces.

The left caudate nucleus also exhibited an activation pattern that selectively discriminated faces of “foes” previously seen with an angry expression on WON as opposed to LOST trials (see Figure 6ab). As caudate activity has previously been related to the representation of rewards in social and non social domains (Kampe, Frith, Dolan, & Frith, 2001), this pattern might reflect a rewarding value that is experienced when re-encountering faces of hostile partners who were outplayed during the social game. These findings converge with data from previous studies investigating brain responses related to altruistic punishment (de Quervain et al., 2004) and processing of social hierarchy in humans (Zink et al., 2008). In these two studies, the ventral striatum (including the caudate nucleus) was more activated when subjects could effectively punish defectors in an economic exchange, or moved to a higher hierarchical position based on their performance against a previously superior player. Our data thus provide new evidence that “superiority” information (i.e., winning against the conflicting goals of a virtual opponent in our pseudo-game) is an important component of social memory traces formed after exposure to faces, which may contribute to define the social status of a previously encountered person, and that such information is at least partly dependent on activity of reward circuits in caudate nucleus.

Finally, of particular interest, in addition to regions typically involved in familiar face recognition, we found that activity in the rostral anterior cingulate cortex (ACC) was significantly modulated by the prior social context of faces, and exhibited a selective increase to faces of “foes” from the two incongruent feedback conditions. This region accords well with a functional segregation recently proposed for the medial frontal cortex (MPFC; Amodio & Frith, 2006), including an anterior region (aMPFC) that may be particularly important to access emotional and social knowledge concerning the self or other persons, such as during mentalizing (Frith & Frith, 1999; Frith & Frith, 2003) or introspection processes (Mitchell, Banaji, & Macrae, 2005; Mitchell, Banaji, & Macrae, 2005). The rostral anterior ACC identified in our study overlapped with this aMPFC subregion (Amodio & Frith, 2006), and with increases previously observed when thinking about the mental states of other people or animals (Kelley et al., 2002; Mitchell, Banaji, & Macrae, 2005; Mitchell, Banaji, &l Macrae, 2005; Schmitz, Kawahara-Baccus, & Johnson, 2004). Another study (Somerville, Heatherton, & Kelley, 2006) reported that a similar region (xyz= -6 45 3) was modulated by the perception of social rejection during a pseudo-interactive paradigm in which participants read the judgment of other virtual participants on them (e.g. “This person said she did not like you”). The aMPFC might therefore play an important role in appraisal processes that serve to predict social and/or emotional outcomes, not only for ongoing social events, but also based on memory traces from past encounters.

Because most of the above-mentioned areas showed the highest levels of activation for familiar faces previously encountered in an incongruent social context, our results could potentially also be interpreted as a simple incongruency/congruency effect on memory. However, an incongruency effect would predict similar influences for both types of incongruent faces irrespective of their previous expression (angry or happy), whereas our results indicated a clear difference as a function of expression, in agreement with a role for more specific factors related to affective/social perception during encoding. Moreover, all brain areas showing these differential effects have previously been related to social processing, further suggesting that incongruency per se is unlikely to explain the present activation patterns.

Implicit vs explicit memories of person impressions

Strikingly, differential brain responses to faces were observed depending on prior social context although our behavioural “sorting” test post-fMRI revealed that participants were unable to explicitly classify familiar faces into the correct condition of past context in which they had been seen. Moreover, we found no significant differences in memory performance between the four distinct feedback conditions (cf. Figure 2b). Although it is not possible to draw conclusions from the absence of significance, these data suggest that the differential responses reflecting past person impressions were likely to be implicit or unconscious, which would corroborate the notion that person impressions may be formed automatically without awareness (Uleman, Blader, & Todorov, 2005), and converge with other fMRI data showing implicit biases in brain responses to specific categories of faces (e.g. outgroup race effects; (Phelps, Cannistraci, & Cunningham, 2003) in the absence of explicit acknowledgement by the participant. Our data therefore also add support to behavioural findings (Todorov & Uleman, 2002) that the retrieval of affective person knowledge may arise without any explicit memory for the behaviours that triggered these affective inferences. Accordingly, a recent fMRI study (Todorov & Uleman, 2002) reported that activation of STS by affective person knowledge during face recognition was dependent on explicit memory for past behaviours, whereas activation in insula was implicit (Todorov, Gobbini, Evans, & Haxby, 2007). In our study, however, both regions were activated by familiar faces despite a lack of subsequent explicit memory for the exact context of past encounters; but our participants were engaged in a memory task (old/new recognition) which did not allow us to directly address the exact nature of memory traces. Further research is necessary to delineate a more specific contribution of different brain areas to distinct components of explicit and implicit representations of person impressions in memory, by including additional behavioural measures like i.e. skin conductance responses.

CONCLUSIONS

Our study shows that a wide network of brain areas involved in face recognition is modulated by familiarity of faces derived from past encounters in different contexts and with different expressions, including left fusiform gyrus, right posterior superior temporal sulcus, left insula, right temporal pole, right precuneus, and posterior cingulate cortex. Furthermore, the activation of some of these regions to new encounters of the same face identities is modulated by the positive (friendly) or negative (hostile) nature of past social context, providing a neural substrate for enduring person impression that can influence emotion and behaviour during subsequent presentation with the same face identities. Critically, we provide new evidence that the ventral anterior cingulate cortex, together with amygdala, are key components in neural systems activated by negative person impressions, possibly engaged by retrieval of affective person knowledge so as to adjust emotional responses to faces associated with conflict behaviour during previous social encounters. In addition, we also show that caudate nucleus may code for information concerning social superiority and selectively activates to faces of outplayed opponents, possibly serving to guide behaviour based on the relative social status of previously encountered people. These results do not only contribute to better delineate brain systems underlying social cognition skills in humans, but also help understand neurobiological factors involved in the perception and reaction to social conflicts and aggression behaviours.

ACKOWLEDGMENT

This research was supported by the National Center of Competence in Research (NCCR) Affective Sciences financed by the Swiss National Science Foundation (n° 51NF40-104897) and hosted by the University of Geneva; and grants of the Swiss National Science Foundation to DS and PV.

REFERENCES

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience, 4*(3), 165-178.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature, 393*(6684), 470-474.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268-277.

Armony, J. L., Chochol, C., Fecteau, S., & Belin, P. (2007). Laugh (or cry) and you will be remembered - Influence of emotional expression on memory for vocalizations. *Psychological Science, 18*(12), 1027-1029.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269-278.

Bargh, J. A. (1994). The four horsemen of automaticity: Intention, awareness, efficiency, and control as separate issues. In R. S. Wyer Jr. & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes* (2nd edition ed., Vol. Vol. 1, pp. 1-40). Hillsdale, NJ: Lawrence.

Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. *Neuroimage, 21*(3), 1155-1166.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17*(5), 875-887.

Burgess, N., Maguire, E. A., Spiers, H. J., & O'Keefe, J. (2001). A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *Neuroimage, 14*(2), 439-453.

Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience, 3*(11), 1077-1078.

Critchley, H. D., Melmed, R. N., Featherstone, E., Mathias, C. J., & Dolan, R. J. (2002). Volitional control of autonomic arousal:

A functional magnetic resonance study. *Neuroimage, 16*(4), 909-919.

D'Argembeau, A., & Van der Linden, M. (2007). Facial expressions of emotion influence memory for facial identity in an automatic way. *Emotion, 7*(3), 507-515.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature, 380*(6574), 499-505.

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition, 92*(1-2), 179-229.

de Quervain, D. J. F., Fischbacher, U., Treyer, V., Schelthammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science, 305*(5688), 1254-1258.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience, 8*(11), 1611-1618.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science, 298*(5596), 1191-1194.

Dubois, S., Rossion, B., Schiltz, C., Bodart, J. M., Michel, C., Bruyer, R., et al. (1999). Effect of familiarity on the processing of human faces. *Neuroimage, 9*(3), 278-289.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science, 302*(5643), 290-292.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience, 19*(9), 1508-1519.

Fletcher, P. C., Frith, C. D., Baker, S. C., Shallice, T., Frackowiak, R. S. J., & Dolan, R. J. (1995). The Minds Eye - Precuneus Activation in Memory-Related Imagery. *Neuroimage, 2*(3), 195-200.

Frith, C. D., & Frith, U. (1999). Cognitive psychology - Interacting minds - A biological basis. *Science, 286*(5445), 1692-1695.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 358*(1431), 459-473.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences, 7*(2), 77-83.

Gilbert, D. T., & Malone, P. S. (1995). The Correspondence Bias. *Psychological Bulletin, 117*(1), 21-38.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia, 45*(1), 32-41.

Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage, 22*(4), 1628-1635.

Gorno-Tempini, M. L., Price, C. J., Josephs, O., Vandenberghe, R., Cappa, S. F., Kapur, N., et al. (2000). The neural systems sustaining face and proper-name processing (vol 121, pg 2103, 1998). *Brain, 123*, 419-419.

Grabowski, T. J., Damasio, H., Tranel, D., Ponto, L. L. B., Hichwa, R. D., & Damasio, A. R. (2001). A role for left temporal pole in the retrieval of words for unique entities. *Human Brain Mapping, 13*(4), 199-212.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*(5537), 2105-2108.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science, 310*(5754), 1680-1683.

Ishai, A., Haxby, J. V., & Ungerleider, L. G. (2000). Distributed neural systems for the generation of visual images. *Journal of Cognitive Neuroscience*, 70-70.

Kampe, K. K. W., Frith, C. D., Dolan, R. J., & Frith, U. (2001). Reward value of attractiveness and gaze - Making eye contact enhances the appeal of a pleasing face, irrespective of gender. *Nature, 413*(6856), 589-589.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience, 14*(5), 785-794.

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *Neuroreport, 14*(18), 2317-2322.

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., et al. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of Cognitive Neuroscience, 16*(10), 1730-1745.

Leibenluft, E., Gobbini, M. I., Harrison, T., & Haxby, J. V. (2004). Mothers' Neural Activation in Response to Pictures of Their Children and Other Children. *Biological Psychiatry, 56*(4), 225-232.

Maddock, R. F., Garrett, A. S., & Buonocore, M. H. (2001). Remembering familiar people: The posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience, 104*(3), 667-676.

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human Brain Mapping, 18*(1), 30-41.

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology, 17*(2), 119-128.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage, 28*(4), 757-762.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The Link between Social Cognition and Self-referential Thought in the Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience, 17*(8), 1306-1315.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*(6603), 812-815.

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., et al. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing - A PET study. *Brain, 123*, 1903-1912.

Oda, R. (1997). Biased face recognition in the Prisoner's Dilemma game. *Evolution and Human Behavior, 18*(5), 309-315.

Phelps, E. A., Cannistraci, C. J., & Cunningham, W. A. (2003). Intact performance on an indirect measure of race bias following amygdala damage. *Neuropsychologia, 41*(2), 203-208.

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F. O., & Vuilleumier, P. (2005). View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity: an event-related fMI study. *Neuroimage, 24*(4), 1214-1224.

Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends in Cognitive Sciences, 11*(4), 168-176.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences, 14*(4), 303-316.

Schmitz, T. W., Kawahara-Baccus, T. N., & Johnson, S. C. (2004). Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *Neuroimage, 22*(2), 941-947.

Shah, N. J., Marshall, J. C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H. J., et al. (2001). The neural correlates of person familiarity - A functional magnetic resonance imaging study with clinical implications. *Brain, 124*, 804-815.

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron, 41*(4), 653-662.

Somerville, L. H., Heatherton, T. F., & Kelley, W. M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nature Neuroscience, 9*(8), 1007-1008.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400*(6747), 869-873.

Takahashi, T. (2005). Social memory, social stress, and economic behaviors. *Brain Research Bulletin, 67*(5), 398-402.

Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia, 45*(1), 163-173.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology, 83*(5), 1051-1065.

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology, 39*(6), 549-562.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology, 87*(4), 482-493.

Trope, Y., & Alfieri, T. (1997). Effortfulness and flexibility of dispositional judgment processes. *Journal of Personality and Social Psychology, 73*(4), 662-674.

Uleman, J. S., Blader, S., & Todorov, A. (2005). Implicit Impressions. In J. S. Hassin, J. S. Uleman & J. A. Bargh (Eds.), *The new unconscious*. New York: Oxford University Press.

Vrtička, P., Andersson, F., Grandjean, D., Sander, D., & Vuilleumier, P. (2008). Individual attachment style modulates human amygdala and striatum activation during social appraisal. *PLoS ONE, 3*(8), e2868.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron, 30*(3), 829-841.

Vuilleumier, P., Mohr, C., Valenza, N., Wetzel, C., & Landis, T. (2003). Hyperfamiliarity for unknown faces after left lateral temporo-occipital venous infarction: a double dissociation with prosopagnosia. *Brain, 126*, 889-907.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592-598.

Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology, 92*(3), 1830-1839.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5*(3), 277-283.

Zink, C. F., Tong, Y. X., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know your place: Neural processing of social hierarchy in humans. *Neuron, 85*(2), 273-283.

VI.3. Behavioral Study

In Preparation for Publication

The behavioral study primarily served to obtain independent ratings of social *versus* non-social emotional stimuli for further use in fMRI experiments N° 2a and 2b. However, as individual AAS scores of all participants could be collected rather easily, it was also analyzed regarding the impact of AX and AV on three different rating scores. Accordingly, the main research question for the behavioral study was:

- **How does AAS modulate the subjective attribution of likeability, intensity and controllability to positive and negative social *versus* non-social emotional scenes?**

Very few previous studies have assessed the subjective experience of such stimulus properties as likeability (valence), intensity (arousal) or control as a function of AAS. One study [193] reported some effects of AV and AX on subjective experience of the intensity of negative and positive film-clips (see Chapter IV.4.6.), but neither this nor other similar stimulus properties in the context of AAS have systematically been investigated up to date.

The results of the behavioral study showed that AV selectively affected the pleasantness (but not intensity and control) ratings of social positive images: the higher subjects scored on AV, the lower their pleasantness scores to this distinct stimulus category. What is concerning AX, the data revealed higher intensity but lower control ratings for social negative (but not positive) emotional images, but also increased intensity ratings for both positive stimulus categories as a function of high AX. Such rating patterns are consistent with AAT and also previous fMRI findings in the context of this doctoral thesis (see Chapters IV.4.5. and VI.2.), implying decreased positive experience of social positive scenarios by people scoring high on AV, and higher intensity experience but lower control capabilities of social negative scenes by high anxious participants. What is concerning the findings of higher intensity ratings of both positive emotional image categories as a function of high AX, not as much evidence bolstering such accounts is available. However, as a central feature of AX is ambivalence, it is possible that people scoring high on AX experience both positive as well as negative (social) scenes as more intense. In any case, more research is needed to clarify particularly the second issue.

Because only AAS scores of the participants were obtained for the behavioral study, the results are not as clear as expected, and no additional measures (i.e. brain correlates) could be collected, these findings were not published yet and currently evaluated if suitable for an ultimate publication. Nonetheless, they are considered important enough to be an integral part of this doctoral thesis.

The Influence of Adult Attachment Style on the Perception of Social and Non-Social Emotional Scenes

Pascal Vrtička (1,2), David Sander (1,3), & Patrik Vuilleumier (1,2)

(1) Swiss Center for Affective Sciences, University of Geneva
(2) Laboratory for Neurology & Imaging of Cognition, Department of Neurology and Department of Neuroscience, University Hospital and Medical School, University of Geneva
(3) Department of Psychology, FPSE, University of Geneva

ABSTRACT

Attachment theory describes how people emotionally bond with others and how these social interactions are modulated by so called secondary attachment strategies, either involving hyper-vigilance (anxious attachment) or de-activation (avoidant attachment). However, it still remains unclear whether individual differences in attachment style influence the appraisal of negative and/or attachment-related information only, or whether they may also affect the processing of positive and/or attachment-unrelated content. Thus, it is unknown whether anxious and/or avoidant attachment modulates emotional ratings for negative and positive scenes as a function of their social content. Accordingly, the present study tested how social and non-social positive or negative emotional scenes were subjectively rated on pleasantness, arousal, and control as a function of individual attachment style differences.

Our data revealed that avoidant attachment had a selective impact on pleasantness ratings of social positive images. This finding extends previous literature by showing that attachment style can also have an effect on the subjective experience of positive images, particularly if they contain social information. Moreover, anxious attachment was found to increase intensity but decrease control ratings for social negative images specifically, and lead to higher intensity ratings for both positive stimulus categories. Whereas the former data are supported by attachment theory characterizing anxious attachment with hyper-vigilance to cues of social threat, the latter results could be explained by the notion of ambivalence as a second hallmark of anxious attachment style.

INTRODUCTION

Attachment refers to the formation and maintenance of emotional bonds with another person, and is shaped by individual personality factors that reflect one's needs and expectations concerning social interactions, as well as more general responsiveness to emotionally significant events [1]. Behavioral research in developmental and social psychology has suggested that humans are endowed with a highly organized attachment system which is strongly dependent on early experiences of children with their caregivers, and whose main function throughout life is to maintain proximity to significant others in times of stress, or in other words, to regulate support-seeking behavior [2, 3]. Moreover, because attachment defines lasting feelings of “psychological connectedness between human beings" [2], it is thought to have a tremendous impact on social processes that continues into adulthood [1-3] and can thus influence how adults perceive and respond to emotional information in a wide range of contexts. Even though the exact links between adult and developmental aspects of the attachment system still remain unresolved [4], it has been proposed that early experiences with primary caregivers during childhood may later give rise to internal working models (IWMs) of the self and others, that become part of a person's implicit procedural knowledge about socio-affective situations and then govern thoughts, feelings, and behaviors throughout the lifespan [1].

Following the early conception of different attachment patterns in infants [5], researchers have extended attachment theory to adults, and identified four main categories of “attachment styles” that define individual differences in social interactions and emotional responses to others [6]: secure, anxious-preoccupied, dismissive-avoidant, and fearful-avoidant (or disorganized). Individuals with secure attachment (SA) report positive and trustful interactions with others; whereas attachment anxiety (AX) is characterized by a strong need for closeness, worries about relationships, and fears of rejection; and the hallmarks of attachment avoidance (AV) include a preference for self-reliance and emotional distance from others. More recently, a bi-dimensional theory has been put forward [7-10], in which these different categories of attachment style are mapped on two basic dimensions defined by orthogonal axes of anxiety and avoidance. Thus, in this model, attachment security corresponds to individuals in whom

---

CORRESPONDING AUTHOR:

Pascal Vrtička
Swiss Center for Affective Sciences
7, rue des Battoirs, 1205 Geneva, Switzerland
Tel: +41 – (0)22 379 9824 Fax: +41 – (0)22 379 9844
Email: pascal.vrticka@unige.ch

both anxiety and avoidance are low. A bi-dimensional organization of the attachment system is supported by recent functional brain imaging studies showing that AX correlates with selective increases in the amygdala response to negative social stimuli (angry faces), whereas AV correlates with activation of the ventral striatum and ventral tegmental area to positive social stimuli (smiling faces) [11, 12].

According to the *control systems model* of the attachment system in adulthood [13], three main modules underlie the functioning of the attachment system. The first module is responsible for the activation of the attachment system through the monitoring and appraisal of signs of social threat. The second module involves the monitoring and appraisal of the availability of an attachment figure, which is closely associated with the sense of felt security. Finally, the third module involves the regulation of proximity-seeking as a way of dealing with perceived attachment insecurity, and operates by either hyperactivating or deactivating strategies to resolve such insecurity. Whereas hyperactivation of affective response to others is associated with AX, deactivation is related to AV. Importantly, the *control systems model* also includes reciprocal influences between modules, such that a recurrent use of hyperactivating or deactivating strategies can affect the appraisal of social threats (first module) and attachment-figure availability (second module). As a consequence, anxiously attached people – who employ hyperactivating strategies and maintain the attachment system in a chronically active state – also tend to exhibit heightened appraisal and enhanced attention towards threatening events. In turn, avoidantly attached people’s tend to use deactivating strategies and maintain their attachment system in a down-regulated or inactive state, which may then lead them to ignore or dismiss threatening events, and to deny the need for an attachment figure availability (for review, see [1]). Moreover, recent neuroimaging data suggest that individual differences in the functioning of the first module might be subtended by changes in brain areas associated with emotional processing, such as the amygdala and ventral striatum [11, 12], whereas differences in the other two modules might implicate prefrontal regions associated with emotional regulation and social conflicts [11, 14] as well as somatosensory areas responsible for emotional contagion and recognition [15].

In the present study, we specifically investigated emotional processes linked to the first module of the *control systems model*, and tested whether individual differences in attachment system activation might influence the appraisal of emotional information with or without social significance. According to attachment theory, the attachment system is activated only in threatening conditions, but the triggering events could be either attachment-related (e.g., the loss of a loved one) or attachment-unrelated (e.g., receiving a frightening medical diagnosis). This has been found in particular for hyperactivating or deactivating strategies linked to an anxious or avoidant attachment style, respectively, because even attachment-unrelated distress-eliciting material may reactivate representations and needs for social proximity and protection [13, 16, 17]. However, despite abundant research on attachment system activation in social conditions, very few studies have directly investigated the monitoring and processing of more general emotional material. In particular, it still remains unclear whether individual differences in attachment style influence the appraisal of negative and/or attachment-related information only, or whether they may also affect the processing of positive and/or attachment-unrelated content. Moreover, although recent brain imaging results have shown that attachment style correlates with differential responses of emotional brain areas to faces [11, 15] and sentences with attachment-related meaning [12, 14], it remains unknown whether these findings would extend to other (non social) material, and whether they may translate into behavioral differences in subjective emotional judgments. Here, we therefore examined how individual attachment style influences emotional ratings for negative and positive scenes as a function of their social content.

Previous behavioral studies investigating emotional processes in relation to attachment have generally focused on either vigilance or attentional monitoring for verbal material or emotional facial expressions. For instance, a study by Mikulincer et al. [18] showed that highly anxious people were faster to make lexical decisions in response to the names of their attachment figures. In another study by Fraley et al. [19], highly anxious subjects perceived the onset as well as the offset of emotional facial expressions in a morph movie paradigm at earlier time-points relative to people with secure attachment. However, no specific patterns could be found for avoidant people, and no data was collected on attachment-unrelated cues in neither of these studies. Other studies on attentive processing using a classic dot-probe paradigm [20, 21] reported that attachment anxiety was associated with a selective hypervigilance toward attachment names, while both attachment anxiety and avoidance were associated with an attentional bias away from general threat-related words. However, in fact, the best predictor for this attentional aversion was the interaction between attachment anxiety and avoidance. These results confirm that attentional processing may differ as a function of secure (low anxious and low avoidant) versus insecure (high anxious and high avoidant) attachment [22, 23], but these effects could result from differential responsiveness at either the appraisal or regulatory stages in the control systems model of attachment [1]. Moreover, attention may be automatically captured by threat-related stimuli in the absence of negative evaluation in explicit or implicit emotional judgments [24]. Therefore, these previous behavioral results did not disentangle the influence of specific attachment dimensions, i.e. anxiety or avoidance, on the subjective appraisal of emotional information. Only one recent study using emotional video-clips with attachment-related content [25] showed that anxious individuals experienced negative emotions of fear and sadness as more arousing as compared with secure individuals, whereas avoidant individuals experienced positive emotions as less arousing than the other two groups. Although this study provides new clues about the effect of attachment system activation on emotional appraisal, it did not distinguish between responses to attachment-related and -unrelated image content.

The goal of our study was to obtain measures that were sensitive to both avoidance and anxiety, and to investigate their influence on the appraisal of more general emotional cues, by using visual scenes with both attachment-related and unrelated content (i.e. social vs non-social), and both positive and negative emotional significance. For this purpose, we asked participants to rate different visual scenes on three standard rating scales, widely used in the field of emotion research [26], which assessed the appraisal of pleasantness, arousal, and control. This enabled us to measure not only the perceptual and relatively automatic aspects of emotion processing (pleasantness and arousal), but also the more cognitive and regulatory aspects (control) that might be modulated by different patterns of attachment system activation during affective appraisals. Importantly, our stimuli made no explicit reference to particular attachment figures (e.g. parents, romantic partners…) or specific attachment concepts (e.g. love, trust...), but rather probed for any effect reflecting an implicit (or unconscious) activation of attachment system when processing general emotional scenes.

METHODS

SUBJECTS

We recruited 57 participants from second-year psychology students at Geneva University (average age 23.57 years, 55 female). Only women were included in the final analysis to ensure homogenous effects without gender-related differences, because of evidence that emotions are typically more intense and more prone to regulation in women (e.g. [27]). In addition, data from one female subject was removed at a later step because of incoherent response pattern (see Stimuli and Rating). Therefore, data from 54 subjects (all female) were included in the final analysis.

ATTACHMENT QUESTIONNAIRE

Individual attachment style was assessed in all subjects using a French version of the Relationship Scale Questionnaire (RSQ; [28], translated and validated in previous work by N. Guédeney (2005). Following recent recommendations in a review paper by Kurdek et al. [8], the RSQ was analyzed according to a specific dimensional model of attachment [10], by calculating two separate scores for AVOIDANCE (AV) and ANXIETY (AX) for each subject, based on 13 items (8 for AV and 5 for AX) out of the total 30 items in the questionnaire. Values were then centered (z-scores) to reduce effects of multicollinearity [29].

STIMULI and RATING

We created a set of 360 emotional pictures that were chosen either from the International Affective Pictures System (IAPS) [30] or from images on the internet, and then sorted according to a 2x2 factorial design, with SCENE CONTENT (SC; either social or non-social, i.e. attachment-related or -unrelated) and VALENCE (VAL; either positive or negative) as factors. This gave rise to four different experimental categories: Social Positive (SP) or Negative (SN) scenes, plus Non-Social Positive (NSP) or Negative (NSN) scenes. Pictures belonging to the social category displayed scenes with a clear social implication mostly depicting two or more individuals, such as two people fighting (negative VAL), or a mother interacting with her baby (positive VAL). In contrast, images belonging to the non-social category depicted scenarios with no people directly involved, like a dead bird covered in oil (negative VAL), or a tropical island scene (positive VAL). All pictures were in colors and appropriately adjusted to obtain similar low-level properties (e.g. size, luminance, and pixel resolution).

From the entire image set, two separate lists of 180 pictures were generated, each including 45 pictures per experimental category. Subjects had to rate all emotional pictures from one list in two sessions, each lasting approximately 15 minutes. Each picture was presented on a computer screen for 3 seconds, and followed by a rating screen that remained visible until a response was made. An empty interval of ** seconds preceded the onset of the next picture. Ratings were made using a continuous scale (from 0 to 100) for three different dimensions [26]: PLEASANTNESS (PLN; from “very negative” to “very positive”), INTENSITY (INT; from “low arousal” to “high arousal”), and CONTROL (CON; “To what extent were you able to control the induction of emotion by the previous picture?”, from “absence of control” to “presence of control”). Since there were no significant differences between rating scores for PLN, INT, and CON between the two lists, our subsequent analysis pooled the rating responses across all 360 emotional pictures and all subjects without keeping this list factor.

By inspecting the raw data, we noticed that one female subject showed abnormal rating scores on the pleasantness scale (with only a small and negative difference between pleasant and unpleasant pictures [mean -1.66 ± 1.09], unlike the large and positive difference for all other subjects [mean ≥ 17.65 ± 0.64; average 51.19 ± 15.41]). Therefore, this subject was excluded from final data analysis (total remaining: n = 54).

ANALYSIS

For each of the rating scores (PLN, INT, CONT), we performed repeated-measure ANOVAs with factors of Scene Content (SC) and Scene Valence (VAL), across all participants, as well as bivariate followed by partial correlations using individual indices for the different dimensions of attachment style (AV, AX).

RESULTS

Relationships scales questionnaire (RSQ)

The mean scores for each attachment dimension across all participants were 19.08 ± 4.97 for AV and 10.96 ± 3.92 for AX. This corresponds to previous findings for a healthy unselected population [8]. There was a mild positive association

PLEASANTNESS rating

We first computed a 2x2 repeated measures analysis of variance (ANOVA) on pleasantness ratings, with VAL and SC as factors, across all participants. Only a significant man effect of VAL [F(1,53)= 5958, p< 0.001], because positive images were rated significantly more pleasant than negative images (Figure 1a) – as expected. There was no main effect of SC [F(1,53)= 1.21, p= .277] and no interaction [F(1,53)= 2.38, p= .121], indicating that our image set was successfully balanced in valence across the different scene contents.

Next, we computed the bivariate correlations between PLN ratings and the different attachment style indices. This analysis revealed a specific negative correlation between the PLN ratings for social positive (SP) images and avoidant attachment style [AV; Pearson r= -.292, p= .032] (Figure 1b). In other words,

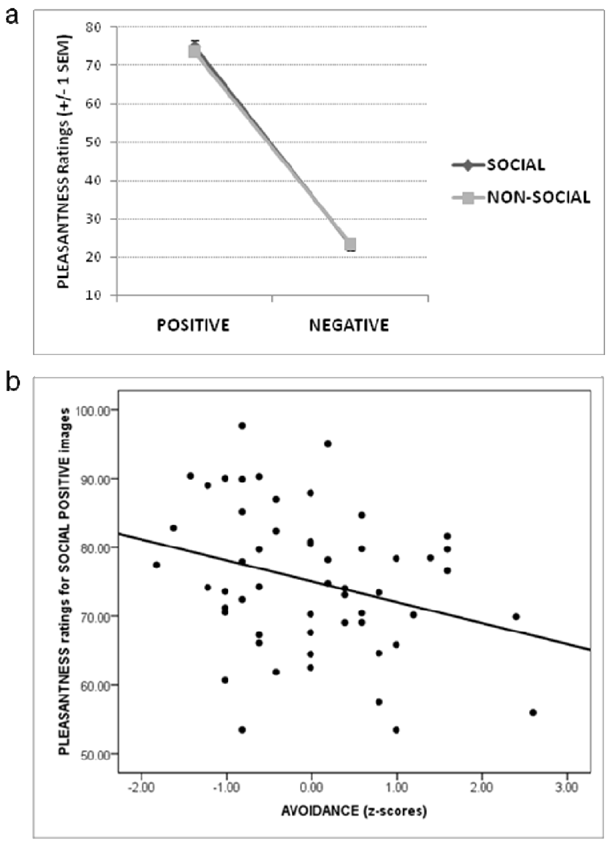


***Figure 1: a)** Main effect of picture valence on the pleasantness ratings, showing that positive images were rated as more pleasant than negative images. **b)** Significant negative correlation between avoidance scores and pleasantness ratings for social positive images.*

the higher subjects scored on AV, the less pleasant they specifically rated the SP images. In contrast, there were no significant correlations between AV and the PLN ratings for the three remaining image categories [Pearson-rs< ± 0.185, ps> .181]. Likewise, no significant correlations were found between AX and PLN ratings for any of the image categories [Pearson -rs< ± 0.151, ps> .276]. Finally, to ensure that the found negative correlation between PLN ratings for SP images and AV was independent of AX, we computed an additional partial correlation, which remained significant even if controlling for AX [r= -.289, p= .036].

INTENSITY rating

As above, a 2x2 repeated measures ANOVA was first computed across all participants. This analysis revealed (1) a main effect of VAL [F(1,53)= 1.474E2, p< 0.001], because negative images were rated as more intense than positive images; (2) a main effect of SC [F(1,53)= 8.376, p= .006], because social images were rated as more intense than non-social images; and (3) a significant interaction between VAL x SC [F(1,53)= 11.138, p= .002], because the difference in intensity ratings between social and non-social images was present only for positive stimuli [t(54,1)= 3.45, p= .001; paired t-test], but not for negative stimuli [t(54,1)= .27, p= .79] (Figure 2a).

The bivariate correlations between INT ratings and attachment indices showed (marginally) significant effects for the anxious dimension only. That is, the higher subjects scored on anxiety (AX), the more intense they rated the two categories of positive images [SP: Pearson r= .262, p= .056; NSP: Pearson r= .30, p= .028], and the more intense they rated the social negative images [SN; Pearson r= .272, p= .047] (Figure 2b). There was no significant correlation between AX scores and INT ratings for NSN images [Pearson-r= .204, p= .138]. Similarly, no correlations were found between AV scores and INT ratings for all image categories [Pearson-rs< ± .129, ps> .353]. Again, to ensure the specificity of the found relations regarding AX, additional partial correlations were derived by controlling for AV, confirming the significance of the effects found for SN [r= .314, p= .022] (increased significance, indicating an even stronger influence of AV if AX is controlled for) and NSP [r= .304, p= .027], and a marginally significant effect for SP [r= .253, p= .067].

CONTROL rating

For control ratings, the 2x2 repeated measures ANOVA showed (1) a significant main effect of VAL [F(1,53)= 46.517, p< .001], because positive images were rated as more controllable than negative images; and (2) a significant main effect of SC [F(1,53)= 4.283, p= .043], because social images were rated as more controllable than non-social images (Figure 3a). There was no interaction [F(1,53)= 1.12, p= .295].

Bivariate correlations with attachment indices revealed a marginally significant negative correlation between AX and control for SN images [Pearson r= -.233, p= .065], meaning that the higher the subjects scored on AX, the less controllable they rated the SN images (Figure 3b). There were no significant correlations between AX and the CON ratings for the three remaining image categories [Pearson-rs< ± 0.136, ps> .327], and no correlations between AV and CON ratings for all image categories [Pearson-rs< .058, ps> .676]. Once more, the partial correlation analysis confirmed – and even increased – the specificity of the found correlation between CON ratings for SN images and AX, even if controlled for AV [r= -.276, p= .046].

DISCUSSION

This study asked the simple but important question of whether adult attachment style has an influence on the appraisal of emotional stimuli, even when these make no explicit reference to attachment-related concepts. We measured the ratings for pleasantness, intensity, and control on a large set of emotional pictures, as a function of their valence (positive or negative) and social content (social or non-social), in a 2x2 factorial design, and found a significant effect of adult attachment style that modulated these rating scores in a systematic manner across individual participants. Our data revealed that attachment avoidance specifically influenced pleasantness ratings for social (and thus attachment-related) positive images, whereas attachment anxiety modulated the intensity ratings (and to a lesser extent the control ratings) for all social images (positive and negative) as well as for positive (but not negative) non-social images. The latter pattern of responses in anxious subjects might reflect a more general effect of ambivalence (see e.g. [31]) when these individuals judged the emotional intensity of social scenes (irrespective of valence) or positive scenes (irrespective of their social content). These distinctive behavioral findings for avoidant and anxious attachment provide the first evidence to our knowledge that adult attachment style may have a pervasive impact on

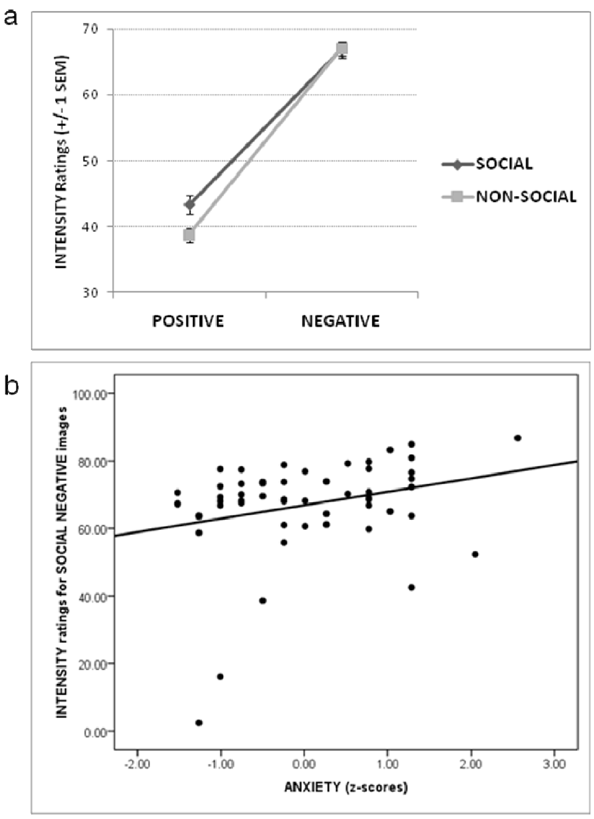


***Figure 2**: **a)** Significant interaction between picture valence and social content for the intensity ratings, reflecting a significant difference in reported intensity for positive [SOC > NSOC; (54,1)= 3.45, p= .001; paired t-test], but not negative [t(54,1)= .27, p= .79] images. **b)** Positive correlations between anxiety scores and intensity ratings for social negative [SN; Pearson r= .272, p= .047], social positive [SP: Pearson r= .262, p= .056], and non-social positive images [NSP: Pearson r= .30, p= .028] (from left to right).*

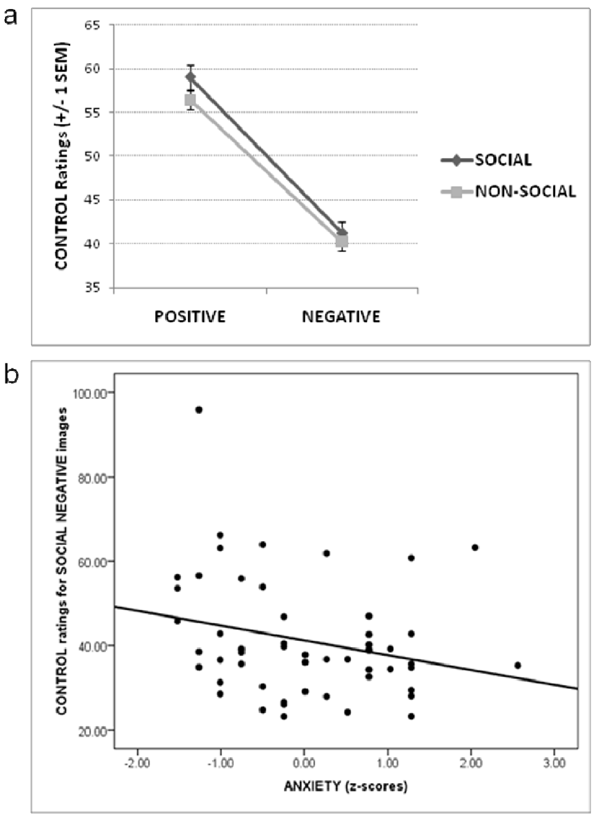


***Figure 3: a)** Main effect of valence [POS > NEG; F(1,53)= 46.517, p< .001] and social context [SOC > NSOC; F(1,53)= 4.283, p= .043] for the control ratings. **b)** Marginally significant positive correlation between anxiety scores and control ratings for social negative images [Pearson r= -.233, p= .065].*

the perception of both social and non-social emotional scenes, but appear consistent with current models of the adult attachment system [1]. Our results also extend the recent behavioral [19-21] and neuroimaging studies [11, 12, 14] that demonstrated strong influences of attachment style on emotional stimulus processing, but always focused on faces or words with socially relevant meaning.

In our study, avoidant attachment style (AV) was found to have a specific effect on pleasantness ratings for social positive images (SP): the higher subjects scored on AV, the less pleasant they rated SP images. There were no other relations between AV and any other stimulus condition or any other aspects of emotional ratings (i.e. perceived intensity and control). In addition, even though there was a negative correlation (r= -.351, p= .009) between pleasantness and control ratings for SP images in general (meaning that the more pleasant subjects rated SP images, the less controllable they judged them to be), our data only showed a significant effect of AV on pleasantness but not control ratings, suggesting not only that this relation was selective, but also that higher pleasantness of SP scenes did not entail judgments of reduced control in avoidant subjects, unlike we observed for the overall sample. Generally speaking, these findings converge with previous research reporting a relative absence of positive behaviors and experiences in avoidant individuals in the context of social relationships (e.g. see [32]). Moreover, avoidantly attached individuals report fewer positive emotions [33] and judge everyday social interactions as more boring and unengaging [34], as compared with securely attached individuals. In addition, avoidant individuals were found to rate positive emotions induced by movie clips with attachment-related content as less arousing than participants with other attachment styles (secure, preoccupied, and fearful/avoidant) [25]. These effects have previously been related to deactivating strategies that avoidant individuals employ to minimize the impact of positive emotions associated with social encounters and attachment needs [1].

Importantly, however, our new data uncovers two important hallmarks in the pattern of decreased positive emotion experience in avoidant individuals. Firstly, their reduced judgment of pleasantness seems to be highly specific to social information, and hence attachment-related representations, without any general effect on their appraisal of other positive stimuli. Secondly, their reduced experience of pleasantness for SP images did not entail an increase in ratings of control for this condition, as could have been expected from traditional attachment theories according to which avoidance primarily depends on deactivation strategies and emotional suppression. Instead, the current results point to a direct effect of avoidance on appraisal of positive emotions in social scenes.

These findings thus nicely dovetail with two recent functional neuroimaging studies that reported a selective correlation between AV and neural activity in brain areas associated with emotional processing. In the first fMRI study [11] using a pseudo-interactive game paradigm where feedback on performance was conveyed by pictures of smiling or angry faces from virtual partners, participants with higher AV were found to exhibit a weaker activation of reward circuits (ventral striatum and ventral tegmental areas) in response to the socially supportive faces (i.e. smiling to convey positive feedback). This modulation by AV style was specific to the perceived social meaning of face expressions in relation to current task goals, because no such effects was produced in ventral striatum by smiling faces presented in a different feedback context, therefore suggesting a selective impassiveness of avoidant individuals to rewards with social meaning (rather than to any type of reward). In the second fMRI study [15], higher AV was found to correlate with a reduced activation of the primary somatosensory cortex to masked faces with a sad expression, possibly indicating a reduction in automatic emotional contagion and recognition functions mediated by somatosensory cortices [35].

In contrast to the above, our data revealed selective effects of attachment anxiety (AX) on ratings of emotional intensity across several image conditions, but not on their pleasantness. Thus, the higher subjects scored on AX, the more intense they rated both social and non-social positive (SP and NSP) images, and the more intense they rated social negative (SN) images. Perceived intensity of non-social negative images was unaffected by AX. In addition, there was a marginal negative correlation between AX and judgments of control for SN images (i.e. the higher our subjects scored on AX, the less controllable they rated SN images), Interestingly, across the whole sample, we also observed a general negative correlation between control and intensity ratings, which indicated that more intense images were generally rated as less controllable. Nevertheless, only negative images with social content displayed significant correlations with AX, for both intensity and control ratings, thus implying a specific effect of anxious attachment style on the perception of social and thus attachment-related stimulus content. These data agree with previous accounts [8] that attachment anxiety as measured by our adult attachment questionnaire (RSQ; [28]) specifically probes the appraisal of negative attachment-related material, but not all aversive stimuli or anxiety in general. On the other hand, however, we found that higher AX also correlated with higher intensity ratings for all positive (social and non-social) images, suggesting a broader impact of anxiety on arousal response to emotional scenes.

Our findings that attachment anxiety entails increased emotional arousal and decreased control / coping judgments to SN images, is highly consistent with traditional assumptions of adult attachment theory. Because anxious people tend to adopt hyperactivating strategies, they display increased fears of rejection and abandonment, which makes them more vigilant and sensitive about signals of social threat, loss, or disapproval. Thereby, unlike the down-regulation of emotions observed in avoidant individuals, anxiously attached people usually present with an up-regulation, or intensification, of their emotions [1]. Several behavioral studies have confirmed this general pattern of increased arousal and decreased control to socially threatening stimuli in anxious attachment. In one study investigating the influence of attachment style on emotional distress elicited by relationship threat [36], female participants who were in a current romantic relationship were asked to rate the degree to which they felt that different scenarios would be a threat to their relationship. These possible scenarios differed in degree of relationship threat and were based on an imaginative event that their partner had cancelled Saturday plans with them to instead spend all day and evening with another person, ostensibly to ‘‘study for an exam’’. Three pictures of the other person were shown - one was a man (low threat), the second a moderately attractive woman (medium threat), and the third a highly attractive woman (high threat). Results showed that anxious attachment was consistently linked to emotional distress, implying that anxious subjects were particularly sensitive to negative social scenarios. Another recent study reported increased arousal in anxiously attached subjects when viewing movie clips with negative relationships scenarios inducing emotions of fear and sadness [25]. More strikingly, in a study on tolerance for experimentally induced pain [37], the authors found that subjects scoring high on AX reported increased catastrophizing and stress, perceived less control, and exhibited weaker ability to decrease pain experience. The authors attributed these effects of high AX to maladaptive coping cognitions and emotional distress.

In keeping with these behavioral findings, recent neuroimaging data indicate that higher AX correlates with enhanced responses of the amygdala to socially threatening cues, including words with rejection meaning [12] and faces with angry expressions [11]. Furthermore, in the latter study, amygdala activation was specifically related to the social value of angry expression and arose only when these were presented as negative feedback from virtual partners, but not in a different feedback context. Because the amygdala is known to be critically implicated in processing threatening and self-relevant affective information [38, 39], and has strong projections to autonomic and attentional systems [40], these imaging results support the notion that a key aspect of anxious attachment involves hypervigilance towards socially-significant cues, and provide a plausible neural substrates for increased arousal responses to socially and emotionally significant stimuli in our study.

Finally, unlike negative images, positive images were rated as more intense by anxious participants with no significant difference between social and non-social content, suggesting a more general impact on all pleasant stimuli. However, these data are also consistent with assumptions of adult attachment theory, stating that the primary cognitive-affective hallmark of anxious people might be ambivalence in emotion regulation strategies, leading to stronger positive *and* negative attitudes, and presumably reflecting a mixture of desire for security and rejection sensitivity [1]. Unfortunately, only few studies have directly investigated ambivalence in anxious attachment style, but there is some evidence that anxiously attached students are more ambivalent toward their parents and romantic partners [31]. Our data indirectly extend this notion by showing that anxiously attached people rated all positive images as more intense, irrespective of their social content, which might reflect stronger vigilance and needs for reassuring or safe situations. However, more research on attachment anxiety and ambivalence is needed to further clarify this issue.

In sum, we show that adult attachment style strongly influences appraisal of images depicting emotional scenes, with distinct effects on different aspects of emotional experience, including pleasantness, intensity, and control. These effects occurred even though our task and stimuli made no explicit reference to attachment figures or concepts, indicating relatively automatic effects due to the implicit activation of the attachment system when processing general emotional scenes. On one hand, the higher subjects scored on avoidant attachment, the less pleasant they rated social positive scenes. On the other hand, the higher subjects scored on anxious attachment, the more intense they rated social negative stimuli as well as positive stimuli, and the less controllable they perceived the social negative scenes,. Our data do not only converge with classic notions on the influence of attachment style on behavior in relationship settings, but also provide new support to more recent findings suggesting more general influences on responses to pain and more abstract stimuli such as emotional words, faces, or movies [18-21, 25]. These data are also consistent with brain imaging results demonstrating that attachment style can modulate the neural responses of fear and reward circuitry during emotional challenges [11, 12, 14, 15]. Finally, our findings provide important information concerning inter-individual differences in emotional appraisal, paving the way for improved approaches to clinical populations with personality and relational disorders, anxiety, and social phobias.

AKNOWLEDGMENT

This research was supported by the National Center of Competence in Research (NCCR) Affective Sciences financed by the Swiss National Science Foundation (n° 51NF40-104897) and hosted by the University of Geneva, and grants of the Swiss National Science Foundation to DS and PV.

REFERENCES

1. Mikulincer, M. and P.R. Shaver, *Attachment in Adulthood: Structure, Dynamics, and Change*. 2007, New York: The Guilford Press.
2. Bowlby, J., *Attachment and loss*. 1969, New York,: Basic Books. v.
3. Bowlby, J., *Attachment and loss*. 2nd ed. 1982, New York: Basic Books. v.
4. Riggs, S.A., et al., *Attachment, personality, and psychopathology among adult inpatients: Self-reported romantic attachment style versus Adult Attachment Interview states of mind.* Development and Psychopathology, 2007. **19**(1): p. 263-291.
5. Ainsworth, M.D.S., *Bowlby-Ainsworth Attachment Theory.* Behavioral and Brain Sciences, 1978. **1**(3): p. 436-438.
6. Bartholomew, K. and L.M. Horowitz, *Attachment Styles among Young-Adults - a Test of a 4-Category Model.* Journal of Personality and Social Psychology, 1991. **61**(2): p. 226-244.
7. Brennan, K.A., C.L. Clark, and P.R. Shaver, *Self-report measurement of adult attachment: An integrative overview.*, in *Attachment theory and close relationships*, J.A. Simpson and W.S. Rholes, Editors. 1998, Guilford Press: New York. p. 46-76.
8. Kurdek, L.A., *On being insecure about the assessment of attachment styles.* Journal of Social and Personal Relationships, 2002. **19**(6): p. 811-834.
9. Simpson, J.A., *Influence of Attachment Styles on Romantic Relationships.* Journal of Personality and Social Psychology, 1990. **59**(5): p. 971-980.
10. Simpson, J.A., W.S. Rholes, and J.S. Nelligan, *Support Seeking and Support Giving within Couples in an Anxiety-Provoking Situation - the Role of Attachment Styles.* Journal of Personality and Social Psychology, 1992. **62**(3): p. 434-446.
11. Vrtička, P., et al., *Individual attachment style modulates human amygdala and striatum activation during social appraisal.* PLoS ONE, 2008. **3**(8): p. e2868.
12. Lemche, E., et al., *Human attachment security is mediated by the amygdala: Evidence from combined fMRI and psychophysiological measures.* Human Brain Mapping, 2006. **27**(8): p. 623-635.
13. Mikulincer, M. and P.R. Shaver, *The attachment behavioral system in adulthood: Activation, psychodynamics, and interpersonal processes.*, in *Advances in experimental and social psychology*, M.P. Zanna, Editor. 2003, Academic Press: New York. p. 56-152.
14. Gillath, O., et al., *Attachment-style differences in the ability to suppress negative thoughts: Exploring the neural correlates.* Neuroimage, 2005. **28**(4): p. 835-847.
15. Suslow, T., et al., *Attachment avoidance modulates neural response to masked facial emotion.* Human Brain Mapping, 2009.
16. Mikulincer, M. and E. Sheffi, *Adult attachment style and cognitive reactions to positive affect: A test of mental categorization and creative problem solving.* Motivation and Emotion, 2000. **24**(3): p. 149-174.
17. Pereg, D. and M. Mikulincer, *Attachment style and the regulation of negative affect: Exploring individual differences in mood congruency effects on memory and judgment.* Personality and Social Psychology Bulletin, 2004. **30**(1): p. 67-80.
18. Mikulincer, M., O. Gillath, and P.R. Shaver, *Activation of the attachment system in adulthood: Threat-related primes increase the accessibility of mental representations of attachment figures.* Journal of Personality and Social Psychology, 2002. **83**(4): p. 881-895.
19. Fraley, R.C., et al., *Adult attachment and the perception of emotional expressions: Probing the hyperactivating strategies underlying anxious attachment.* Journal of Personality, 2006. **74**(4): p. 1163-1190.
20. Dewitte, M., et al., *What's in a name? Attachment-related attentional.* Emotion, 2007. **7**(3): p. 535-545.
21. Dewitte, M., et al., *Attentive processing of threat and adult attachment: A dot-probe study.* Behaviour Research and Therapy, 2007. **45**(6): p. 1307-1317.
22. Kirsh, S.J. and J. Cassidy, *Preschoolers' attention to and memory for attachment-relevant information.* Child Development, 1997. **68**(6): p. 1143-1153.
23. Van Emmichoven, I.A.Z., et al., *Selective processing of threatening information: Effects of attachment representation and anxiety disorder on attention and memory.* Development and Psychopathology, 2003. **15**(1): p. 219-237.
24. Purkis, H.M. and O.V. Lipp, *Automatic attention does not equal automatic fear: Preferential attention without implicit valence.* Emotion, 2007. **7**(2): p. 314-323.
25. Rognoni, E., et al., *Relationship between adult attachment patterns, emotional experience and EEG frontal asymmetry.* Personality and Individual Differences, 2008. **44**(4): p. 909-920.
26. Bradley, M.M. and P.J. Lang, *Measuring Emotion - the Self-Assessment Mannequin and the Semantic Differential.* Journal of Behavior Therapy and Experimental Psychiatry, 1994. **25**(1): p. 49-59.
27. Fujita, F., E. Diener, and E. Sandvik, *Gender Differences in Negative Affect and Well-Being - the Case for Emotional Intensity.* Journal of Personality and Social Psychology, 1991. **61**(3): p. 427-434.
28. Griffin, D. and K. Bartholomew, *Models of the Self and Other - Fundamental Dimensions Underlying Measures of Adult Attachment.* Journal of Personality and Social Psychology, 1994. **67**(3): p. 430-445.
29. Aiken, L.S. and S.G. West, *Multiple regression: Testing and interpreting interactions*. 1991, Newbury Park, CA: Sage.
30. Lang, P.J., M.M. Bradley, and B.N. Cuthbert, eds. *International affective picture system (IAPS): instruction manual and affective ratings*. Technical report A-4, the center for research in psychophysiology. 1999, University of Florida.
31. Maio, G.R., F.D. Fincham, and E.J. Lycett, *Attitudinal ambivalence toward parents and attachment style.* Personality and Social Psychology Bulletin, 2000. **26**(12): p. 1451-1464.
32. Cohen, M.X. and P.R. Shaver, *Avoidant attachment and hemispheric lateralisation of the processing of attachment- and emotion-related words.* Cognition & Emotion, 2004. **18**(6): p. 799-813.
33. Searle, B. and N.M. Meara, *Affective dimensions of attachment styles: Exploring self-reported attachment style, gender, and emotional experience among college students.* Journal of Counseling Psychology, 1999. **46**(2): p. 147-158.
34. Tidwell, M.C.O., H.T. Reis, and P.R. Shaver, *Attachment, attractiveness, and social interaction: A diary study.* Journal of Personality and Social Psychology, 1996. **71**(4): p. 729-745.
35. Adolphs, R., et al., *A role for somatosensory cortices in the visual recognition of emotion as revealed by three-*

dimensional lesion mapping. Journal of Neuroscience, 2000. **20**(7): p. 2683-2690.
36. Meyer, B., L. Olivier, and D.A. Roth, *Please don't leave me! BIS/BAS, attachment styles, and responses to a relationship threat.* Personality and Individual Differences, 2005. **38**(1): p. 151-162.
37. Meredith, P.J., J. Strong, and J.A. Feeney, *The relationship of adult attachment to emotion, catastrophizing, control, threshold and tolerance, in experimentally-induced pain.* Pain, 2006. **120**(1-2): p. 44-52.
38. Adolphs, R., et al., *A mechanism for impaired fear recognition after amygdala damage.* Nature, 2005. **433**(7021): p. 68-72.
39. Sander, D., J. Grafman, and T. Zalla, *The human amygdala: an evolved system for relevance detection.* Rev Neurosci, 2003. **14**(4): p. 303-16.
40. Vuilleumier, P., *How brains beware: neural mechanisms of emotional attention.* Trends in Cognitive Sciences, 2005. **9**(12): p. 585-594.

VI.4. fMRI Study N°2a

In Preparation for Publication

fMRI study N°2a combined two new approaches in social neuroscience. It not only applied a direct comparison of the two most prominent emotion regulation (ER) strategies – re-appraisal (REAP) *versus* suppression (SUP) – but also assessed their differential impact on social *versus* non-social stimulus content, in addition to the prevalently used positive *versus* negative stimulus valence distinction. Along these lines, the main research question was:

- **How do re-appraisal and suppression affect the regulation of emotions as a function of social content as well as stimulus valence?**

The direct comparison between REAP and SUP (and vice versa) revealed activity in three prefrontal cortical areas displaying either selective increases during cognitive or behavioral ER. These effects were independent of valence or social content, which speaks in favor of a general role of these brain areas in ER.

Additional specific analyses of the differential activation as a function of either valence or social content during different ER conditions revealed several interesting patterns. Thereby, brain areas were identified in which REAP and/or SUP had a selective impact on responses to negative scenes (medial fusiform, anterior insula, dorsomedial prefrontal cortex [PFC]), non-social images (middle insula), or social scenarios (amygdala, medial PFC, posterior cingulate, lateral fusiform).

Most interestingly, whereas activity was modulated by ER selectively for social images in amygdala bilaterally, the left amygdala turned out to be more sensitive to REAP, whereas the right amygdala was more influenced by SUP. Such a result could demonstrate for the first time in a direct manner already previously suggested lateralization effects in the amygdala regarding ER. According to such accounts, activity in the left amygdala was linked with more cognitive and activity in the right amygdala with more basic, arousal-driven emotion processing. These notions fit well with the new data, showing that the activity in the left amygdala was decreased more during cognitive ER (REAP), whereas BOLD signal change in the right amygdala decreased stronger during behavioral ER (SUP).

Differential Brain Activation during Emotion Regulation as a Function of Valence and Social Content

Pascal Vrtička (1,2), David Sander (1,3), & Patrik Vuilleumier (1,2)

(1) Swiss Center for Affective Sciences, University of Geneva

(2) Laboratory for Neurology & Imaging of Cognition, Department of Neurology and Department of Neurosciences, University Hospital and Medical School, University of Geneva

3) Department of Psychology, FPSE, University of Geneva

ABSTRACT

Emotion Regulation (ER) refers to different mechanisms aiming at volitionally modulating emotional responses, including cognitive re-evaluation (re-appraisal; REAP) or behavioral inhibition of emotion expression (suppression; SUP). However, despite the importance of these ER strategies, previous functional magnetic resonance imaging (fMRI) studies have not directly disentangled the specific neural impact of REAP *versus* SUP on brain responses to different kinds of emotional events. Although different effects have been reported regarding valence (positive *versus* negative stimuli), no study has systematic investigated changes in emotional processing as a function of particular stimulus content variables.

Our fMRI study directly compared brain activation to visual scenes during the use of either ER strategies, relative to a "natural" viewing condition, and tested the differential effects of ER as a function of the social *versus* non-social content of scenes, in addition to their negative versus positive valence (by manipulating these factors orthogonally in a systematic 2x2 design).

We found that several prefrontal cortical areas were differentially recruited during either REAP or SUP, independent of the valence and scene content of images. In addition, selective modulations by either REAP or SUP were found depending on the negative valence of scenes (medial fusiform gyrus, anterior insula, dmPFC), and on their non-social (insula) or social (bilateral amygdala, mPFC, posterior cingulate, lateral fusiform gyrus) significance. Furthermore, we observed a significant lateralization in the amygdala for the effect of the two different ER strategies, with a predominant modulation by REAP on the left side but by SUP on the right side.

Taken together, these results not only highlight the distributed nature of neural changes induced by emotion regulation, but also reveal the selectivity of these changes for different strategies (REAP or SUP) and for different dimensions of emotional information (SOC or NEG).

CORRESPONDING AUTHOR:

Pascal Vrtička
Swiss Center for Affective Sciences
7, rue des Battoirs, 1205 Geneva, Switzerland
Tel: +41 – (0)22 379 9824 Fax: +41 – (0)22 379 9844
Email: pascal.vrticka@unige.ch

INTRODUCTION

Emotions play a vital part in human life, shaping many personal and social processes during almost every moment of our existence. However, besides being influenced by emotions in our actions, we in turn have the ability to modulate our emotional experiences by different mechanisms. The importance of such emotion regulation capacities can be observed in various domains or situations, either in terms of increasing positive health outcomes when used appropriately, or by promoting mood disorders and anxiety when malfunctioning [1, 2]. Uncovering the neural mechanisms that can regulate affect and their influence on the processing of emotionally significant information is therefore of great importance, to better understand both the predispositions to affective disorders and the effect of specific therapeutic interventions.

Behavioral research on emotion regulation (ER) [3] has highlighted two major kinds of ER strategies, which are conceptualized to have their impact at distinct stages during emotion processing.

On the one hand, *cognitive reappraisal (REAP)*, or antecedent-focused ER, is thought to intervene at a relatively early stage of emotion processing by modulating the meaning of an emotional event. This typically involves the intentional (conscious) generation of alternative interpretations in response to an event, allowing one to modify (e.g. minimize) its emotional significance. Accordingly, REAP has been shown to alter subjective emotion experience (as measured by physiological arousal or verbal report) to both negative and positive emotional cues, and for both up- and down-regulation of affect [4]. Moreover, studies using functional magnetic resonance imaging (fMRI) during REAP have shown increased activity in a widespread network of prefrontal and orbitofrontal regions, together with simultaneous decreases in activity in areas implied in emotion processing, such as the amygdala, the posterior cingulate cortex (PCC), and insula [4-7].

On the other hand, *suppression of behavioral expression (SUP)*, or response-focused ER, is thought to operate only after emotion generation, and directed towards the inhibition of emotional responses (e.g. facial expressions, physiological changes, etc.). Thus, the triggering event is emotionally appraised as usual but the overt behavioral manifestations are voluntarily suppressed. Unlike REAP, SUP has been found to increase sympathetic activation, to produce little effect on negative emotion experience, and to even decrease positive affect [3]. Even though brain imaging data on

SUP are still scarce, two studies have reported decreased negative emotion expression and experience in combination with increased prefrontal cortical activity during SUP of negative emotions (either sadness or disgust) [7, 8]. One of these studies also found increased amygdala and insula responses during SUP, but only in a late period of prolonged exposure to disgusting film-clips [7], contrasting with the decreases observed in these regions during REAP [4-6]. Finally, whereas the use of REAP has been linked to enhanced control of emotion, better interpersonal functioning, and higher psychological and physical well-being, the frequent use of SUP is thought to result in diminished control of emotion, worse interpersonal functioning, poor memory, and greater risks for depression [9].

Despite the fact that our knowledge regarding the neural correlates of ER has steadily increased during the last decade, several important issues still remain largely unresolved. Firstly, most of the previous brain imaging studies on ER have investigated only one type of ER strategy at a time [4-6, 8, 10] or compared REAP and SUP using indirect measures based on putative differences in their specific temporal dynamics [7]. Hence, very little is known about the *differential* impact of REAP and SUP on brain activity during emotional processing, particularly in relation to the down-regulation of emotions. This does not only concern cortical brain areas exerting top-down control during ER, predominantly located within prefrontal cortex, where a functional segregation between behavioral (SUP) and cognitive (REAP) control has already previously been suggested [7]; but also the lower-level areas in sensory cortices and subcortical regions that are crucially involved in emotional responses, such as the amygdala, striatum, or insula. Thus, it still remains to be elucidated whether insula activity during down-regulation of affect can only decrease through REAP but increases during SUP, as would be predicted by ER theory [3, 9], or whether it can also be down-regulated during SUP just like during REAP. Therefore, in our study, we examined both cognitive (REAP) as well as behavioral (SUP) strategies of ER, besides a “natural viewing” (or “watch”) condition (EMEXP; see Methods) that served as an emotional baseline condition.

Secondly, previous fMRI experiments on ER have only differentiated modulatory effects as a function of the valence (VAL; positive *versus* negative) and/or arousal (low *versus* high intensity) of emotional stimuli, but did not consider other important stimulus properties. Thus, the emotional material used in previous studies typically mixed together various types of cues associated with animate *versus* inanimate, or social *versus* non-social stimulus features, even though data from a number of previous experiments point to major differences in emotional reactions (and their corresponding neural signatures) depending on the social *versus* non-social nature of information [11-14]. In addition, there is evidence that the activation of some limbic brain areas (i.e. the amygdala) is more sensitive to social and thus interpersonal aspects rather than to non-social dimensions of emotionally arousing situations [15, 16], presumably implying different processes for successful ER strategies. As a consequence, in our study, emotional images were systematically varied according to both their affective valence (VAL: positive *versus* negative) and their social content (SC: social *versus* non-social), as separate experimental factors in a 2x2 design.

Finally, we used a systematic approach combining whole-brain imaging with analysis of functionally defined regions of interest (ROIs). In previous work, activation differences related to ER have typically been derived from comparisons between REAP *versus* baseline or SUP *versus* baseline, and this for a single emotional dimension (i.e. negative stimuli) [6, 7]. To our knowledge, only one imaging study so far also compared the *relative differences* in brain responses to positive *versus* negative images during ER [4]. In the present study, we were especially interested in identifying brain areas that were not only involved in processing a specific stimulus property (i.e. VAL or SC), but also significantly modulated by the different ER strategies (Viewing condition [VCON]: EMEXP, REAP, or SUP), and that thus exhibited a significant VCON x VAL or VCON x SC interaction. For this purpose, we first used a whole-brain RFX analysis to determine functional ROIs whose activity was modulated by either VAL or SC, and then computed an additional second-level RFX analysis using a paired t-test design in SPM to identify only those voxels within the ROIs that displayed significant effects related to the different ER strategies (VCON – see Methods). This approach allowed us to determine brain regions where the processing of specific affective cues (valence or social content) was modulated by the different kinds of ER.

METHODS

Subjects

We recruited 19 healthy paid volunteers (all right-handed women, mean age 24.82 ± 4.0), who all had a normal or corrected to normal vision, no history of neurological or psychiatric disease, and gave informed written consent according to the local ethical committee regulation. Only women were included because of evidence that emotions are typically more intense and more prone to regulation in women (e.g. [17]), but also to increase comparability with previous emotion regulation studies that also included women only [4] and to avoid any potential sex differences that could have modulated the regulation effects of primary interest.

Experimental material and procedure

STIMULI

A total number of 360 emotional pictures were initially chosen either from the International Affective Pictures System (IAPS) or from the internet. All were in colors, and adjusted to obtain similar size, contrast, and pixel resolution. Half of the pictures displayed scenes with a clear social content, such as two people fighting or a mother interacting with her baby. The other half represented objects or landscapes (non-social), like a dead bird covered in oil or a tropical island scene. All 360 pictures were rated in a separate behavioral study by 54 female students on three continuous rating scales (from 0 to 100), including PLEASANTNESS (PLN, from very negative to very positive), INTENSITY (INT; from low to high arousal), and CONTROL (CON; from absence to full presence). According to the averaged rating results from this sample, 240 pictures were finally chosen for the fMRI study, and sorted by their SOCIAL CONTENT (SC; either social or non-social) and VALENCE (VAL; either positive or negative). This gave rise to four stimulus categories (60 pictures each): Social Positive (SP) or Negative (NSP), and Non-Social Positive (NSP) or Negative (NSN). The final distribution of pictures in these four categories as a function of emotional rating scores (PLN, INT, and CON) showed that NEG images had significantly lower pleasantness scores as compared to POS images [F(1,59)= 3919.25, p< .001], but higher intensity [F(1,59)= 1554.01, p< .001] and lower control [F(1,59)= 452.97, p< .001] scores. However, most importantly, there were no significant differences for all three rating scales between SOC *versus* NSOC images [Fs(1,59)< 1.36, ps> .25], and no significant VAL x SC interactions [Fs(1,59)< 2.65, ps> .11] (as shown by a 2 [VAL] x 2 [SC] repeated-measure analysis of variance [ANOVAs]). Note that the differences in intensity between negative and positive stimuli could not be avoided in order to match pairs of social and non-social scenes in both valence conditions, because social material is otherwise typically judged as much more intense than non-social material [18]. Finally, we also selected 40 neutral images from the IAPS database (20 including humans, 20 without humans) to be used in a baseline control condition (see below), with average valence ratings of 4.97 ± 0.17 (on the same scale of 1 to 10).

The final experimental set of emotional images showed no differences in luminance across categories [all Fs< 2.83; ps> .098 in a 2 [VAL] x 2 [SC] repeated-measure ANOVA], and all social images were comparable in terms of the average number of people depicted per image [SOC POS *versus* SOC NEG: Fs(1,59)= .468; p= .497].

EXPERIMENTAL CONDITIONS

Before entering the fMRI scanner, we told all participants that the purpose of the experiment would be to investigate how the brain reacts to different types of images (e.g. real scenes, TV or movie scenes) and to which degree people can voluntarily influence their emotional effects. Accordingly, the experimental layout comprised four different viewing conditions (VCON), in which pictures were presented with different task instructions to induce different emotion regulation strategies.

The first condition served a control baseline (BASE), and was introduced to the participants as “a photographic quality” judgment, where they had to indicate on each trial (by button press, using a 4-point scale – see below) whether the image was of good quality (e.g. well focused, properly lighted, etc.). All images in this condition were neutral, but could display either scenes with humans (i.e. social content), or inanimate settings and landscapes (i.e., non-social content). This condition was later used to provide a baseline for general differences in brain activation to social vs non-social stimuli, irrespective of emotional processing demands and valence. This condition was presented as the first block of the first and the last block of the last scanning run.

The three other conditions constituted the main experimental design and included emotional images only. To assess brain responses during “natural” viewing [4, 6], participants were asked to watch and experience the depicted emotional scenarios as if they were living them for real (EMOTION EXPERIENCE condition; EMEXP). To assess the effect of cognitive re-evaluation (RE-APPRAISAL condition; REAP), participants were instructed to view the depicted emotional scenes as parts of a movie clip or TV show that displayed fake or artificially set-up situations created to give rise to emotions. The latter strategy (“pretend unreal”) has been one of the most often used in order to down-regulate emotional reactions to negative and positive images [4]. Finally, to assess the effect of behavioral inhibition (SUPPRESSION condition; SUP) the participants were instructed to watch the pictures similarly to the EMEXP condition, but with the important difference that they were told not to display any felt emotions become visible on the outside (i.e. through breathing frequency, heart rate, and skin conductance responses – which were actually monitored during fMRI scanning; or through facial expression – which was pretended to be recorded via an eyetracker camera). After each picture, participants were shown a rating display and asked to report the feeling state evoked by the preceding stimulus (“How did you feel while seeing the last image”?), using a 4-point scale (see below).

Importantly, all emotional images were counterbalanced across participants, so that the same images seen in one viewing conditions by a given subject were seen in the other viewing conditions by different subjects.

PROCEDURE

The fMRI experiment was divided into three successive scanning runs. Each run included two of the three experimental viewing conditions, presented in blocks of 40 emotional images (duration = 294 sec per block), whereas the first and the last run also included an additional block of 20 neutral images (baseline condition, duration = 151 sec). Within each block, images were pseudo-randomized and equally probable for the different stimulus categories (social *vs* non-social content, positive *vs* negative valence). The first and the third runs lasted approximately 13 min, and the middle session 10 minutes.

Each VCON block began with an instruction display (7 sec), followed by psudo-randomized images. Every individual trial started with a fixation cross at the screen center (average duration= 1125 ms jittered between 790 and 1485 ms), followed by an emotional or neutral image for 2 sec, and then a response display to rate the image (4 sec; see Figure 1). Ratings were made by using a 4-button response box according to a 4-point scale ranging from very and slightly negative (buttons 1 and 2, respectively), to slightly and very positive (buttons 3 and 4, respectively),

MRI ACQUISITION

MRI data were acquired on a 3 T whole-body INTERA system (Philips Medical Systems), using standard head-coil configuration. For each participant, a structural image was obtained with a MPRAGE T1-weighted sequence

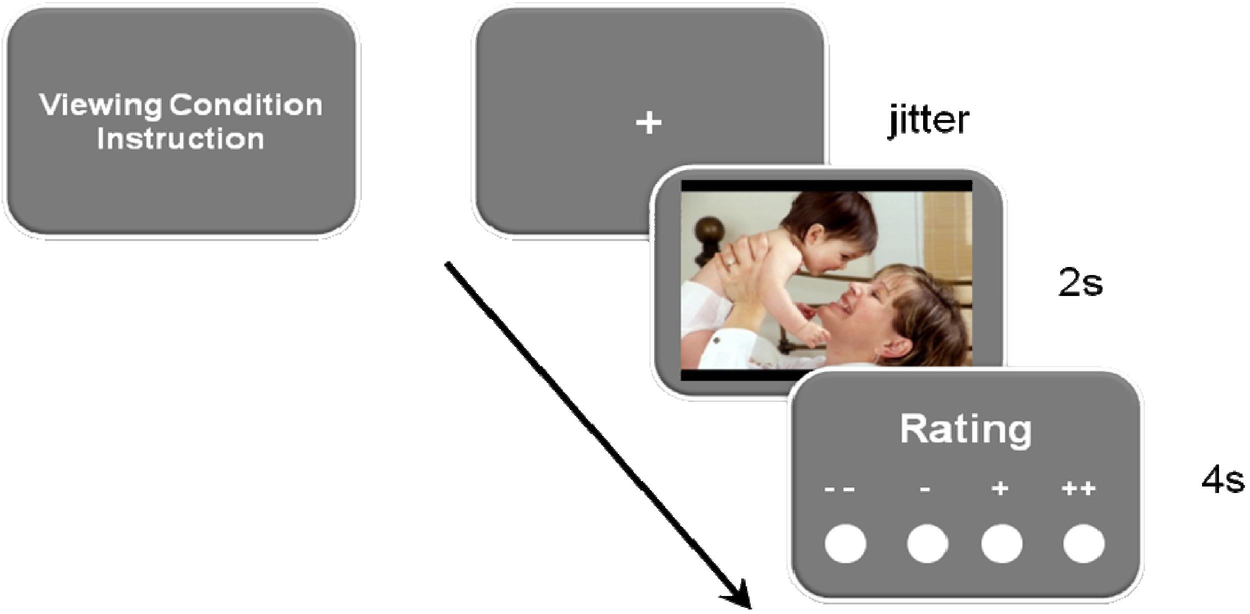


Figure 1: Illustration of the Within-Block Event-Related fMRI Paradigm. *Each block in each viewing condition began with an instruction slide (7 sec), followed by a series of pictures. Each picture was shown for 2 seconds, preceded by a central fixation cross (jittered between 790 and 1485 ms), and followed by an emotional rating display (4 seconds).*

(TR/TE/Flip = 2200 ms/30ms/85°, parallel acquisition (GRAPPA) with acceleration factor 2, FOV = 235mm x 235mm, matrix = 128x84, resulting voxel size is 2.8 x 1.8 x 3.4 mm3). Functional images (TI/TR/TE/flip = 900/1900/2.32/9°, parallel acquisition (GRAPPA) with acceleration factor 2, FOV = 230 x 230 x 173 mm3, Matrix = 256 x 246 x 192) covered the whole brain, composed of 36 contiguous 4mm axial slices parallel to the inferior edge of the occipital and temporal lobes, and acquired continuously for a total of 975 images per participant (two sessions with 350 and one session with 275 images).

Image processing was performed with SPM2 (www.fil.ion.ucl.ac.uk) using standard procedures for realignment of the time-series, slice-timing correction, normalization to a standard brain template in MNI space, and smoothing with a 8 mm FWHM Gaussian kernel. Statistical analysis was performed using the general linear model implemented in SPM2, with a separate regressor for each event type convolved with a canonical hemodynamic response function. Twelve event types from the emotion regulation task (4 image categories: SP, NSP, SN and NSN; for each of the three experimental viewing conditions: EMEXP, REAP, and SUP), plus two additional event types (social and non-social) from the baseline condition (BASE) were modeled for each participant, using the three scanning runs in a fixed-effect analysis at the single-subject level. Movement parameters from realignment corrections were entered as additional covariates of no interest for each scanning run, in order to account for residual movement artifacts after realignment. Statistical parametric maps were then generated from linear contrasts between the different conditions in each participant.

A first second-stage random-effect (RFX) analysis was performed using one-sample t-tests on contrast images obtained in each subject for each comparison of interest. All contrasts were performed across the whole brain using standard threshold criteria (Worsley et al., 1996) of significant activation at a voxel-level of p< .001 uncorrected (except for bilateral amygdala, p< .005 given a priori predictions), and a cluster size greater than 5 voxels (135 mm3). Average parameter estimates of activity (betas) for each condition were extracted from all voxels in regions of interest (ROIs), defined by the full-extent clusters showing significant activation at a voxel level of p< .001 (uncorrected, T-value > 3.61; except for bilateral amygdala: p< .005, T-value > 2.88) in the SPM group analysis (random-effect contrasts). These beta values were then used for subsidiary repeated-measure ANOVAs and t-tests performed in SPSS with

the factors of stimulus content, valence, and viewing condition, when appropriate.

Additional RFX paired t-tests were also performed to reveal which regions showed differential responses to SOC *versus* NSOC and/or POS *versus* NEG images (and vice-versa) during the three experimental viewing conditions (i.e. [$\mathrm{EMEXP_{SOC}}$ – $\mathrm{EMEXP_{NSOC}}$] > [$\mathrm{REAP_{SOC}}$ – $\mathrm{REAP_{NSOC}}$]), and thus identify the selective effect of different emotion regulation strategies on different stimulus types (i.e. stimulus x viewing condition interactions). These comparisons were made using a paired t-test procedure applied to second-stage contrasts (at p< .025), as previously described elsewhere [19, 20], which allowed us to assess specific non-crossed interactions between the factors of interest (unlike unconstrained interaction tests that typically probe for the two sides of an interaction in a single whole-brain contrast). Subsequently, the paired t-tests were inclusively masked with the SOC *versus* NSOC contrast or POS *versus* NEG contrast (at p< .025), and vice-versa, to ensure that the observed interactions involved the initial ROIs found to activate in the main effects of our RFX analysis, and then thresholded using

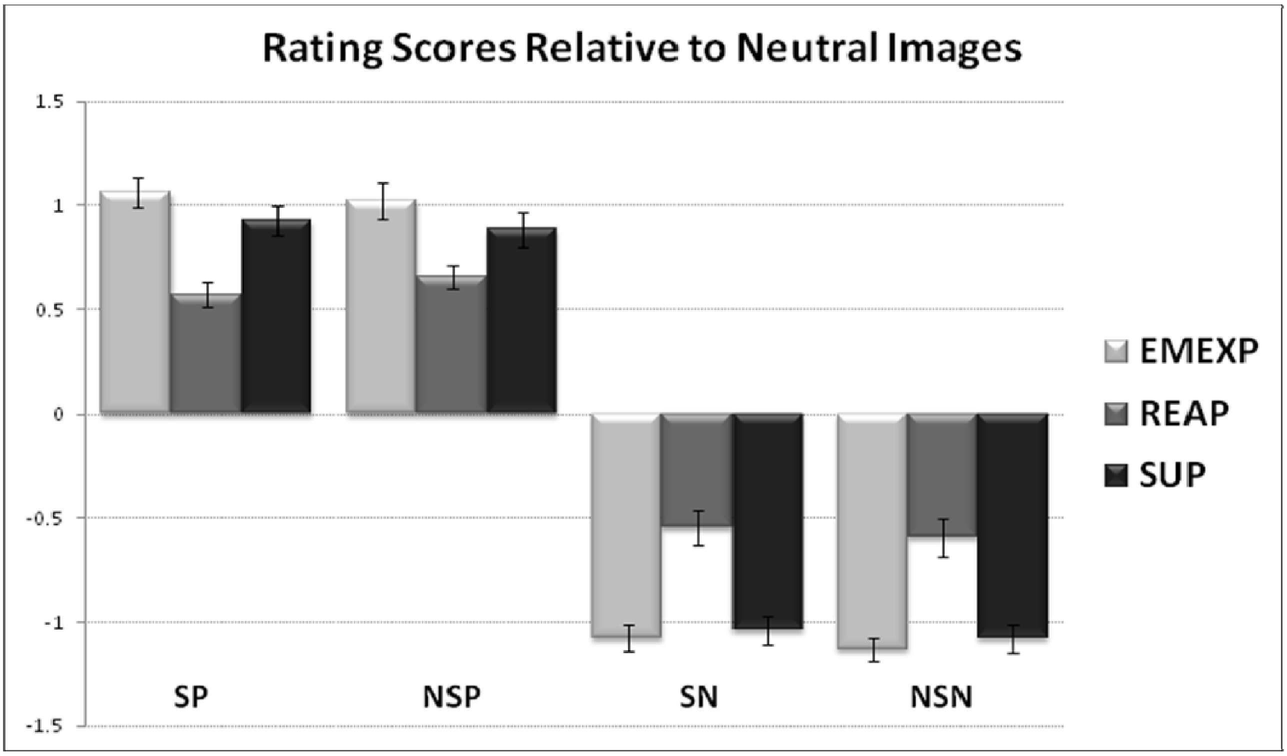


Figure 2: Behavioral Rating Results. *Subjects rated each image on a 4 point scale (from -2 and -1 for very and slightly negative, to +1 and +2 for very and slightly positive, respectively). While there was a clear main effect of picture valence, emotion regulation effects were significantly stronger during REAP than during SUP and significant (relative to EMEXP) for all four stimulus categories during REAP [p< .001], but only for POS stimuli [p< .020], not NEG stimuli [p> .17], during SUP. All values are averages of individual means displayed with ± 1 S.E.M. Abbreviations: SP= Social Positive, NSP= Non-Social Positive, SN= Social Negative, NSN= Non-Social Negative, EMEXP= Emotion Experience, REAP= Re-Appraisal, SUP= Suppression.*

a T-value > 3.61 (p< .001), and a cluster size of k>10 voxels [270 mm3]). Even though probabilities are not completely independent, this procedure results in an effective threshold that approaches the joint probability estimate of p< .000025 [19, 20], which is inferior to the conventional statistical threshold of p< .001 at the voxel level.

RESULTS

Behavioral Data

During fMRI scanning, participants rated their emotional feeling in response to each picture on a four point scale (from very negative to very positive, see Methods). These rating scores confirmed that positive images were more pleasant and negative images less pleasant than neutral images, across all viewing conditions (average = 3.28 ± .17, 2.43 ± .21, and 1.52 ± .15 for positive, neutral, and negative, respectively; all pairwise comparisons t≥ 14.44, p≤ .001). This difference was found regardless of stimulus categories [main effect of VAL, F(1,18)> 6.24, p< .001], without any effect of the social content for any of the emotional conditions [SP *versus* NSP: t(18)< 1.84, p> .083; SN *versus* NSN: t(18)< 1.47, p> .16, paired t-tests], and no interaction (VAL x SC x VCON repeated-measure ANOVA).

More critically, however, emotional ratings of the pictures were significantly modulated by the emotional viewing conditions induced by our task instructions (even though pictures were counterbalanced across participants and equally presented in each condition). As compared with the "natural" EMEXP condition, emotion regulation was successfully achieved during both REAP and SUP conditions, leading to less intense emotional rating scores for all four stimulus categories during REAP [t(18)> 4.35, p< .001], and for POS stimuli during SUP [t(18)> 2.56, p< .020]. There was no significant difference for NEG stimuli in the latter condition: t(18)< 1.45, p> .17]. Moreover, as shown in Figure 2, the effects of REAP on emotional ratings were stronger than those of SUP for all four picture categories [paired t-tests, all t(18)> 2.72, p< .014). This pattern is consistent with our emotion regulation instructions for the SUP condition, which required suppressing "external" emotional expression but not "internal" experience (unlike REAP).

fMRI Data

1. MAIN EFFECTS

In a first step, we computed the main effects of scene content (SOC *versus* NSOC, and vice-versa) and emotional valence (POS *versus* NEG, and vice-versa), across all viewing conditions; as well as the direct comparison between the two emotion regulation conditions (REAP *versus* SUP, and vice-versa), across all picture types. Findings are described below and summarized in Tables 1 to 3.

SOC > NSOC

Region	*BA*	*Voxel*	*T*	*xyz*	*ROI*
Amygdala L *		13	3.55	-21 -9 -18	x
Amygdala R *		26	4.21	21 -6 -18	x
mOFC	11	40	5.26	-3 54 -18	x
mPFC	10	28	4.36	3 57 15	x
PCC / PREC	7 / 23	393	7.32	0 -51 33	x
FG R	37	98	6.59	42 -42 -27	x
dlPFC L	6	6	4.99	-39 3 60	
dmPFC L	9	9	4.77	-9 48 51	
dlPFC R	44	16	4.15	51 21 27	
ATP R	20	9	4.37	33 12 -36	
Temp Inf L	20 / 21	109	6.85	-57 -3 -27	
Temp Inf R	21	111	6.29	60 -6 -24	
pSTS L	19	458	7.82	-45 -84 0	
pSTS R	19	711	7.07	45 -48 18	
FG L	18	18	6.5	-15 -60 -6	
FG L	19 / 37	25	4.43	-42 -63 -21	
FG L	20	10	4.4	-42 -33 -24	
Occipital L	17	331	7.74	-6 -102 9	

NSOC > SOC

Region	*BA*	*Voxel*	*T-Value*	*xyz*	*ROI*
Insula middle R		151	6.42	57 3 -6	x
Insula middle R		39	4.78	33 18 -6	x
dlPFC L	6	32	4.71	-24 6 54	
dlPFC L	9	9	4.09	-21 54 24	
PFC R	11	17	5.38	12 39 -3	
dlPFC R	46	29	5.84	48 48 15	
dlPFC L	45	85	5.63	-42 33 18	
vlOFC L	47	76	5.43	-36 36 -15	
vlOFC R	47	22	4.78	24 24 -15	
Cingulate L	23	81	5.71	-3 -30 42	
Cingulate L	24	136	5.41	-3 3 30	
Cingulate R	23	9	5.39	9 -9 42	
ACC L	32	38	4.31	-6 36 18	
Insula medial L		94	7.28	-39 -6 6	
Insula posterior R		7	4.33	39 -18 15	
Insula anterior L		30	5.54	-39 15 -9	
Insula posterior R		18	5.79	39 3 9	
Temp sup R	22	53	5.04	63 -18 9	
Supramarginal L	40	293	5.39	-54 -45 39	
Supramarginal R	40	301	7.41	51 -39 54	
Heschl R	41	24	5.27	45 -30 15	
Calcarine L	17	25	5.47	-15 -60 15	
Calcarine R	17	62	6.3	18 -57 15	
FG R	37	923	9.51	30 -45 -12	
Occipital L	18	415	7.92	-33 -87 15	
Occipital L	37	43	7.22	-48 -63 -9	
Occipital L	17 / 18	5	4.12	-15 -99 -3	
Occipital R	18	265	7.61	30 -81 15	

Table 1: Brain Areas activated for the Comparison between Social and Non-Social Emotional Image Categories. *Peak coordinates are given in MNI space and listed with best estimates of anatomical location. p≤ .001 for all clusters (except fot p< .005). BA = Brodmann's area, ROI = region of interest, mOFC = medial orbitofrontal cortex, mPFC= medial prefrontal cortex, FG= Fusiform gyrus, dlPFC = dorso-lateral prefrontal cortex, dmPFC = dorso-medial prefrontal cortex, ATP = anterior temporal pole, pSTS = posterior temporal sulcus, vlOFC = ventro-lateral orbitofrontal cortex, ACC = anterior cingulate cortex.*

1.1. Comparison of Social and Non-Social Pictures

As expected [12, 13, 21-23], social scenes produced greater activation in widespread brain networks including extrastriate visual cortex, temporal lobe, and ventromedial prefrontal areas, all previously associated with face and body perception, person recognition, mentalizing, and/or social cognition. Conversely, non-social scenes produced greater activation in visual areas associated with object and place recognition, as well as insula, anterior cingulate, and more

POS > NEG					
Region	*BA*	*Vox*	*T*	*xyz*	*ROI*
ACC subcallosal R	25	69	6.07	3 9 -9	
Hippocampus L	37	7	4.58	-36 -39 -3	
Postcentral L	3	129	6.26	-42 -30 63	
Occipital L	18	17	5.54	-138	
Occipital R	17	6	4.31	15 -90 12	
NEG > POS					
Region	*BA*	*Voxel*	*T-Value*	*xyz*	*ROI*
Insula anterior R		874	8.05	30 24 -12	x
dmPFC	8	309	6.83	0 42 48	x
FG/LG L	17	1581	9.13	-3 -84 6	x
dmPFC L	6	7	4.41	-15 12 63	
dlPFC R	8 / 9	8	4.38	27 30 42	
Cing med R	32	37	5.06	9 18 36	
Caudate L		62	5.46	-9 6 3	
Caudate R		202	5.88	15 3 15	
Insula lateral L		662	7.86	-42 21 3	
PREC R	17 / 18	796	7.53	21 -57 6	
Heschl R	42	5	3.94	57 -39 24	
Postcentral L	3	12	4.61	-51 -18 51	
Precentral L	6	55	5.7	-45 -6 45	
aSTS R	20	17	4.91	51 -24 -15	
aSTS R	21	7	3.8	54 -39 -3	
Parietal sup. R	7	136	5.35	12 -57 69	
Parietal sup. L	40	53	4.54	-60 -39 39	
Temp inf L	20	10	4.46	-51 -3 -24	
FG R	19	15	3.97	24 -66 -9	
FG R	37	13	3.92	30 -51 -12	
FG L	37	17	4.56	-42 -42 -21	
Occipital R	18	54	5.68	21 -93 -6	
Brainstem		6	3.93	9 -21 -36	

Table 2: Brain Areas activated for the Comparison between Positive and Negative Emotional Image Categories. *Peak coordinates are given in MNI space and listed with best estimates of anatomical location. p≤ .001 for all clusters. BA = Brodmann's area, ROI = region of interest, ACC = anterior cingulate cortex, dmPFC = dorso-medial prefrontal cortex, FG = fusiform gyrus, LG = lingual gyrus, dlPFC = dorso-lateral prefrontal cortex, aSTS = anterior superior temporal sulcus.*

lateral areas in prefrontal cortex (see Table 1 and Figures 1 and 2 in Supplemental material).

1.2. Comparison of Negative and Positive Pictures

Negative images increased in activity mainly in right insula, prefrontal, parietal, dorsal cingulate, and extrastriate visual cortex, all brain areas known to be involved in arousal, attention, and object recognition [4, 24-26]. Positive images produced greater activation in subgenual anterior cingulate cortex, hippocampus, and occipital regions (see Table 2 and Figure 1 in Supplement).

1.3. Comparison of Re-Appraisal *versus* Suppression

Of greater interest for the current study were the specific activations related to different emotion regulation strategies. Distinct clusters were identified in right prefrontal cortex (PFC) for the direct comparison between REAP *versus* SUP (anterior portion of the superior frontal gyrus [SFG] and posterior portion of the middle frontal gyrus [MFG], Figure 3a and Table 3), and for the comparison between SUP *versus* REAP (superior frontal sulcus and supplementary motor area [SMA], Figure 3d and Table 3).

For both PFC clusters found for the contrast REAP *versus* SUP (Figure 3a), we conducted further statistical tests using a full 3 (VCON) x 2 (VAL) x 2 (SC) repeated-measure ANOVA on beta values extracted from activated voxels (in SPSS, see methods). This analysis confirmed a main effect of VCON (Fs(1,18)> 6.09, ps< .005), which was due to significant increases during REAP relative to SUP [all t(18)> 5.86, ps< .001], and to a lesser extent during EMEXP relative to SUP [t(18)≥ 2.24 ps≤ .038]. However, in both clusters, only the REAP condition produced significant increases as compared with the neutral baseline condition [BASE; all t(18)> 2.32, ps< .032] (Figure 3b). Moreover, in the posterior right MFG cluster (xyz= 39 24 54), there was a significant SC x VAL interaction during REAP [F(1,18)= 6.51, p= .02], indicating that this activation was predominantly driven by the social negative (SN) image category, with a significant effect of picture content for negative [SN > NSN, t(18)= 2.62, p= .017], but not positive [SP > NSP; t(18)= .0443, p= .97] images. In the SFG, there was also a marginally significant SC x VAL interaction (p= .08), but no main effects of picture content [p≥ .14] in pairwise comparisons.

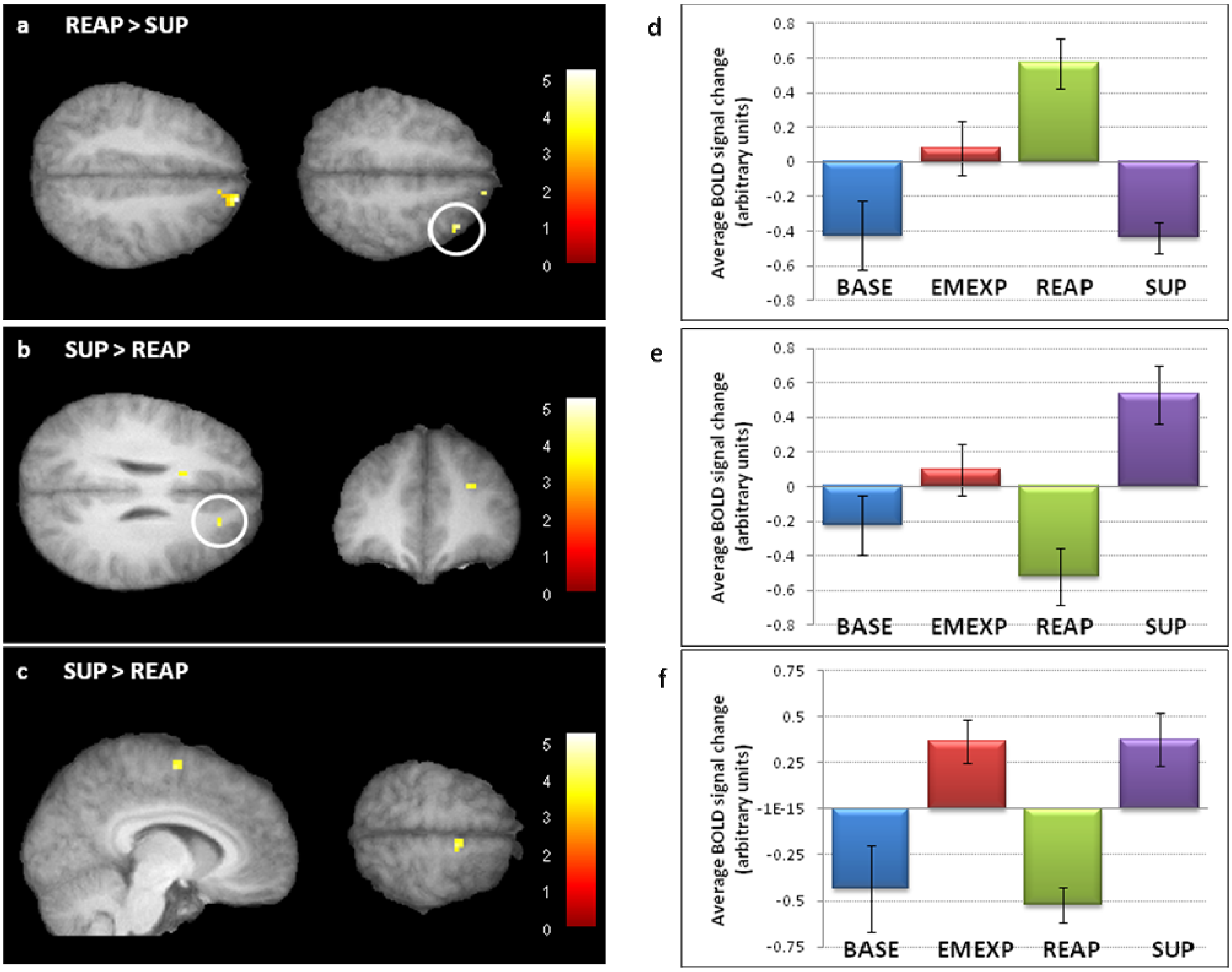


Figure 3: Distinct PFC and SMA activations for the two Emotion Regulation Strategies. a) *Statistical parametric map for the contrast REAP > SUP, showing activation in two areas of the right dorsal PFC (superior and middle frontal gyri [SFG and MFG]).* ***b)*** *Statistical parametric map (p< .005) for the contrast SUP > REAP, showing activation in right superior frontal sulcus (SFS).* ***c)*** *Statistical parametric map (p< .005) for the contrast SUP > REAP, showing activation in right supplementary motor area (SMA).* ***d)*** *Parameter estimates (beta values) extracted from the MFG cluster in a), averaged across voxels and participants, showing increases during REAP relative to SUP, but also decreases during SUP relative to EMEXP. A similar pattern of activity was also found for the right SFG cluster (see text).* ***e)*** *Parameter estimates (beta values) extracted from SFS cluster in b), displaying significant increases during SUP relative to both REAP and EMEXP.* ***f)*** *Parameter estimates (beta values) extracted from the SMA in c), revealing decreases during REAP as compared to both EMEXP and SUP. BASE = neutral condition with neutral pictures (see text). All values are displayed with ± 1 S.E.M., p< .001)*

Conversely, in the right superior frontal sulcus cluster identified by the contrast SUP *versus* REAP (Figure 3b and Table 3), a 3x2x2 repeated measure ANOVA on beta values confirmed a significant main effect of VCON [F=(1,18)= 9.42, p= .001], reflecting a marked difference between SUP and REAP [t(18)= 5.07, p< .001] Moreover, only SUP showed significant increases as compared with the neutral baseline [BASE; t(18)= 3.16, p= .005] (Figure 3e). Finally, for the right supplementary motor area (SMA; Figure 3c and Table 3), a similar ANOVA again indicated a main effect of VCON [F= 13.86, p< .001], while pairwise comparisons revealed relative decreases during REAP as compared with both SUP [t(18)= 5.34, p< .001] and EMEXP [t(18)=5.03, p< .001]. Moreover, both SUP and EMEXP showed significant increases relative to BASE [t(18)> 2.38, p≤ .028] (Figure 3f). No effect or interaction involving SC was found in these two clusters. In all these regions, the repeated-measure ANOVAs showed no main effects of VAL [all Fs(1,18)≤ .09, ps≥ .76] and no other interactions.

2. SELECTIVE REGULATION EFFECTS IN DIFFERENT STIMULUS CONDITIONS

In a second step, we performed a random-effect SPM analysis using a masked paired t-test design on specific contrasts of interest (see Methods), allowing us to test for the differential effects of each regulation condition as a function of the social content (VCON x SC interaction) and emotional valence (VCON x VAL interaction) of pictures. Parameters of activity (betas) were then extracted from each region showing such interactions and submitted to a full 3 (VCON) x 2 (VAL) x 2 (SC) repeated-measure ANOVA, followed by post-hoc t-test comparisons. Findings are described below for the main clusters of activation and summarized in Table 4.

2.1. Regulation as a function of valence (VCON x VAL interactions)

There were no brain areas displaying a significant interaction for the response to POS > NEG images.

REAP>SUP					
Region	*BA*	*Voxel*	*T-Value*	*xyz*	*ROI*
dlPFC R	9	12	5.13	18 51 45	x
dlPFC R	9	5	4.42	39 24 54	x
Pariet sup R	7	6	3.87	18 -57 69	
SUP>REAP					
Region	*BA*	*Voxel*	*T-Value*	*xyz*	*ROI*
lPFC R	46	6	4.1	27 42 24	x
SMA R	6	10	4.3	9 -6 63	x

Table 3: Brain Areas activated for the direct Comparison between the two different Emotion Regulation Strategies. *Peak coordinates are given in MNI space and listed with best estimates of anatomical location. p≤ .001 for all clusters. BA = Brodmann's area, ROI = region of interest, dlPFC = dorso-lateral prefrontal cortex, lPFC = lateral prefrontal cortex, SMA = supplementary motor area.*

However, three brain regions showed a significant VCON x VAL interaction for the response to NEG > POS images, as listed below. The *dorso-medial prefrontal cortex (dmPFC,* xyz= -3 33 51, Brodman area 8, see Figure 4a and Table 4) was significantly activated in the critical interaction test comparing the "natural" viewing condition EMEXP to REAP (i.e. [$\text{EMEXP}_{\text{NEG}}$ – $\text{EMEXP}_{\text{POS}}$] > [$\text{REAP}_{\text{NEG}}$ – REAP_{POS}]). A subsequent 3 (VCON) x 2 (VAL) x 2 (SC) ANOVA on beta values (in SPSS) confirmed this significant VCON x VAL interaction [F(1,18)= 4.89, p= .023] (see Figure 4d), and post-hoc comparisons showed that it was due to a selective elimination of increases to negative scenes during REAP [NEG > POS, t = 1.95, n.s.], unlike increases seen in the other two conditions [t(18)≥ 3.33, ps≤ .004]. Thus, the magnitude of the negative valence effect (activation *difference* between NEG *versus* POS images) was reliably reduced during REAP relative to EMEXP [t(18)= 2.96, p= .008] and to SUP [t(18)= 2.04, p= .057]; but there was no significant reduction of the valence effect during SUP compared to EMEXP [t(18)= 1.45, p= .16]. An ANOVA for each VAL condition separately confirmed that this pattern was produced by a selective modulation of responses to NEG scenes, as indicated by a main effect of VCON for NEG [F(1,18)= 3.96, p= .029] but not POS pictures [F(1,18)= .22, p= .80]. Further pairwise comparisons also confirmed that activation to NEG images was specifically decreased during REAP compared to EMEXP [t(18)= 2.84, p= .011], whereas the reduction during SUP relative to EMEXP failed to reach significance [t(18)= 1.71, p= .10]. There were no such modulation for POS images across VCON conditions [all t(18)≤ .727, ps≥ .48]. In addition, the full 3x2x2 ANOVA also indicated a clear main effect of valence [NEG > POS; F(1,18)= 24.35, p< .001], but no effect of social content [SOC > NSOC; F(1,18)= .365,p= .553] and no other interaction [VCON x SC: F(1,18)= 3.02, p= .10].

A second cluster in the *medial fusiform/lingual gyrus (mFG/LG,* xyz= -24 -66 -9) showed a similar interaction effect between valence and viewing condition (see Figure 4b and Table 4). The full 3x2x2 ANOVA on beta values also confirmed an interaction of VAL x VCON [F(1,18)> 5.16, p< .011] (see Figure 4e), while post-hoc comparisons indicated that it was due to a reduced valence effect (difference between NEG *versus* POS images) during REAP as compared with EMEXP [all t(18)> 3.21, p< .005], but also during SUP as compared with EMEXP [Figure 4c, yellow peak; t(18)= 3.25, p= .004]. Moreover, an ANOVA for each VAL condition again indicated that these changes were selective for the NEG images (main effect of VCON: [F(1,18)> 6.76, p< .003]), with significant decreases during both REAP and SUP relative to EMEXP [t(18)> 3.24, p< .005, and t(18)> 2.90, p< .01, respectively], and no significant difference between REAP and SUP [t(18)< 1.08, p> .30]. There were no such effects for POS images (main effect of VCON: [F(1,18)=.587, t= .561; all pairwise tests non significant]. In addition, the 3x2x2 ANOVA showed main effects of both the valence [NEG > POS; F(1,18)> 33.52, p< .001] and social content [NSOC > SOC; F(1,18)> 27.06, p< .001] of pictures, but no other interactions [VCON x SC: Fs(1,18)< 1.43, ps> .25].

The third activation was found in the right *anterior insula (aINS; xyz = 36 27 -3; see Figure 4c [yellow cluster] and Table 4).* A significant VCON x VAL interaction was confirmed by the full 3x2x2 ANOVA on beta values [F(1,18)> 4.76, p< .015], reflecting a significant decrease of the valence effect (difference between NEG *vs.* POS images) during both REAP [t(18)> 2, p< .05] and SUP [t(18)> 2.1, p< .05; paired t-tests], relative to EMEXP (see Figure 4f). Additional ANOVAs and pairwise comparisons again showed that these effects were specific to NEG scenes [main effect of VCON: F(1,18)> 5.47, p< .008; EMEXP *vs*. REAP: t(18)= 3.21, p= .005 and EMEXP *vs*. SUP: t(18)=.2.67, p= .016]; whereas no such modulation was present for POS scenes [main effect of VCON: F(1,18)≤ .739, p≥ .485]. An additional main effect of valence [NEG > POS; Fs(1,18)> 46.6, ps< .001] was also present, but there was no main effect of SC [NSOC > SOC; Fs(1,18)< 3.42, ps> .081] and no other interactions [VCON x SC: F(1,18)< 1.92, p> .16].

Taken together, these results indicate that both REAP and SUP had a selective impact on responses to negative scenes in the fusiform and insula, whereas in addition REAP produced a reduction of responses to negative scenes in the dorsomedial PFC. No differential impact of emotion regulation strategies was observed for the positive scenes. Only the medial fusiform was differentially sensitive to nonsocial content (presumably reflecting the predominance of objects [27] and places [28] in these images, see above), but this factor did not interact with the valence condition.

2.2. Regulation as a function of social content (VCON x SC interactions)

For differential responses to NSOC > SOC images, a selective interaction with viewing condition was found in the right middle insula (see Figure 5a and Table 4). An interaction effect was also observed in the mid-dorsal cingulate and left prefrontal cortex, but primarily driven by relative changes of the responses to SOC images across viewing conditions (data not shown), rather than reflecting modulations for NSOC images. Therefore, only data from the right middle insula (mINS) were further analyzed.

This right mINS cluster (Figure 5a) showed a selective activation to non-social emotional scenes that was significantly reduced during SUP relative to other viewing condition (i.e. [$\text{EMEXP}_{\text{NSOC}}$ – $\text{EMEXP}_{\text{SOC}}$] > [$\text{SUP}_{\text{NSOC}}$ – SUP_{SOC}], yellow cluster in Fig 5).

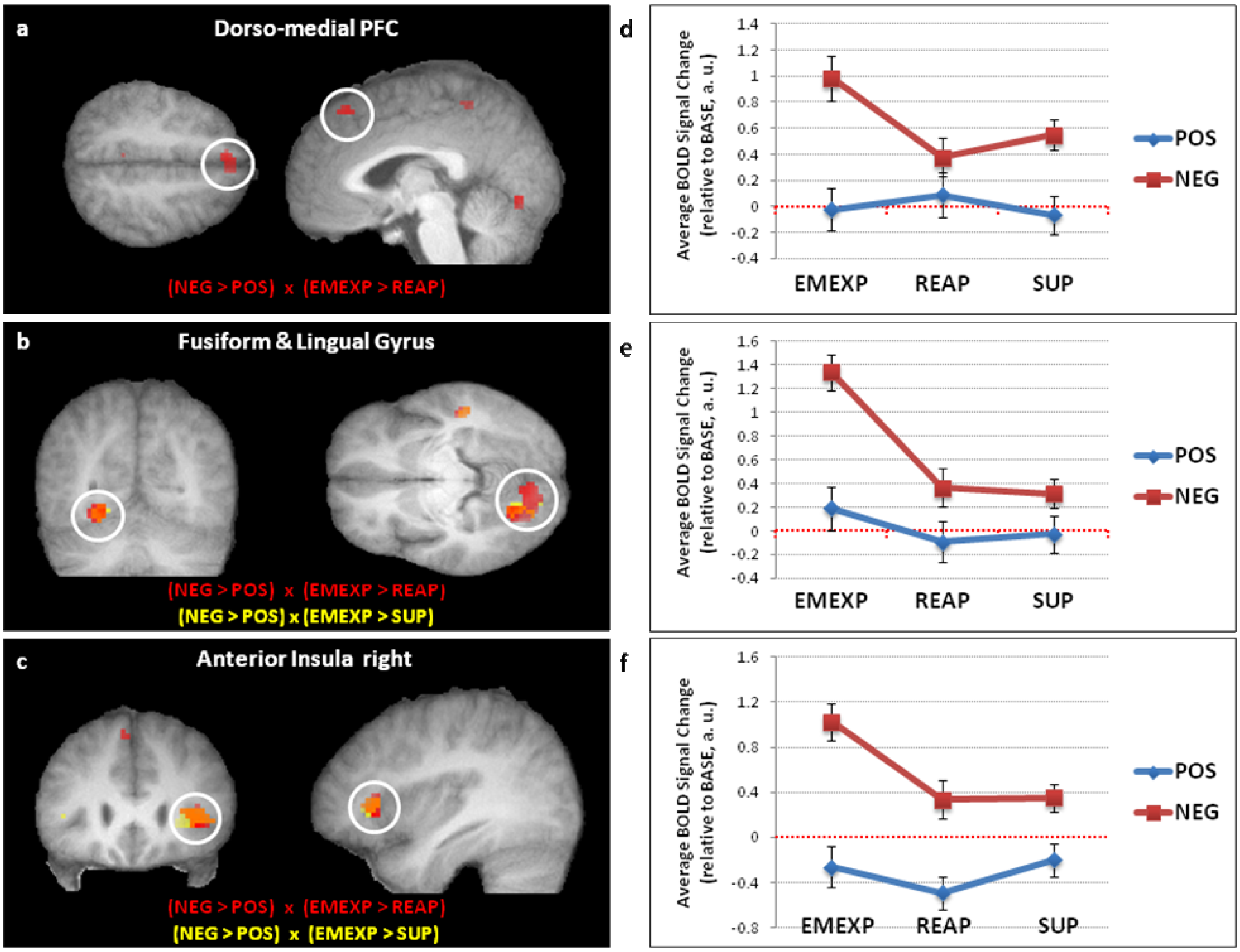


Figure 4: Interactions of emotion regulation with differential responses to negative versus positive scenes (VCON x VAL). *Statistical parametric maps are illustrated (threshold p= .025) for the paired t-test model comparing valence-specific responses across viewing conditions [NEG versus POS x EMEXP versus REAP (red); and/or NEG versus POS x EMEXP versus SUP (yellow)]. Significant effects were found in* ***a)*** *left dmPFC (xyz= -3 33 51),* ***b)*** *left medial fusiform (xyz= -24 -66 -9), and* ***c)*** *right anterior insula (xyz = 39 30 -3). Parameter estimates (beta values) extracted from the clusters in* ***d)*** *dmPFC,* ***e)*** *fusiform, and* ***f)*** *insula, averaged across voxels and participants, are plotted for the different stimulus and viewing conditions, showing greater activation to NEG versus POS pictures during EMEXP relative to REAP and SUP in all cases, but a total lack of valence effect in the dmPFC during REAP, and a significant reduction of responses to NEG valence in other areas and other conditions. All values are displayed relative to the neutral lbaseline condition (BASE), with ± 1 S.E.M.*

This VCON x SC interaction was confirmed by a full 3x2x2 repeated-measures ANOVAs on extracted beta values [F(1,18)= 7.40, p= .002], and observed because the relative activation difference between NSOC *versus* SOC images was selectively abolished during SUP as compared to both EMEXP [peak xyz= 39 12 -3; t(18)= 3.99, p= .001; Figure 5b, red cluster] and REAP conditions [peak xyz= 51 -3 -3; t(18)= 2.45, p= .025; Figure 5b, yellow cluster]. Additional ANOVAs and pairwise comparisons revealed that only responses to NSOC images were significantly modulated by the viewing condition [red cluster: main effect of VCON: F(1,18)= 7.57, p= .002; paired t-tests: EMEXP > REAP, t(18)= 3.28, p= .004; EMEXP > SUP, t(18)= 4.29, p< .001; REAP > SUP, t(18)= 1.71, p= .11]. But there was no significant modulation for SOC images [main effect of VCON: F(1,18)< 2.0 ps> .15; all paired t-tests n.s.].

Finally, in addition to a main effect of scene content [NSOC > SOC: Fs(1,18)> 17.55, ps< .001], the full repeated-measure ANOVA also revealed main effects of valence [NEG > POS; Fs(1,18)> 19.78, ps< .001] and viewing condition [Fs(1,18)> 5.11, ps< .01], but there was no VCON x VAL interaction [Fs(1,18)< .38, ps> .69].

For the opposite contrast SOC > NSOC, a significant interaction with VCON was found in several cortical brain areas as well as in the amygdala. First, two distinct clusters were found in the *medial PFC and OFC (mPFC, xyz = 3 54 21; and mOFC, xyz = 6 54 -14)* when comparing REAP to EMEXP or to SUP (Figure 6a, red and yellow voxels, respectively; and Table 4). The repeated-measure ANOVAs in SPSS confirmed a significant VCON x SC interaction [Fs(1,18)> 6.65, ps< .004] in both clusters, which was driven by the fact that the relative activationto SOC *versus* NSOC images was significantly reduced during REAP [all t(18)> 3.42, ps< .003] and to a lesser extent during SUP [all t(18)> 2.11, ps< .049] as compared to EMEXP, but not when comparing REAP with SUP [all t(18)< 1.81, ps> .09] (Figure 6d). Additional ANOVAs and pairwise comparisons confirmed that viewing conditions modulated the neural response to SOC images only [main effect of VCON: Fs(1,18)> 5.29, ps< .01; EMEXP > REAP in both clusters, all ts(18)> 2.84, ps< .011; EMEXP > SUP only in mPFC, t(18)= 3.65, p= .002;

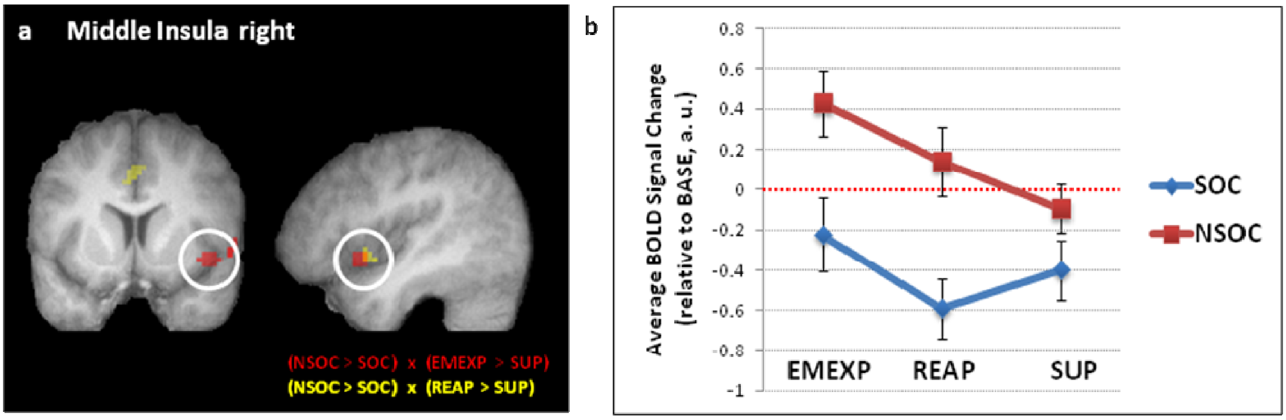


Figure 5: Interactions of emotion regulation with differential responses to non-social versus social scenes (VCON x SC). a) *Statistical parametric map (threshold p= .025) for the paired t-test model comparing selective responses to non-social scenes in the different viewing conditions [NSOC versus SOC x EMEXP versus SUP (red); and NSOC versus SOC x REAP versus SUP (yellow)], showing a significant effect in the right mid insula.* ***b)*** *Parameter estimates (beta values) for this region (as defined by the second cluster above, in yellow), averaged across voxels and participants, show a selective decrease of the NSOC versus SOC activation difference during SUP. All values are displayed relative to neutral baseline (BASE), with ± 1 S.E.M.*

REAP > SUP: all paired t-tests n.s.], but no significant modulation was found for NSOC images [main effect of VCON: Fs(1,18)< 3.24, ps> .06; paired t-test n.s.]. In both the mPFC and mOFC, there was also a clear main

A second effect was found in the *Posterior Cingulate Cortex (PCC)* (xyz= 0 -51 33, Figure 6b and Table 4)*:* The ANOVA in SPSS on extracted beta values also revealed a highly significant VCON x SC interaction [red voxels: F(1,18)= 10.71, p< .001; yellow: F(1,18)= 8.18, p= .001], reflecting a reduced activation difference between SOC *versus* NSOC images during REAP as compared to both EMEXP [red: t(18)= 5.54, p< .001; yellow: t(18)= 3.76, p= .001] and SUP [red: t(18)= 3.22, p= .005; yellow: t(1,18)= 4.44, p< .001], but not for SUP as compared to EMEXP [ts(18)≤ 1.22, ps≥ 2.38] (Figure 6d). Additional ANOVAs and pairwise comparisons confirmed the specificity of this modulation for SOC images [main effect of VCON: F(1,18)> 6.54, ps< .004]), with significant decreases for EMEXP > REAP [all t(18)> 2.97, ps< 0.008] and SUP > REAP [all t(18)> 2.12, ps< .048], but not for SUP > REAP [paired tests n.s.]. The modulation effects for NSOC images were not significant [main effect of VCON: F(1,18)< 2.01; p> .15; paired t-test: all ps≥ .06]. Finally, except for the additional main effect of social content [SOC > NSOC; Fs(1,18)> 68.94, ps< .001], the repeated-measure ANOVA showed no other effect or interaction [Fs(1,18)< .338, ps> .72].

A third region showing a marginal interaction effect was in the *right lateral fusiform gyrus (lFG,* xyz= 42 -42 -27; see Figure 6c and Table 4*):* Even though this region just failed to reach significance in the paired t-test models (p= .006) we nevertheless examined its activation pattern because of its major role in processing social stimuli such as faces and bodies [29], and its overlap with the fusiform ROI found in the initial simple main contrast SOC *versus* NSOC. Moreover, in the full ANOVA in SPSS using betas from this region, we also found a marginally significant VCON x SC interaction [F(1,18)= 3.21, p= .052]. This interaction reflected a specific reduction in the relative activation difference to SOC *versus* NSOC images during REAP, as compared with SUP [t(18)= 3.28, p= .004] but not EMEXP [t(18)= 1.45, t= .16]., and there was no significant reduction of this relative difference during

effect of social content [SOC > NSOC, Fs(1,18)> 36.12, ps< .001], but no other interaction [VCON x VAL, Fs(1,18)< 1.74, ps> .19].

SUP as compared with EMEXP [t(18)= .7, p= .49], despite an overall decrease in activity during SUP (see Figure 6f). The additional ANOVAs and pairwise comparisons revealed a significant effect of VCON for both SOC [F(1,18)= 5.26, p= .01] and NSOC images [F(1,18)= 3.59, p= .04], with reliable decreases when comparing EMEXP and SUP for both stimulus categories [all t(18)> 2.44, ps< .025], but a significant reduction between EMEXP and REAP only for SOC images [t(18)= 2.78, p= .012]. The full ANOVA also indicated main effect of VCON [F(1,18)= 4.62, p= .02], and no other interaction [F(1,18)= 2.43, p= .10].

Finally, an interaction was found in *bilateral Amygdala* (see Figures 7ac and Table 4), but with a different pattern of effects for each regulation condition on each side.

In the left amygdala, a significant VCON x SC interaction emerged [F(1,18)= 6.19, p= .005, red] because the relative activation to SOC *versus* NSOC scenes was entirely abolished during REAP [SOC > NSOC, t(18)= 1.76, n.s.] and thus significantly differed from the activation difference to SOC vs NSOC during EMEXP [t(18)= 3.46, p= .003], This preferential amygdala activation to SOC scenes still persisted during SUP [SOC > NSOC, t= 2.81, p= .012], even though it was also reduced in this condition relative to EMEXP [t(18)= 2.27, p= .04] (see Figure 7c). Again, these modulations were specific for SOC images, as demonstrated by the additional ANOVAs showing a significant main effect of VCON for the SOC stimuli only [F(1,18)= 4.68, p= .016]; but not for NSOC stimuli [F(1,18)= .675, p= .456]. Likewise, paired t-tests indicated significant changes for SOC stimuli when comparing EMEXP > REAP [t(18)= 2.54, p= .02] and EMEXP > SUP [t(18)= 3.25, p= .004]; but no such changes for NSOC stimuli [all ts(18)< 1.30, ps> .21].

In the right amygdala (Figure 7b), however, a significant VCON x SC interaction [F(1,18)=6.41, p= .004, yellow] was found because of a selective elimination of the relative activation to SOC *versus* NSOC scenes during SUP [t(18)= .61, n.s.]. This

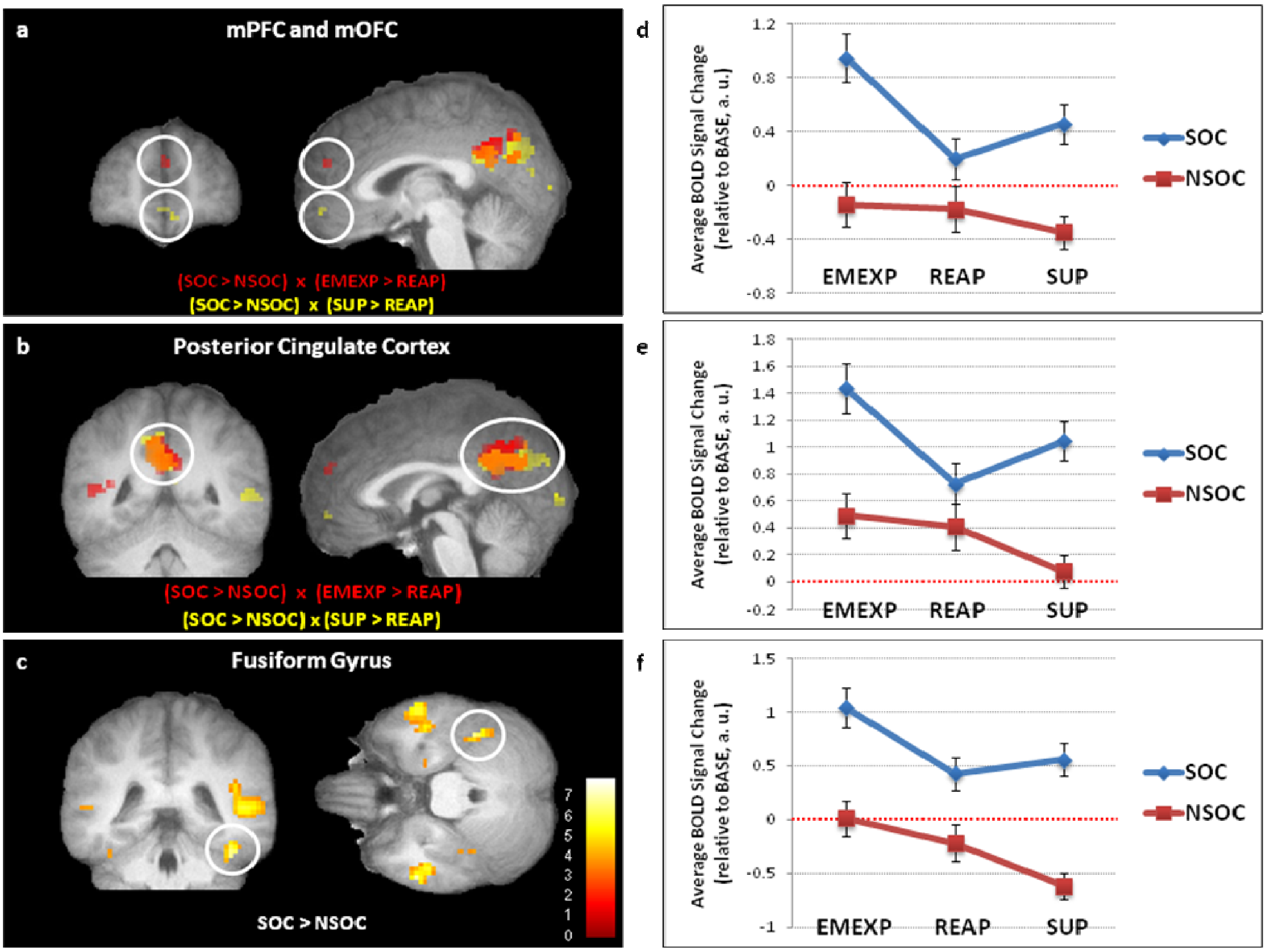


Figure 6: Interactions of emotion regulation with differential responses to social versus non-social scenes (VCON x SC). *Statistical parametric maps are illustrated (threshold p= .025) for the paired t-test model comparing activations to social stimuli across different viewing conditions, showing interaction effects **a)** in mPFC (xyz= 3 54 21) for the contrast [SOC versus NSOC x EMEXP versus REAP (red)] and in mOFC (xyz= 6 54 -15) for the contrast [SOC versus NSOC x SUP versus REAP (yellow)]; **b)** in PCC (xyz= 0 -51 33) for the contrasts [SOC versus NSOC x EMEXP versus REAP (red)] and [SOC versus NSOC x EMEXP versus SUP (yellow)]; and **c)** in right lateral fusiform gyrus (xyz= 42 -42 27) for the contrast SOC > NONSOC. Parameter estimates (beta values) extracted from **d)** the mPFC, **e)** PCC, and **f)** lateral fusiform clusters, averaged across voxels and participants, showed consistent modulations for SOC images during REAP in all three regions. All values are displayed relative to neutral baseline (BASE), with ± 1 S.E.M.*

differed significantly from the relative activation to SOC stimuli during both EMEXP [t(18)= 3.32, p= .004] and REAP [t(18)= 2.64, p= .017]. There was no significant difference between EMEXP and REAP [t(1,18)= 1.16, p= .26] (see Figure 7d). Additional ANOVAS and paired t-tests also revealed a selective effect of VCON on responses to SOC images [F(1,18)= 3.38, p= .05], but no effect for NSOC images [F(1,18)= .50, p= .61], and no other interactions.

To verify this hemispheric asymmetry between left and right amygdala, we computed an additional 2 (SIDE) x 3 (VCON) x 2 (VAL) x 2 (SC) repeated-measures ANOVA, which revealed a significant triple interaction between SIDE x VCON x SC [F(1,18)= 5.05, p= .012]. This triple interaction was driven by two factors: Firstly, the activation difference between SOC *versus* NSOC images showed a significant decrease during REAP as compared with EMEXP in the left amygdala only [left: t(18)= 3.46, p= .003; right: t(18)= 1.16, p= .26], leading to a marginally significant 2 (SIDE) x 2 (VCON) interaction [F(1,18)= 3.60, p= .072]; Seond, the activation difference between SOC *versus* NSOC images showed a significant decrease during SUP as compared to REAP in the right amygdala only [right: t(18)= 2.64, p= .017; left: t(18)= 1.35, p= .193], leading to a significant 2 (SIDE) x 2 (VCON) interaction [F(1,18)= 9.85, p= .006].

The repeated-measure ANOVA indicated additional main effects of social content bilaterally [SOC > NSOC; left: F(1,18)= 19.13, p< .001; right: F(1,18)= 10.11, p= .005], but no main effects of VAL and no other interactions [VCON x VAL: Fs(1,18)< 1.63, ps> .21].

Thus, altogether, these data reveal distinct neural effects of each emotion regulation condition as a function of the stimulus content. There was a selective effect of SUP on responses to non-social emotional scenes in the insula, whereas responses to social scenes were specifically modulated by SUP in the right amygdala, but by REAP in the left amygdala as well as several cortical areas implicated in affect and social cognition, including medial PFC, PCC, and (to a lesser extent) lateral fusiform gyrus.

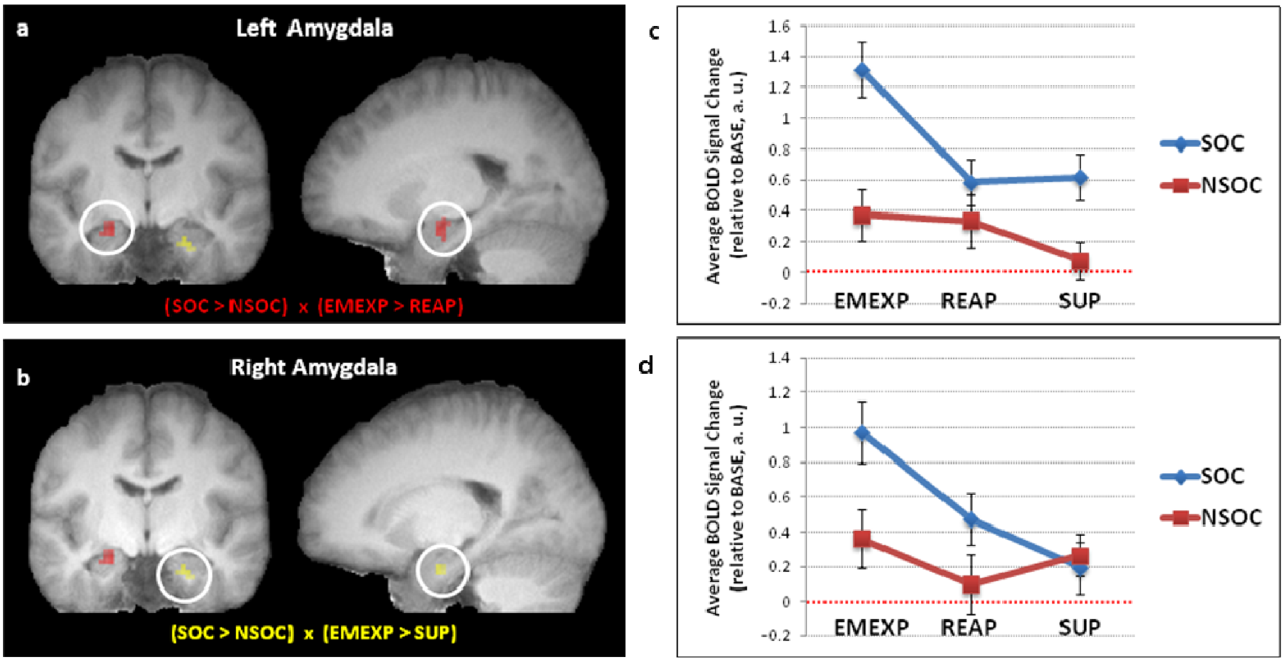


Figure 7: Amygdala Lateralization during Emotion Regulation.*Statistical parametric maps for the paired t-test model (threshold p= .025) comparing activations to social stimuli across different viewing conditions, showing a significant interaction effects in* ***a)*** *the left amygdala (xyz= -21 -9 -18) for the contrast [SOC versus NSOC x EMEXP versus REAP (red)]; and* ***b)*** *the right amygdala (xyz= 24 -9 -27) for the contrast [NSOC versus SOC x EMEXP versus SUP (yellow)]. Parameter estimates (beta values) extracted from the* ***c)*** *left and* ***d)*** *right amygdala clusters, averaged across voxels and participants, show different regulation effects in each hemisphere. The left amygdala displayed a selective elimination of differential responses to social (vs non-social) images during REAP only. Conversely, the right amygdala exhibited a selective elimination of the social content effect during SUP only. This pattern led to a significant three-way interaction of hemisphere x scene content x regulation condition (see text). All values are displayed relative to neutral baseline (BASE), with ± 1 S.E.M.*

DISCUSSION

The current fMRI study aimed at differentiating the impact of emotion regulation (ER) on activity in limbic and cortical brain areas, as a function of the nature of affective cues, i.e. when related to social (SOC) *versus* non-social (NSOC) information or when associated with negative (NEG) *versus* positive (POS] valence. To this aim, ER using either reappraisal (REAP) or suppression (SUP) strategies were systematically compared with a natural “spontaneous” viewing condition (EMEXP), across different stimulus conditions

Behavioral ratings of visual scenes during fMRI scanning indicated that REAP was more efficient than SUP in modulating the subjective intensity of emotion experience, for all types of scenes. This finding is consistent with previous data on the volitional control of emotions, particularly in response to negative stimuli [1, 6, 7]. By contrast, SUP produced a significant reduction of emotion intensity ratings for positive scenes only (not negative scenes). However, the social content (SC) of stimuli had no influence on subjective emotion experience, as evidenced by similar effects of REAP and SUP on the rating scores for both SOC and NSOC emotional scenes. This different impact of the two ER conditions accords with the notion that SUP primarily regulates behavioral expressions of emotions but not the experienced emotional feelings [3]. But it is also possible that the lack of effect of SUP on the ratings of negative stimuli reflected their higher arousal values relative to the positive stimuli (see methods), although this difference had no such effect in the REAP condition. Because social stimuli typically tend to be judged as more arousing than non-social stimuli [18], we could not totally match emotion intensity between positive and negative scenes in order to balance the more crucial comparisons of social and non-social content in both valence conditions. Thus, SOC and NSOC images were well matched on arousal (and control) properties, and showed similar changes in subjective ratings across the different emotion regulation conditions. Nevertheless, despite this similarity in the ratings, our fMRI results revealed specific effects of ER on the response of several brain areas as a function of the different content of emotional scenes. These results therefore point to specific neural sites where emotional processing can be modified by ER (e.g. in limbic and sensory areas), in addition to the possible neural sources where modulatory influences are presumably generated (e.g. in prefrontal cortices).

PREFRONTAL CORTICAL REGIONS AS A PUTATIVE SOURCE OF EMOTION REGULATION

By directly contrasting brain activation in the two ER conditions with each other (REAP > SUP and SUP > REAP), we identified a set of regions in the right prefrontal cortex (PFC) whose activity was differentially increased either during REAP (superior and middle frontal gyrus; SFG and MFG, respectively) or during SUP (superior frontal sulcus and supplementary motor area; i.e. SFS and SMA, respectively), but regardless of the content of emotional scenes (no main effects or interactions due to the factors valence or social content). Thus, these regions were modulated by the nature of the regulation strategy but not by the nature of emotional information to be regulated (except for MFG that was more

VCON x SC Interactions

NSOC > SOC

REGION	*Contrast*	*BA*	*Voxel*	*T-Value (Max Vox)*	*xyz*
Insula middle right	EMEXP > SUP		30	3.92	39 12 -9
Insula middle right	REAP > SUP		20	5.06	51 -3 -3

SOC > NSOC

REGION	*Contrast*	*BA*	*Voxel*	*T-Value (Max Vox)*	*xyz*
Amygdala left *	EMEXP > REAP		22	3.55	-21 -9 -18
Amygdala right	EMEXP > SUP		11	3.77	24 -9 -27
mOFC	EMEXP > SUP	11	14	4.19	6 54 -15
mPFC	EMEXP > REAP	10	11	3.94	3 54 21
PCC / PC	EMEXP > REAP	23	332	7.32	0 -51 33
PCC / PC	SUP > REAP	23	381	7.32	0 -51 33
FG lateral right	SOC > NSOC (*)	37	98	6.59	42 -42 -27

VCON x VAL Interactions

NEG > POS

REGION	*Contrast*	*BA*	*Voxel*	*T-Value (Max Vox)*	*xyz*
Insula anterior right	EMEXP > REAP		65	7.49	39 30 -3
Insula anterior right	EMEXP > SUP		81	6.29	36 27 -3
dmPFC	EMEXP > REAP	8	25	5.28	-3 33 51
FG/LG left	EMEXP > REAP	18	251	8.12	-24 -66 -9
FG/LG left	EMEXP > SUP	18	40	8.12	-24 -66 -9

Table 4: Brain Areas displaying significant VCON x SC or VCON x VAL interactions. Peak coordinates are given in MNI space and listed with best estimates of anatomical location. p≤ .001 for all clusters (T-value for maximal voxel > 3.61, except fot *T= 3.55). BA = Brodmann's area, ROI = region of interest, mOFC = medial orbitofrontal cortex, mPFC= medial prefrontal cortex, PCC= posterior cingulate cortex, PC = precuneus, FG = fusiform gyrus, LG = lingual gyrus.(*)= derived from main effects analysis and not from a paired t-test model.

activated by negative social scenes than other pictures). This pattern suggests a general role of the right PFC in processes subserving ER, rather than in the representation of emotional information per se.

Anatomically, these activations accord with previous studies describing increased activity in dorsolateral PFC during cognitive up- and down-regulation of both positive and negative emotion [4-7], as well as during behavioral suppression of negative affect [7, 8]. Such activations are thought to subserve the maintenance of appraisal instructions or scripts used to re-interpret stimuli, as well as the resistance to interference by emotionally salient inputs, implicating in particular the SFG [4, 5], in addition to behavioral inhibition and voluntary suppression of bodily expressions, implicating more inferior frontal areas in SFS and lateral OFC [8]. Several regions in right dorsolateral PFC have also been involved in controlling task-set or stimulus-response mapping in various conditions, suggesting a general role in the regulation of both cognitive and affective processing[30]. On the other hand, activation in the SMA was similarly increased during SUP and natural viewing (EMEXP), relative to both the neutral baseline condition and REAP. This is consistent with a role of medial prefrontal areas in the planning and monitoring of affective motor behavior [31], which might operate during both SUP and EMEXP but not when emotion experience was diminished by REAP (as indicated by subjective ratings).

Thus, our data provide new evidence for a functional anatomical segregation within PFC, including regions more specifically implicated in REAP (SFG and MFG) and others implicated in SUP (SFS and SMA). These functions could be distinguished independent of any difference in ther temporal dynamics as postulated in a previous study comparing REAP and SUP [7]. In the latter study, brain regions associated with different ER strategies were primarily determined by imputing early responses after stimulus onset to REAP, while later responses after stimulus onset were ascribed to SUP. Here, because ER was applied in a block-design, we could not reliably compare the time-course of stimulus-evoked responses in the different viewing conditions.

Even though there were no main effects of VAL or SC in right PFC, the MFG was the only area that displayed some degree of stimulus-related specificity during REAP. Signal changes in right MFG during REAP showed a selective SOC *versus* NSOC effect for negative but not positive images, implying that its activity was the highest for negative social information. These findings suggest that the down-regulation of NEG emotion during REAP might require more cognitive processing or more effort for SOC as opposed to NSOC stimuli, presumably because negative social information has a stronger intrinsic relevance for affect and behavior than non-social features [11].

EMOTION REGULATION AS A FUNCTION OF VALENCE

We were able to identify three brain areas where different ER conditions selectively influenced the processing of negative valence information (NEG *versus* POS emotional images), including the dorso-medial PFC (dmPFC), anterior insula (aINS), and medial fusiform / lingual gyrus (FG/LG). By contrast, no specific effects were found for the processing of positive valence (POS *versus* NEG emotional pictures). Because NEG images were generally perceived as more arousing than POS images (see above), these effects of valence (VAL) might also be at least partly attributed to the arousing value of visual scenes. However, both arousal and valence represent the most salient affective dimensions that can influence perception and attention towards emotional stimuli [32, 33]. Importantly, these differences in arousal did not confound our critical comparisons between social and non-social stimuli (see below).

In the dmPFC, the relative increase in response to NEG *versus* POS scenes was selectively abolished during REAP (as compared to EMEXP as well as SUP). Activation of this pre-frontal brain area in the context of ER has previously been described during up- and down-regulation of either negative [6] or positive [4] emotions. In addition, the dmPFC has previously been reported to mediate self-monitoring and self-evaluation processes through the computing of internal representations of one's own and others' mind [34, 35], and to contribute to the continuous updating of the value of actions (or errors) in order to regulate behavior [36, 37]. Furthermore, previous work on dmPFC responses to emotional stimuli also suggests that this brain region may critically be involved in explicit affective judgments [38-40], and preferentially activates to emotionally arousing conditions [41, 42]. Accordingly, our new data suggests that REAP might reduce the cognitive aspects of emotion processing in dmPFC, related to self-monitoring and evaluation of behavioral actions, when emotional significance of visual scenes is lessened by the "pretend unreal" instructions. This reduction in activity nicely dovetails with our behavioral data showing a significant decrease of subjective emotion ratings during REAP (but not SUP).

By contrast, the anterior insula (aINS) activation to negative stimuli was attenuated during both REAP and SUP. Previous imaging work has consistently demonstrated that the aINS is involved in the representation of aversive stimuli, particularly disgust [43, 44] and pain [45]. In particular, aINS responses are specifically linked to the affective, but not sensory, components of pain [45]. Moreover, its activation reflects the degree of bodily arousal evoked by the negative stimulus content, which might play a mediating role to interface cognitive control with changes in autonomic responses during ER [24]. Because negative scenes in our study often displayed people experiencing pain or distressing conditions (SOC category), as well as situations representing harmful or disgust conditions (NSOC category), the decreases in aINS activation might be interpreted in terms a reduction of the affective components of empathic responses and/or disgust. However, inspection of aINS responses across conditions showed a non-significant preference for NSOC over SOC stimuli (see result section 2.1.), and a selective effect of REAP and SUP on the processing of NSOC stimuli (but not SOC) was also found in a partly overlapping region of the (middle) insula (see result section 2.2.). Taken together, these data indicate that insula responses were mainly driven by non-social negative scenes, and presumably related to disgust cues more than to pain or empathy. This interpretation is corroborated by the main effects of NEG > POS scenes and NSOC > SOC scenes that both revealed significant activations in the insula (see Tables 1 and 2). Hence, our results demonstrate that both REAP and SUP strategies produced a convergent modulatory effect on emotional processing of non-social aversive information in the insula, a brain area known to play a central role in the autonomic and subjective dimensions of affective responses [24].

We also found that the processing of negative stimuli was selectively reduced by both REAP and SUP in the medial fusiform / lingual cortex. These regions showed preferential responses to non-social stimuli, consistent with their role in the visual processing of complex objects and scenes [27, 28]. Their modulation by ER accords with top-down influences on sensory processing in extrastriate visual areas for emotionally arousing and threat-related (e.g. fear conditioned) stimuli [32, 46, 47], which may be driven by greater attention allocation and/or direct influences from amygdala [26, 33]. Accordingly, the reduction of differential responses to NEG *versus* POS scenes during both REAP as well as SUP in visual cortex indicates a decrease in attention allocation to the sensory content of emotional scenes as a consequence of the cognitive or behavioral processes engaged by ER. During REAP, participants had to look at the scenes while imaging an artificial setting ("pretend unreal" as in TV or cinema); whereas during SUP, they had to concentrate on the inhibition of external behavioral manifestations of emotion, thereby altering emotional responses and limiting attentional resources.

In sum, we found distinct effects of ER on the representation of negative emotional information, partly reflecting a modulation of the high arousal value of these stimuli, and involving several processing stages in distinct brain areas. These effects included changes at the perceptual (visual cortex), affective (insula), and cognitive levels (dmPFC). Whereas the cognitive processes mediated by dmPFC were more specifically affected during REAP, the visual processes in extrastriate cortex and affective responses in insula were modulated during both REAP and SUP.

EMOTION REGULATION AS A FUNCTION OF SOCIAL CONTENT

Differential responses to social (*versus* non-social) emotional images were attenuated by ER in medial prefrontal (mPFC) and medial orbitofrontal (mOFC) cortex, posterior cingulate cortex (PCC), and lateral fusiform gyrus (see Figure 6). All these regions have previously been shown to be implicated in social cognition, although related to distinct aspects of the latter. In addition, all these regions were selectively modulated by REAP (not SUP), except for mPFC that was equally modulated by both ER conditions.

Abundant evidence points to a major role of mPFC in mentalizing about self and others [35, 48], including in particular the representation of personal emotional experiences and dispositions (called reflective awareness; [49]). Likewise, increased activity in PCC has been observed during tasks involving the attribution of emotion to self and/or other [50, 51] and during theory of mind [52], besides an activation to social stimuli and affective information about familiar persons in general [12]. Therefore, the increased activity in mPFC and PCC for SOC *versus* NSOC images in our study was most likely related to mentalizing processes (including theory of mind) and the attribution of emotional experiences to humans seen in social images, as opposed to the inanimate displays seen in nonsocial images. Moreover, our results suggest that these attributional processes were attenuated by both REAP and SUP strategies. In contrast, mOFC activation was significantly reduced during REAP only, and is generally considered to play a major role in reinforcement-guided decision making, especially in terms of the context-sensitive evaluation of outcomes [53]. The reduced activation of mOFC to SOC images during REAP might therefore correspond to a decrease in cognitive operations required to compute possible outcomes of social scenarios displayed in these scenes, involving for example intentionality – unlike the mental representation of more basic emotional cues such as appetite/food desire and disgust, as typically depicted in non-social scenes [11].

Finally, we observed that REAP also reduced the differential responses to SOC *versus* NSOC scenes in the right lateral fusiform gyrus, an area known to be preferentially activated to images depicting faces (so called fusiform face area, FFA) and/or bodies and body parts (so called fusiform body area, FBA) as opposed to objects [54, 55]. This modulation of FFA/FBA is likely to represent changes in emotional processing and attention for human animate characters in these images, which by essence included many faces and bodies – unlike the non-social images. Such a pattern most likely emerged through top-down influences from other brain areas on visual information encoding in lateral fusiform [47, 56, 57] during REAP, similar to the effects observed in more medial extrastriate visual areas for negative non-social images (see above, modulation of emotional valence).

AMYGDALA LATERALIZATION DURING EMOTION REGULATION

Finally, and importantly, our results unveil for the first time a significant hemispheric lateralization in amygdala responses as a function of both the stimulus content and the ER strategy used (see Figure 7). While the bilateral amygdala was generally more responsive to social than non-social scenes, this activation showed a distinct impact of REAP and SUP in the left and right hemisphere. Firstly, REAP (but not SUP) led to a complete elimination of the differential activation to SOC *versus* NSOC images only in left amygdala. Conversely, in the right amygdala, the differential responses to SOC *versus* NSOC images were selectively abolished during SUP (but not REAP). In other words, cognitive re-evaluation (REAP) of social emotional information predominantly influenced neural activity in the left amygdala, whereas the right amygdala activation to the same scenes was more influenced by behavioral inhibition of emotional expression (SUP).

Although this lateralization of emotion regulation effects in the amygdala was unexpected, it is consistent with asymmetries in activation reported by a few previous fMRI studies investigating cognitive up- and down-regulation of negative and positive emotion. On the one hand, two studies found that up- and down-regulation of negative emotions using REAP affected activity in the left amygdala more significantly than the right amygdala [4, 6]. On the other hand, another study found that down-regulation of sexual arousal influenced right amygdala activation selectively [10]. These data therefore suggest that up- and down-regulation of emotion in left amygdala may reflect the use of more cognitive, perhaps verbally mediated strategies, whereas regulation in the right amygdala may be more related to nonverbal strategies specific to the non-dominant hemisphere [4, 6]. Accordingly, a recent meta-analysis on lateralized amygdala activations in relation to emotional memory [58] proposed that: 1) the right amygdala would be involved in the fast, relatively automatic detection of emotional stimuli, and thus capable of modulating other cognitive processes such as attention and memory; and 2) the left amygdala would be involved in a more sustained and detailed cognitive processing stage. Taken together, these previous accounts accord with our new data indicating that the left amygdala responses to social (negative) emotional stimuli might be more strongly influenced by cognitive re-evaluation (REAP) using verbal scripts, whereas emotion processing in the right amygdala might more readily be modulated by behavioral inhibition (SUP) without any explicit cognitive evaluative component.

Importantly, it should be emphasized that these bilateral amygdala activations were driven by distinctive responses to SOC *versus* NSOC stimuli, but we did not find significant amygdala activation for the contrasts between negative versus positive scenes. This pattern is consistent with data from several previous experiments describing preferential amygdala activation to social stimuli like faces, images, or film-clips [11-13, 21]. More generally, such results also accord with recent proposals that the amygdala might be particularly tuned to the intrinsic salience or high biological importance of social stimuli [59], rather than to threat or other specific emotion categories.

Finally, our findings do not entirely corroborate a previous fMRI study that also compared REAP and SUP, but found increased or sustained insula and amygdala activation during the suppression of negative emotions [7]. However, this study identified the effects of SUP at a late period (10.5. – 15 sec) after exposure to brief negative emotional film-clips, based on the assumption that SUP should operate only after emotions have been generated, and therefore later than REAP effects [3]. Because here, we applied a standard event-related paradigm with an image exposure time of 2 sec, the results of these two studies are hard to compare. In addition, unlike ours, the study of Goldin et al. [7] did no distinguish between responses to social and non-social emotional stimuli, a dimension that clearly modulated neural responses in both the amygdala and insula. In any case, more research is needed to further elucidate the distinct time-courses of SUP and REAP on activity in subcortical and cortical brain regions involved in emotional processing and subsequent physiological (e.g. autonomic) responses.

CONCLUSION

By directly comparing cognitive re-evaluation (REAP) and behavioral inhibition (SUP) as emotion regulation strategies, and differentiating their effect on distinct dimensions of emotional stimuli (including social *versus* non-social significance besides positive *versus* negative valence factors), the present study yields important new insights into the neural mechanisms of emotion regulation in humans.

On the one hand, we found that several regions in right dorsolateral PFC were differentially activated by emotional scenes during REAP or SUP, regardless of the content of emotional images, suggesting a general role in controlling regulatory processes. Only the right MFG showed greater increases for negative social scenes compared to other image categories, presumably reflecting the greater emotional saliency of these stimuli, and hence the greater demands for successful regulation.

On the other hand, we identified several sites where emotion regulation had an impact on specific features of emotion processing. Some regions were selectively modulated by REAP, including the right dmPFC, which was preferentially engaged by negative valence information, as well as the left amygdala, lateral fusiform, PCC, and medial OFC, which all were involved in processing social information. Altogether, these modulations are likely to reflect changes in self-monitoring, evaluation of personal relevance, and anticipation of potential outcomes, under the influence of REAP strategies. Because emotional responses to the NEG or SOC conditions in those areas were selectively attenuated during REAP but not SUP, our findings suggest that a predominant modulation of high-level cognitive representations, social judgment, mental imagery, and/or self-relevant association processes when viewing emotional images during the “pretend unreal” or “TV movie” conditions. Conversely, the right amygdala was the only region to be more strongly modulated by SUP than REAP, possibly reflecting a specific impact of SUP on behavioral and autonomic physiological components of emotions mediated by the right amygdala. This predominant right-sided effect of SUP contrasted with the stronger left-sided effect of REAP, although both amygdalae were equally responsive to social information overall. This result reveals a significant lateralization in emotion regulation effects on the amygdalae, probably related to a differential impact of verbal, cognitive strategies (REAP) and non-verbal, arousal-related strategies (SUP) on the left and right side, respectively.

Finally, several brain regions were similarly modulated by both REAP and SUP. These included the right insula and extrastriate visual areas (medial fusiform / lingual gyrus), which were most sensitive to negative non-social scenes involving disgust and other harmful situations, but also the right mPFC which preferentially responded to social stimuli. These activation patterns reflect common influences of the two ER strategies on attention and perception, modulating low-level visual processing, changes in affective processes subserving arousal and visceral or pain-related representations, as well as higher cognitive effects related to mentalizing about self and others.

Taken together, our findings therefore do not only highlight the distributed nature of neural changes induced by emotion regulation, but also reveal the selectivity of impact for different strategies (REAP or SUP) and for different dimensions of emotional information (SOC or NEG). This in turn underscores the importance of comparing different cognitive strategies and different stimulus categories when investigating emotion regulation, in order to better understand the specificity and underlying mechanisms of these effects. Only by carefully taking into account these different factors, we will be able to better apprehend the effective impact of emotion regulation and eventually develop more efficient intervention therapies in clinical disorders related to emotion regulation dysfunctions.

ACKNOWLEDGMENT

This research was supported by the National Center of Competence in Research (NCCR) Affective Sciences financed by the Swiss National Science Foundation (n° 51NF40-104897) and hosted by the University of Geneva, and grants of the Swiss National Science Foundation to DS and PV.

REFERENCES

1. Gross, J.J., *Emotion regulation: Affective, cognitive, and social consequences.* Psychophysiology, 2002. **39**(3): p. 281-291.
2. Jackson, D.C., et al., *Suppression and enhancement of emotional responses to unpleasant pictures.* Psychophysiology, 2000. **37**(4): p. 515-522.
3. Gross, J.J., *Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology.* Journal of Personality and Social Psychology, 1998. **74**(1): p. 224-237.
4. Kim, S.H. and S. Hamann, *Neural correlates of positive and negative emotion regulation.* Journal of Cognitive Neuroscience, 2007. **19**(5): p. 776-798.
5. Ochsner, K.N., et al., *Rethinking feelings: An fMRI study of the cognitive regulation of emotion.* Journal of Cognitive Neuroscience, 2002. **14**(8): p. 1215-1229.
6. Ochsner, K.N., et al., *For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion.* Neuroimage, 2004. **23**(2): p. 483-499.
7. Goldin, P.R., et al., *The neural bases of emotion regulation: Reappraisal and suppression of negative emotion.* Biological Psychiatry, 2008. **63**(6): p. 577-586.
8. Levesque, J., et al., *Neural circuitry underlying voluntary suppression of sadness.* Biological Psychiatry, 2003. **53**(6): p. 502-510.
9. Gross, J.J. and O.P. John, *Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being.* Journal of Personality and Social Psychology, 2003. **85**(2): p. 348-362.
10. Beauregard, M., J. Levesque, and P. Bourgouin, *Neural correlates of conscious self-regulation of emotion.* Journal of Neuroscience, 2001. **21**(18).
11. Hariri, A.R., et al., *The amygdala response to emotional stimuli: A comparison of faces and scenes.* Neuroimage, 2002. **17**(1): p. 317-323.
12. Britton, J.C., et al., *Neural correlates of social and nonsocial emotions: An fMRI study.* Neuroimage, 2006. **31**(1): p. 397-409.
13. Norris, C.J., et al., *The interaction of social and emotional processes in the brain.* Journal of Cognitive Neuroscience, 2004. **16**(10): p. 1818-1829.
14. Goossens, L., et al., *Selective Processing of Social Stimuli in the Superficial Amygdala.* Human Brain Mapping, 2009.
15. Killgore, W.D.S. and D.A. Yurgelun-Todd, *Social anxiety predicts amygdala activation in adolescents viewing fearful faces.* Neuroreport, 2005. **16**(15): p. 1671-1675.
16. Vrtička, P., et al., *Individual attachment style modulates human amygdala and striatum activation during social appraisal.* PLoS ONE, 2008. **3**(8): p. e2868.
17. Fujita, F., E. Diener, and E. Sandvik, *Gender Differences in Negative Affect and Well-Being - the Case for Emotional Intensity.* Journal of Personality and Social Psychology, 1991. **61**(3): p. 427-434.
18. Ewbank, M.P., et al., *The amygdala response to images with impact.* Social Cognitive and Affective Neuroscience, 2009. **4**(2): p. 127-133.
19. Ritchey, M., F. Dolcos, and R. Cabeza, *Role of Amygdala Connectivity in the Persistence of Emotional Memories Over Time: An Event-Related fMRI Investigation.* Cerebral Cortex, 2008. **18**(11): p. 2494-2504.

20. Lazar, N.A., et al., *Combining brains: A survey of methods for statistical pooling of information.* Neuroimage, 2002. **16**(2): p. 538-550.

21. Goossens, L., et al., *Selective Processing of Social Stimuli in the Superficial Amygdala.* Human Brain Mapping, 2009.

22. Harris, L.T., et al., *Regions of the MPFC differentially tuned to social and nonsocial affective evaluation.* Cognitive Affective & Behavioral Neuroscience, 2007. **7**(4): p. 309-316.

23. van den Bos, W., et al., *Dissociating affective evaluation and social cognitive processes in the ventral medial prefrontal cortex.* Cognitive Affective & Behavioral Neuroscience, 2007. **7**(4): p. 337-346.

24. Critchley, H.D., et al., *Volitional control of autonomic arousal: A functional magnetic resonance study.* Neuroimage, 2002. **16**(4): p. 909-919.

25. Vuilleumier, P., et al., *Effects of attention and emotion on face processing in the human brain: An event-related fMRI study.* Neuron, 2001. **30**(3): p. 829-841.

26. Vuilleumier, P., et al., *Distant influences of amygdala lesion on visual cortical activation during emotional face processing.* Nature Neuroscience, 2004. **7**(11): p. 1271-1278.

27. Pourtois, G., et al., *Object Representations for Multiple Visual Categories Overlap in Lateral Occipital and Medial Fusiform Cortex.* Cerebral Cortex, 2009. **19**(8): p. 1806-1819.

28. Haxby, J.V., et al., *Distributed and overlapping representations of faces and objects in ventral temporal cortex.* Science, 2001. **293**(5539): p. 2425-2430.

29. Peelen, M.V. and P.E. Downing, *Within-subject reproducibility of category-specific visual activation with functional MRI.* Human Brain Mapping, 2005. **25**(4): p. 402-408.

30. Brass, M., et al., *The role of the inferior frontal junction area in cognitive control.* Trends in Cognitive Sciences, 2005. **9**(7): p. 314-316.

31. Morecraft, R.J., K.S. Stilwell-Morecraft, and W.R. Rossing, *The motor cortex and facial expression: New insights from neuroscience.* Neurologist, 2004. **10**(5): p. 235-249.

32. Sabatinelli, D., et al., *Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance.* Neuroimage, 2005. **24**(4): p. 1265-1270.

33. Vuilleumier, P. and Y.M. Huang, *Emotional Attention: Uncovering the Mechanisms of Affective Biases in Perception.* Current Directions in Psychological Science, 2009. **18**(3): p. 148-152.

34. Harris, L.T., A. Todorov, and S.T. Fiske, *Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind.* Neuroimage, 2005. **28**(4): p. 763-769.

35. Mitchell, J.P., C.N. Macrae, and M.R. Banaji, *Dissociable medial prefrontal contributions to judgments of similar and dissimilar others.* Neuron, 2006. **50**(4): p. 655-663.

36. Amodio, D.M. and C.D. Frith, *Meeting of minds: the medial frontal cortex and social cognition.* Nature Reviews Neuroscience, 2006. **7**(4): p. 268-277.

37. Botvinick, M.M., *Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function.* Cognitive Affective & Behavioral Neuroscience, 2007. **7**(4): p. 356-366.

38. Lane, R.D., et al., *Neural activation during selective attention to subjective emotional responses.* Neuroreport, 1997. **8**(18): p. 3969-3972.

39. Lane, R.D., et al., *Neuroanatomical correlates of happiness, sadness, and disgust.* American Journal of Psychiatry, 1997. **154**(7): p. 926-933.

40. Lane, R.D., et al., *Neuroanatomical correlates of pleasant and unpleasant emotion.* Neuropsychologia, 1997. **35**(11): p. 1437-1444.

41. Kensinger, E.A. and D.L. Schacter, *Processing emotional pictures and words: Effects of valence and arousal.* Cognitive Affective & Behavioral Neuroscience, 2006. **6**(2): p. 110-126.

42. Dolcos, F., K.S. LaBar, and R. Cabeza, *Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study.* Neuroimage, 2004. **23**(1): p. 64-74.

43. Phillips, M.L., et al., *Differential neural responses to overt and covert presentations of facial expressions of fear and disgust.* Neuroimage, 2004. **21**(4): p. 1484-1496.

44. Mataix-Cols, D., et al., *Individual differences in disgust sensitivity modulate neural responses to aversive/disgusting stimuli.* European Journal of Neuroscience, 2008. **27**(11): p. 3050-3058.

45. Singer, T., et al., *Empathy for pain involves the affective but not sensory components of pain.* Science, 2004. **303**(5661): p. 1157-1162.

46. Pourtois, G., et al., *Neural systems for orienting attention to the location of threat signals: An event-related fMRI study.* Neuroimage, 2006. **31**(2): p. 920-933.

47. Vuilleumier, P. and G. Pourtois, *Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging.* Neuropsychologia, 2007. **45**(1): p. 174-194.

48. Gilbert, S.J., et al., *Distinct regions of medial rostral prefrontal cortex supporting social and nonsocial functions.* Social Cognitive and Affective Neuroscience, 2007. **2**(3): p. 217-226.

49. Lane, R.D., *Theory of emotional awareness and brain processing of emotion.* Journal of Psychosomatic Research, 2005. **58**(6): p. S9-S9.

50. Ochsner, K.N., et al., *Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other.* Journal of Cognitive Neuroscience, 2004. **16**(10): p. 1746-1772.

51. Johnson, M., et al., *Dissociating medial frontal and posterior cingulate activity during self-reflection*
Soc Cogn Affect Neurosci, 2006. **1**(1): p. 56-64.

52. Gobbini, M.I., et al., *Two takes on the social brain: A comparison of theory of mind tasks.* Journal of Cognitive Neuroscience, 2007. **19**: p. 1803-1814.

53. Rushworth, M.F.S., et al., *Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour.* Trends in Cognitive Sciences, 2007. **11**(4): p. 168-176.

54. Schwarzlose, R.F., C.I. Baker, and N. Kanwisher, *Separate face and body selectivity on the fusiform gyrus.* Journal of Neuroscience, 2005. **25**(47): p. 11055-11059.

55. Peelen, M.V., et al., *Emotional modulation of body-selective visual areas.* Social Cognitive and Affective Neuroscience, 2007. **2**(4): p. 274-283.

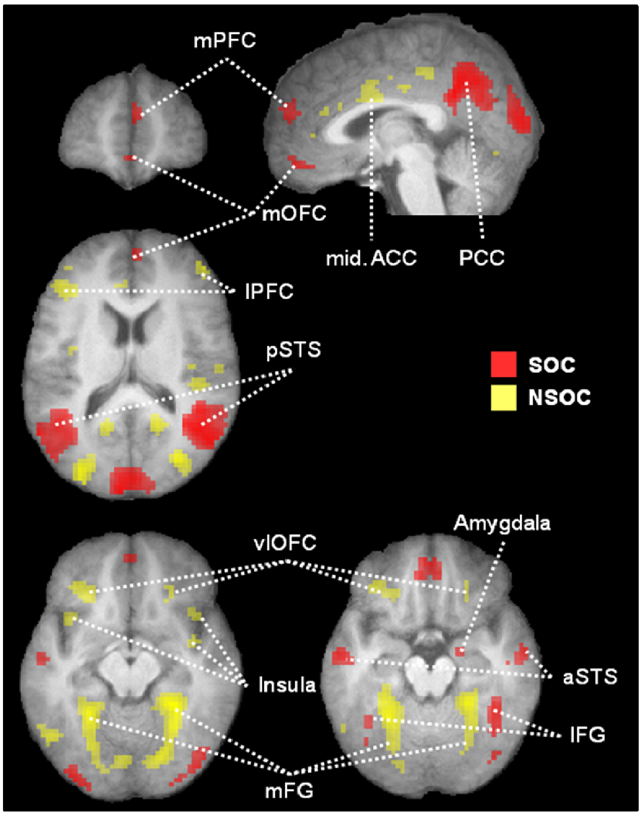
56. Gobbini, M.I. and J.V. Haxby, *Neural systems for recognition of familiar faces.* Neuropsychologia, 2007. **45**(1): p. 32-41.

57. Vuilleumier, P., *How brains beware: neural mechanisms of emotional attention.* Trends in Cognitive Sciences, 2005. **9**(12): p. 585-594.

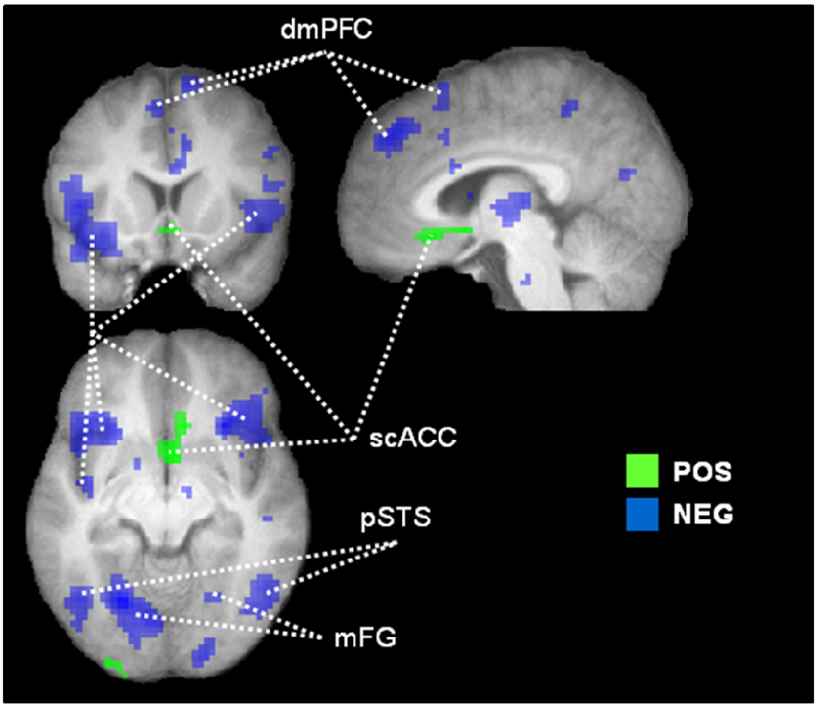
58. Sergerie, K., M. Lepage, and J.L. Armony, *A process-specific functional dissociation of the amygdala in emotional memory.* Journal of Cognitive Neuroscience, 2006. **18**(8): p. 1359-1367.

59. Sander, D., J. Grafman, and T. Zalla, *The human amygdala: an evolved system for relevance detection.* Reviews in the Neurosciences, 2003. **14**(4): p. 303-316.

SUPPLEMENTAL FIGURES



Supplemental Figure S1: Main Effects of Scene Content. *The Statistical parametric maps (threshold p= .001) are illustrated for the contrasts SOC > NSOC (red) and NSOC > SOC (yellow), pooled across all viewing conditions. Abbreviations: mPFC = medial prefrontal cortex, mOFC = medial orbito-frontal cortex, vlOFC = ventro-lateral OFC, lat. PFC = lateral PFC, mid. ACC = middle anterior cingulate cortex, PCC = posterior cingulate cortex, pSTS = posterior superior temporal sulcus, aSTS = anterior STS, FG = fusiform gyrus, m= medial, l= lateral.*



Supplemental Figure S2: Main Effects of Valence. *The statistical parametric maps (threshold p= .001) are illustrated for the contrasts POS > NEG (green) and NEG > POS (blue), pooled across all viewing conditions. Abbreviations: dmPFC = dorso-medial prefrontal cortex, pSTS = posterior superior temporal sulcus, scACC = subcallosal ACC, mFG = medial fusiform gyrus.*

VI.5. fMRI Study N°2b

In Preparation for Publication

fMRI study N° 2b relied on the same experimental paradigm already described in Chapter VI.4., but added the investigation of individual differences, as measured by adult attachment style (AAS), on the neural correlates of emotion regulation (ER) – again, for both cognitive (re-appraisal; REAP) as well as behavioral (suppression; SUP) ER. Accordingly, the main point of interest was:

- **How does AAS modulate the differential regulation (re-appraisal *versus* suppression) of social positive and negative emotions?**

Up to date, there has been only one fMRI study [194] that investigated the influence of AAS on ER, and specifically the suppression of negative relationship-related thoughts. Even though the latter found some hints on how avoidance or anxiety could lead to ER deficits, many questions still remained unresolved.

The new results revealed that attachment avoidance (AV) was correlated with increased prefrontal cortical activity during natural viewing ("watch", see e.g. [87]) of social negative images, in line with previous findings implicating increased attempts to regulate or suppress emotions (PFC), evaluate possible outcomes (mOFC), as well as process and resolve cognitive and emotional conflict (ACC). In addition, sustained amygdala activation during REAP (again when seeing social negative scenes) most likely represented the inability of avoidantly attached participants to successfully down-regulate social negative emotions with the help of a cognitive ER strategy (REAP). Finally, the new data also showed effects of AV on the processing and regulation of social *positive* emotions, particularly by means of increased activity in the reward circuitry and supplemental motor area (SMA) during SUP as well as EMEXP, and thus most likely the inhibition of emotion expression through behavioral ER strategies.

Attachment anxiety (AX) was only found associated with the modulation of activity in left parahippocampus to SP during EMEXP. This might be interpreted by stronger efforts to recall situations/places where SP emotions comparable to the ones shown during the experiment have been experienced before.

The Influence of Adult Attachment Style on the Experience and Regulation of Social Emotions: a fMRI Study

Pascal Vrtička (1,2), David Sander (1,3), & Patrik Vuilleumier (1,2)

(1) Swiss Center for Affective Sciences, University of Geneva
(2) Laboratory for Neurology & Imaging of Cognition, Department of Neurology and Department of Neuroscience, University Hospital and Medical School, University of Geneva
(3) Department of Psychology, FPSE, University of Geneva

ABSTRACT

Adult attachment theory puts forward an elaborated framework on how the regulation of support-seeking behavior is modulated by either avoidance (de-activation of the attachment system) or anxiety (hyper-vigilance to attachment-related cues). Translated to the widely used model of emotion regulation, previous behavioral data suggests that avoidant attachment style may be associated with suppression to prevent experience and expression of emotion, whereas anxious attachment style may be linked with the up-regulation of particularly negative affect by means of re-appraisal and a deficit in negative emotion suppression. Yet, despite such clear predictions, no fMRI study has investigated the modulation of neural responses to social emotions during re-appraisal and suppression as a function of attachment style.

This new fMRI experiment for the first time shows that avoidantly attached participants show increased pre-frontal and anterior cingulate activation to social negative scenes when not using any emotion regulation attempts; persistent high anterior cingulate and amygdala activation for the same stimuli during re-appraisal; and heightened signal in supplemental motor area and ventral striatum during suppression of emotions induced by social positive images. These findings imply that avoidantly attached people experience higher emotional as well as cognitive conflict if they cannot use suppression to regulate social negative emotions; that re-appraisal is not effective in reducing the affective response to the latter; and that the suppression of social positive emotions is linked with decreased pre-motor and increased reward-related activity. Almost no relations were found for anxious attachment style on the regulation of social emotions, but further research is needed to clarify the absence of such effects.

CORRESPONDING AUTHOR:

Pascal Vrtička
Swiss National Center for Affective Sciences
7, rue des Battoirs, 1205 Geneva, Switzerland
Tel: +41 – (0)22 379 9824 Fax: +41 – (0)22 379 9844
Email: pascal.vrticka@unige.ch

INTRODUCTION

From the moment of birth, the main function of the attachment system is to maintain proximity to significant others in times of stress, or in other words, to regulate support-seeking behavior [1, 2]. Because such regulatory efforts of attachment processes are tightly linked with general emotion regulation (ER) mechanisms, some have even argued that “the attachment system is, in itself, an emotion regulation device” [3]. Accordingly, adult attachment theory has come up with an elaborated theoretical framework on how individual differences in adult attachment style (AAS) are mirrored in the use of at least two distinct ER strategies.

Ideally, in the case of secure attachment, ER efforts can be directed to the emotion-generation process by cognitively re-evaluating the meaning and impact of a potentially threatening event. Such beneficial ER has been called “antecedent-focused ER” or re-appraisal (REAP) [4], and is characterized by the successful management and resolution of social negative emotions, which are viewed as goal-incongruent and undesirable by securely attached individuals [3]. Even though avoidantly attached people also tend to view social negative emotions as goal-incongruent and undesirable, their main concern is to keep the attachment system deactivated, which cannot be achieved by REAP. Instead, they refer to so called behavioral inhibition of emotional expression, or in short, suppression (SUP; also called “response-focused ER” [4]), mainly aimed at the prevention of any conscious experience and expression of social emotions. Such ER mechanisms are also referred as to deactivating strategies in relation with AAS [3]. Finally, because anxiously attached individuals often perceive social negative emotions as goal-congruent and thus worth sustaining or even exaggerating, they tend to intensify them through so called hyperactivating strategies. Such ER attempts can be conceptualized as *up*-regulation of emotion (through REAP) – in contrast to *down*-regulation of affect as seen above for secure attachment (through REAP) and avoidant attachment (by SUP) –, involving increased vigilance to signs of rejection or support and their related emotional states, and eased access to memories of and rumination about actual and potential threats [3]. In addition, it has been found that attachment-anxious individuals have difficulties with suppressing negative thoughts and emotions in general, even when asked to do so [5].

As clearly apparent from the last paragraph, most associations between ER strategy use and AAS are conceptualized on the basis of negative emotions. This comes from the fundamental notion in adult attachment theory (AAT) that the attachment system is primarily activated by the presence of a potential threat, and thus particularly by negative attachment-related scenarios. However, one central aspect of attachment theory is the subjective nature of attachment system activation, depending on the subjective appraisal and not only on the actual occurrence of threats. Along these lines, “every event *perceived* by a person … as threatening tends to activate the attachment system” [3]. Therefore, activating triggers would not necessarily have to be equivalent with negative, but could also represent positive emotions. This primarily applies to avoidant attachment (AV), because the main goal of avoidant individuals is to maintain the attachment system in a generally deactivated state. Avoidant persons might thus experience social positive emotions as equally disturbing as their negative counterparts, because both of them are potential – undesired – attachment system activators.

Despite such clear theoretical propositions on the link between attachment processes and ER, very little is known about its neural correlates. To our knowledge, there is only one functional magnetic resonance imaging (fMRI) study up to date that has directly investigated brain activity during ER as a function of AAS, looking at attachment-style differences in the ability to suppress negative thoughts [6]. The latter has found that anxious attachment (AX) was associated with increased reactivity to thoughts of negative relationship scenarios in anterior temporal pole (implicated previously in sadness [7]), and decreased activity in orbitofrontal cortex (normally activated during down-regulation and suppression of emotions [7-10]. In addition, the same study revealed that AV was related with less activity in subcallosal cingulate and lateral prefrontal cortex during SUP of negative thoughts – two brain areas also known to play an important role in ER [6, 7] –, implying that avoidant people’s ability to suppress was less complete or efficient [6].

Even though this first study [11] revealed interesting effects of individual differences in AAS on brain activation during ER, it still leaves several questions unresolved.

Firstly, as the experimental design only included SUP as ER strategy, no conclusions can be drawn on potential differences in brain activation between REAP and SUP as a function of AAS. This would be particularly interesting, because we have only recently been able to show in a fMRI study directly contrasting REAP and SUP (independent of the influence of individual differences) that the latter have different neural substrates in frontal cortical regions, and that REAP and SUP differently modulate activity in visual, limbic and cortical areas (Vrtička et al., submitted). Somewhat related findings have also been reported in another recent article, though in relation to a more complex model of BOLD signal change regarding the two different ER strategies [12]. Moreover, as already noted above, theory on AAS would actually predict different patterns of ER strategy use as a function of AAS. On one hand, AV should more likely be associated with behavioral inhibition (SUP), whereas on the other hand, AX should more strongly involve cognitive *up*-regulation (REAP) and at the same time difficulties in SUP of emotions.

Second, and related to the first remark just mentioned above, Gillath et al. (2005) derived their results from comparing activity between different “think” *versus* “don’t think” contrasts, or emotion processing *versus* suppression, correlated with AAS measures. Even though such a methodological approach is valid, it runs with a risk to be too much concerned with activation differences instead of investigating effects of AAS on one experimental condition, i.e. SUP, alone. This is especially problematic when later on showing correlation graphs of extracted beta *contrast* values with AAS measures, because it is not clear whether such relations are only significant for the “think” *versus* “don’t think” *activation difference*, or for one of these two conditions alone. Moreover, no conclusions can be drawn on the direction of the observed effects, i.e. whether “don’t think” is correlated positively or negatively with AV or AX. Such limitations can be partially overcome by post-hoc statistical analysis on extracted beta values of regions whose activity was found to be correlated with AAS measures – as done by Gillath et al. (2005) –, but it would be more advantageous to already initially compute contrasts within experimental conditions.

Finally, the study by Gillath et al. (2005) used negative relationship scenarios exclusively – such as conflict, breakup or death of a partner –, not taking into account any positive emotions. Even though negative attachment-related scenarios are probably the most potent activators of the attachment system, positive social emotions might also trigger its activation (see above). Accordingly, a previous behavioral study investigating the perceptual processing of positive (happy) and negative (angry, sad) facial expressions changing from emotional to neutral showed that avoidant individuals – under normal conditions – saw the offset of both happy and angry faces relatively late. This finding therefore suggests that AV is characterized by increased vigilance to and preoccupation with *both* positive and negative social cues [13]. In addition, AV has been reported to be associated with a more negative view of others [14], making avoidant people more likely to distrust affective signals from others. This predisposition might increase the vigilance to positive social cues in AV, because the latter might be perceived less genuine and thus more negative. We already tested such relations between AV and positive emotion processing in a recent fMRI study, and found diminished activation in reward-related brain areas as a function of AV during situations representing social support [15]. Similarly, in a behavioral study where participants had to rate the pleasantness of positive/negative and social/non-social images, our results revealed a selective significant negative correlation between pleasantness ratings for social positive images and AV (Vrtička et al., unpublished data). Considering all the described evidence on AV and the processing of positive emotions, it would be promising to extend the findings of Gillath et al. (2005) by not only using negative but also positive attachment-related stimuli in an ER task in relation with AAS.

To address the abovementioned issues, we designed a new study where participants were presented with both positive and negative attachment-related (i.e. social) emotional scenes while they underwent within-block event-related fMRI scanning (see Figure 1 and Methods). We also included neutral images as baseline, as well as positive and negative non-social (attachment-unrelated) emotional scenes as control conditions. Image categories of particular interest included social/non-social and positive/negative

(SP, NSP, SN, NSN) combinations; and social *versus* non-social image categories were matched by pleasantness, intensity, and control ratings to avoid any low level stimulus property constraints (see Methods). Subject's task in the scanner was to rate each picture on a pleasantness scale (from very negative to very positive; 4-point) during three different viewing conditions (VCON); either natural experience of the depicted emotions (EMEXP), using cognitive re-evaluation to down-regulate any emotional impact of emotional scenes (REAP), or applying behavioral suppression to not display any emotional reaction on the outside (i.e. emotional facial expression) while naturally processing the shown emotion inside (SUP; see Methods). All participants filled in a French version of the Relationships Scales Questionnaire (RSQ) several days up to several weeks before the fMRI experiment to assess their AAS scores (in particular AV and AX). Other psychological measures were collected at the same time as well (see Methods). fMRI data was analyzed later on by VCON and VAL, always contrasting social (i.e. attachment-related) *versus* non-social (i.e. attachment-unrelated) emotional scenes (i.e. EMEXP, SP>NSP) correlated with AAS (AV and AX). Thereby, special attention was given to the directionality of any AAS effects on brain activity for each stimulus category and VCON separately, thereby maximally increasing data specificity (see Methods).

METHODS

Subjects

We recruited 19 healthy paid volunteers (all right-handed women, mean age 24.82 ± 4.0), who all had a normal or corrected to normal vision, no history of neurological or psychiatric disease, and gave informed written consent according to the local ethical committee regulation. Only women were included because of evidence that emotions are typically more intense and more prone to regulation in women (e.g. [16]), but also to increase comparability with previous emotion regulation studies that also included women only [10] and to avoid any potential sex differences that could have modulated the regulation effects of primary interest.

Experimental material and procedure

STIMULI

A total number of 360 emotional pictures were initially chosen either from the International Affective Pictures System (IAPS) or from the internet. All were in colors, and adjusted to obtain similar size, contrast, and pixel resolution. Half of the pictures displayed scenes with a clear social content, such as two people fighting or a mother interacting with her baby. The other half represented objects or landscapes (non-social), like a dead bird covered in oil or a tropical island scene. All 360 pictures were rated in a separate behavioral study by 54 female students on three continuous rating scales (from 0 to 100), including PLEASANTNESS (PLN, from very negative to very positive), INTENSITY (INT; from low to high arousal), and CONTROL (CON; from absence to full presence). According to the averaged rating results from this sample, 240 pictures were finally chosen for the fMRI study, and sorted by their SOCIAL CONTENT (SC; either social or non-social) and VALENCE (VAL; either positive or negative). This gave rise to four stimulus categories (60 pictures each): Social Positive (SP) or Negative (NSP), and Non-Social Positive (NSP) or Negative (NSN). The final distribution of pictures in these four categories as a function of emotional rating scores (PLN, INT, and CON) showed that NEG images had significantly lower pleasantness scores as compared to POS images [F(1,59)= 3919.25, p< .001], but higher intensity [F(1,59)= 1554.01, p< .001] and lower control [F(1,59)= 452.97, p< .001] scores. However, most importantly, there were no significant differences for all three rating scales between SOC *versus* NSOC images [Fs(1,59)< 1.36, ps> .25], and no significant VAL x SC interactions [Fs(1,59)< 2.65, ps> .11] (as shown by a 2 [VAL] x 2 [SC] repeated-measure analysis of variance [ANOVAs]). Finally, 40 neutral images were chosen from the IAPS database (20 including humans) to be used in a baseline-task (see below), with average valence ratings of 4.97 ± 0.17 (on the same scale of 1 to 10). Emotional images showed no differences in luminance across categories [all Fs< 2.83; ps> .098 in a 2 [VAL] x 2 [SC] repeated-measure ANOVA], and all social images were comparable in terms of the average number of people depicted per image [SOC POS *versus* SOC NEG: Fs(1,59)= .468; p= .497].

EXPERIMENTAL CONDITIONS

Before entering the fMRI scanner, we told all participants that the purpose of the experiment would be to investigate how the brain reacts to different types of images (e.g. real scenes, TV or movie scenes) and to which degree people can voluntarily influence their emotional effects. Accordingly, the experimental layout comprised four different viewing conditions (VCON), in which pictures were presented with different task instructions to induce different emotion regulation strategies.

The first condition served a control baseline (BASE), and was introduced to the participants as "a photographic quality" judgment, where they had to indicate on each trial (by button press, using a 4-point scale – see below) whether the image was of good quality (e.g. well focused, properly lighted, etc.). All images in this condition were neutral, but could display either scenes with humans (i.e. social content), or inanimate settings and landscapes (i.e., non-social content). This condition was later used to provide a baseline for general differences in brain activation to social vs non-social stimuli, irrespective of emotional processing demands and valence. This condition was presented as the first block of the first and the last block of the last scanning run,.

The three other conditions constituted the main experimental design and included emotional images only. To assess brain responses during "natural" viewing [9, 10], participants were asked to watch and experience the depicted emotional scenarios as if they were living them for real (EMOTION EXPERIENCE condition; EMEXP). To assess the effect of cognitive re-evaluation (RE-APPRAISAL condition; REAP), participants were instructed to view the depicted emotional scenes as parts of a movie clip or TV show that displayed fake or artificially set-up situations created to give rise to emotions. The latter strategy ("pretend unreal") has been one of the most

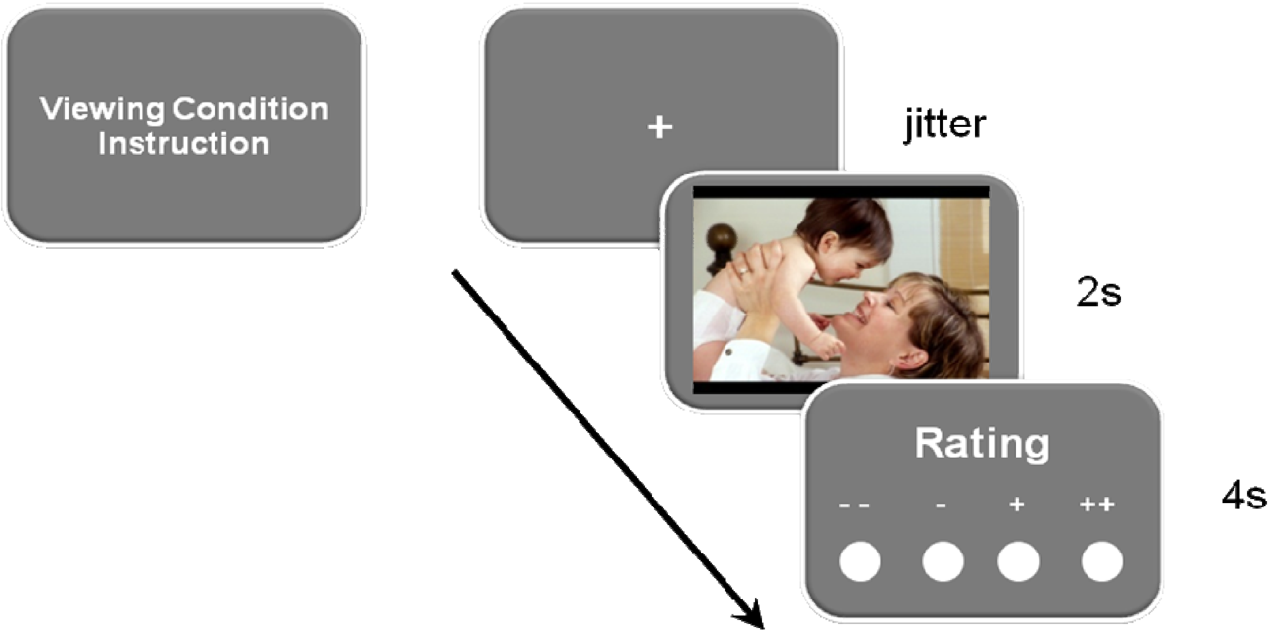


Figure 1: Illustration of the within-block event-related fMRI paradigm. *Each VCON block began with an instruction slide (7 sec), followed by a centered fixation cross (jittered between 790 and 1485 ms, average 1125 ms). The participants were then shown an emotional or neutral image for 2 sec, and subsequently asked to rate the image (see experimental conditions, above) during the display of a rating slide.*

often used in order to down-regulate emotional reactions to negative and positive images [10]. Finally, to assess the effect of behavioral inhibition (SUPPRESSION condition; SUP) the participants were instructed to watch the pictures similarly to the EMEXP condition, but with the important difference that they were told not to display any felt emotions become visible on the outside (i.e. through breathing frequency, heart rate, and skin conductance responses – which were actually monitored during fMRI scanning; or through facial expression – which was pretended to be recorded via an eyetracker camera). After each picture, participants were shown a rating display and asked to report the feeling state evoked by the preceding stimulus (“How did you feel while seeing the last image”?), using a 4-point scale (see below). Importantly, all emotional images were counterbalanced across participants, so that the same images seen in one viewing conditions by a given subject were seen in the other viewing conditions by different subjects.

PROCEDURE

The fMRI experiment was divided into three successive scanning runs. Each run included two of the three experimental viewing conditions, presented in blocks of 40 emotional images (duration = 294 sec per block), whereas the first and the last run also included an additional block of 20 neutral images (baseline condition, duration = 151 sec). Within each block, images were pseudo-randomized and equally probable for the different stimulus categories (social *vs* non-social content, positive *vs* negative valence). The first and the third runs lasted approximately 13 min, and the middle session 10 minutes.

Each VCON block began with an instruction display (7 sec), followed by psudo-randomized images. Every individual trial started with a fixation cross at the screen center (average duration= 1125 ms jittered between 790 and 1485 ms), followed by an emotional or neutral image for 2 sec, and then a response display to rate the image (4 sec; see Figure 1). Ratings were made by using a 4-button response box according to a 4-point scale ranging from very and slightly negative (buttons 1 and 2, respectively), to slightly and very positive (buttons 3 and 4, respectively).

MRI ACQUISITION

MRI data were acquired on a 3 T whole-body INTERA system (Philips Medical Systems), using standard head-coil configuration. For each participant, a structural image was obtained with a MPRAGE T1-weighted sequence (TR/TE/Flip = 2200 ms/30ms/85°, parallel acquisition (GRAPPA) with acceleration factor 2, FOV = 235mm x 235mm, matrix = 128x84, resulting voxel size is 2.8 x 1.8 x 3.4 mm^3). Functional images (TI/TR/TE/flip = 900/1900/2.32/9°, parallel acquisition (GRAPPA) with acceleration factor 2, FOV = 230 x 230 x 173 mm^3, Matrix = 256 x 246 x 192) covered the whole brain, composed of 36 contiguous 4mm axial slices parallel to the inferior edge of the occipital and temporal lobes, and acquired continuously for a total of 975 images per participant (two sessions with 350 and one session with 275 images).

Statistical analysis was performed using the general linear model implemented in SPM2, with a separate regressor for each event type convolved with a canonical hemodynamic response function. Twelve events from the emotion regulation task (SP, NSP, SN and NSN image categories; for each of the three viewing conditions, EMEXP, REAP and SUP) and two events from the artistic quality judgment (social and non-social) were modeled. Movement parameters from realignment corrections were entered as additional covariates of no interest to account for residual movement artifacts after realignment. Statistical parametric maps were generated from linear contrasts

SOC > NSOC

Region	*BA*	*Voxel*	*T-Value*	*p-Value*	*xyz*	*Correlation with AV*
Amygdala left (†)		13	3.55	0.001	-21 -9 -18	REAP_SN; r= .564; p= .012 *
Amygdala right (†)		26	4.21	< 0.001	21 -6 -18	REAP_SN; r= .476, p= .040 *
mOFC	11	40	5.26	< 0.001	-3 54 -18	EMEXP_SP; r= .478, p= .038 *
mPFC	10	28	4.36	< 0.001	3 57 15	
PCC / PC	23 / 7	393	7.32	< 0.001	0 -51 33	
pSTS / Occipital left	19	458	7.82	< 0.001	-45 -84 0	
pSTS / Occipital right	19	711	7.07	< 0.001	45 -48 18	
FG left	18	18	6.5	< 0.001	-15 -60 -6	
FG left	20	10	4.4	< 0.001	-42 -33 -24	
FG left	19 / 37	25	4.43	< 0.001	-42 -63 -21	
FG right	37	98	6.59	< 0.001	42 -42 -27	

TABLE 1: ROIs of the Main Effects Analysis (SOC > NSOC). *Peak coordinates are given in MNI space and listed with best estimates of anatomical location. BA = Brodmann's area, AV = avoidant attachment style, mOFC = medial orbitofrontal cortex, mPFC = medial prefrontal cortex, PCC = posterior cingulate cortex, PC = precuneus, FG = fusiform gyrus, pSTS = posterior superior temporal sulcus. (†) = p< .005; * = p< .05.*

A first second-stage random-effect analysis was then performed using one-sample t-tests on contrast images obtained in each subject for each comparison of interest. All contrasts were performed across the whole brain using standard threshold criteria (Worsley et al., 1996) of significant activation at a voxel-level of p< .001 uncorrected (except for bilateral amygdala, p< .005) and cluster size greater than 5 voxels (135 mm^3). Average parameter estimates of activity (betas) for each feedback condition were extracted from all voxels in regions of interest (ROIs), defined by the full-extent clusters showing significant activation at a voxel level of p< .001 (uncorrected, T-value > 3.61; except for bilateral amygdala: p< .005, T-value > 2.88) in the SPM group analysis (random-effect contrasts).

Statistical correlations with AAS and other personality traits were performed in two stages. Firstly, we tested for a relation between the extracted beta values (whole cluster averages) from activated ROIs (contrast SOC *versus* NSOC) with standardized questionnaire scores (AAS from RSQ, ERQ, RQ, and STAI-T) using two-tailed Pearson product moment coefficients in SPSS 17.0 (SPSS, Chicago, Illinois, United States). Secondly, we performed whole-brain multiple regression analysis on the contrast image of interest using AV and AX scores as linear parametric factors in SPM2, allowing us to test for any voxels throughout the brain where activation in the given contrast varied as a function of behavioral measures [17]. To further insure data specificity regarding only one AAS dimension, we applied exclusive masking procedures at a statistical threshold of p= .01 for the other AAS dimension (i.e. contrast EMEXP SP *versus* NSP correlated positively with AX, masked exclusively with the contrast EMEXP SP *versus* NSP correlated negatively with AV at p= .01). For this second-stage correlation analysis, significant effects were identified using a statistical threshold of p< .001 at the voxel-level (uncorrected) and cluster size equal or greater than 5 or 10 voxels (see Results). Subsequently, we calculated two-tailed Pearson product moment coefficients of extracted beta values (whole cluster averages) from activated ROIs for each stimulus category separately to assess whether the found correlation with AAS measures was only present for the difference (i.e. SP *versus* NSP), or for one and/or two of the stimulus conditions exclusively, in addition to its direction. Finally, we derived two-tailed Pearson product moment coefficients between extracted beta values from activated ROIs with ERQ, RQ, and STAI-T measures, applying the same procedure as for AAS measures.

Psychological Questionnaires

ADULT ATTACHMENT STYLE

Initially, individual differences in adult attachment orientations were described as either secure, anxious, or avoidant attachment styles (AAS) [18]. Such a categorical adult attachment organization was pursued by subsequent experimental investigations, leading to models proposing four or even five different AAS categories [14, 19, 20]). However, more recently, AAS has been conceptualized with the help of two continuous orthogonal axes of avoidance (AV) and anxiety (AX) making up a bi-dimensional space, with the secure attachment style (SA) being situated where both anxiety and avoidance are low [14, 21].

To measure individual AAS of our subjects, we used a French version of the original Relationships Scales Questionnaire (RSQ; [22], translated and validated in a thesis work by N. Guédeney, 2005). According to a recent review paper [23], the RSQ was analyzed with the help of the most appropriate model [24] by calculating scores for avoidance (AV) and anxiety (AX) for each subject, taking 13 items (8 for AV and 5 for AX) out of the total 30 items into account. Values were then centered (z-scores) to reduce effects of multicollinearity [25].

GENERAL USE OF EMOTION REGULATION

To obtain a subjective measure of the preferred emotion regulation (ER) strategy generally used by our subjects, we made them fill in a French version of the Emotion Regulation Questionnaire (ERQ; [26]; provided at http://www-psych.stanford.edu/~psyphy/pdfs/erq10-french.pdf), including 6 items for re-appraisal (REAP), and 4 items for suppression (SUP).

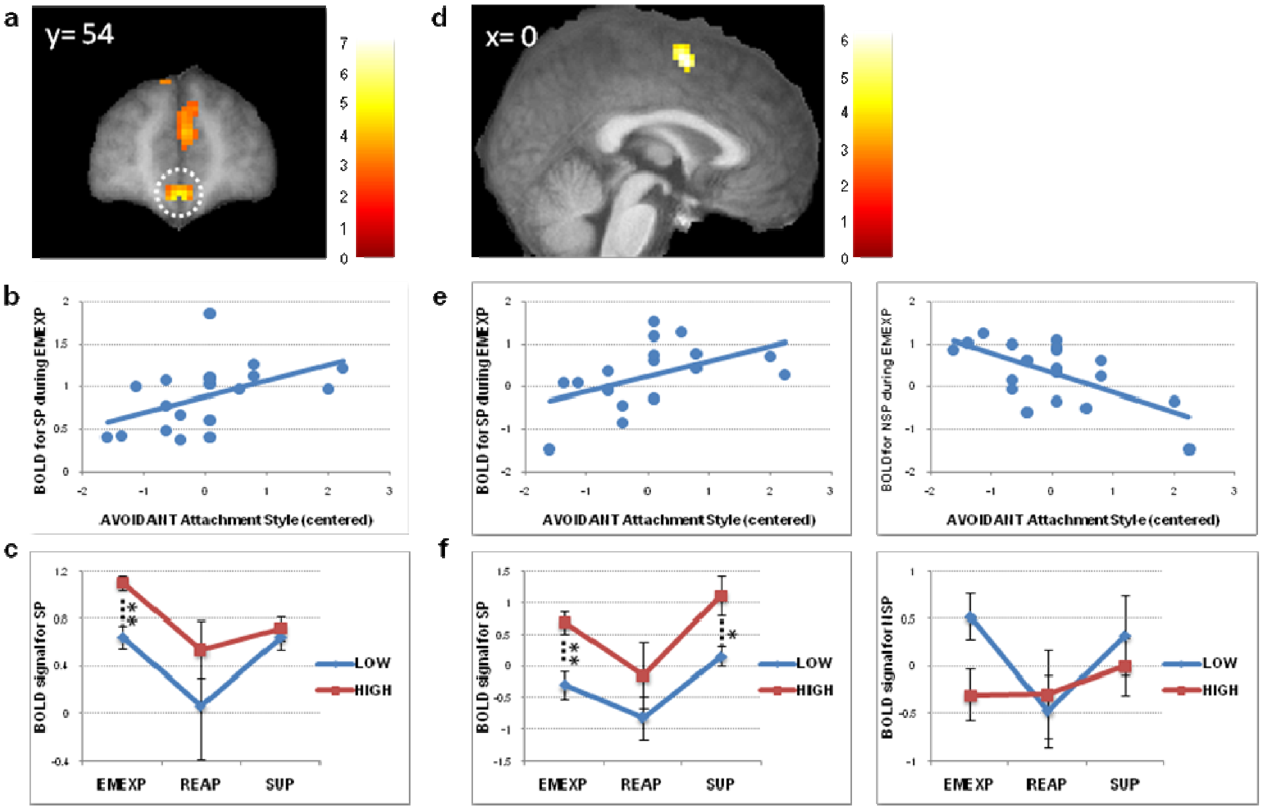


Figure 2: Brain Activations to Social Positive (SP) Emotions during Emotion Experience (EMEXP) in relation with Avoidant AttachmentStyle (AV) *a) Statistical parametric map for the initial contrast SOC versus NSOC, showing increased activity in medial orbitofrontal cortex (mOFC; xyz= -3 54 -18). b) Significant correlation between AV and parameter estimates (beta values) extracted from the mOFC cluster during EMEXP for SP images, averaged across voxels and participants (see Table 1). c) Median split analysis for high (red) versus low (blue) avoidance in mOFC, showing a significant difference between activity to SP emotions during EMEXP only. d) Statistical parametric map for the whole brain multiple regression analysis using the contrast EMEXP SP versus NSP positively correlated with AV, depicting increased activity in an area overlapping with the supplementary motor area (SMA; xyz= 0 0 57). e) Significant correlations between AV and parameter estimates (beta values) extracted from the SMA cluster during EMEXP for SP (left) and NSP (right) images, averaged across voxels and participants (see Table 1). f) Median split analysis for high (red) versus low (blue) avoidance in SMA, showing a significant difference in activity to SP emotions during EMEXP and SUP (left), and a marginally significant difference in activity to NSP emotions during EMEXP (right). All median split analysis graphs are depicted with ± 1 S.E M. * = p< .01, ** = p< .001.*

SELF AND OTHER MODELS

Together with the RSQ, we also administered a French version of the Relationships Questionnaire (RQ, [14]; translated and validated in a thesis work by N. Guédeney, 2005), which can be used to calculate estimates of the positivity and/or negativity of self and other perception.

GENERAL ANXIETY

To differentiate between AX in relation to AAS and general anxiety, we also distributed a French version of the State-Trait Anxiety Inventory (STAI-T; [27, 28]), including 20 items to assess how subjects feel generally regarding the trait anxiety construct.

RESULTS

Questionnaire Data

For the fMRI study at hand, we made use of four personality questionnaires, assessing 1) adult attachment style (AAS; Relationships Scales Questionnaire [RSQ]), 2) self and other models related to attachment style (Relationships Questionnaire [RQ]), 3) general employment of re-appraisal and suppression (Emotion Regulation Questionnaire [ERQ]), and 4) general anxiety (State Trade Anxiety Inventory [STAI-T]). For details and references regarding these questionnaires, please refer to Methods.

Avoidant (AV) and anxious (AX) attachment style measures obtained by the RSQ were not correlated [Pearson-r= .283, p= .24]. For AV, we only found a marginally significant positive relation between employment of suppression [Pearson-r= .404, p=. 096], implying that the higher subjects scored on AV, the more they indicated to use suppression as general emotion regulation (ER) strategy. AX scores showed a marginally negative relation with re-appraisal use [Pearson-r= -.433, p= .073], a marginally positive correlation with general anxiety [Pearson-r= .411, p= .09], and a significant negative correlation with the RQ-self scale [p= .004]. These findings suggest that anxiously attached participants less often use re-appraisal as general ER strategy (to *down*-regulate emotions), had a more negative self model, and were generally more anxious. Finally, we also found a marginally significant negative relation between general anxiety scores and general re-appraisal use [Pearson-r= -.424, p= .08], and a significant negative correlation between general anxiety scores and the RQ self model [Pearson-r= -.770, p< .001].

Behavioural Data

During fMRI scanning, participants were asked to rate neutral and emotional images on a four-point pleasantness scale from very negative to very positive (see Methods).

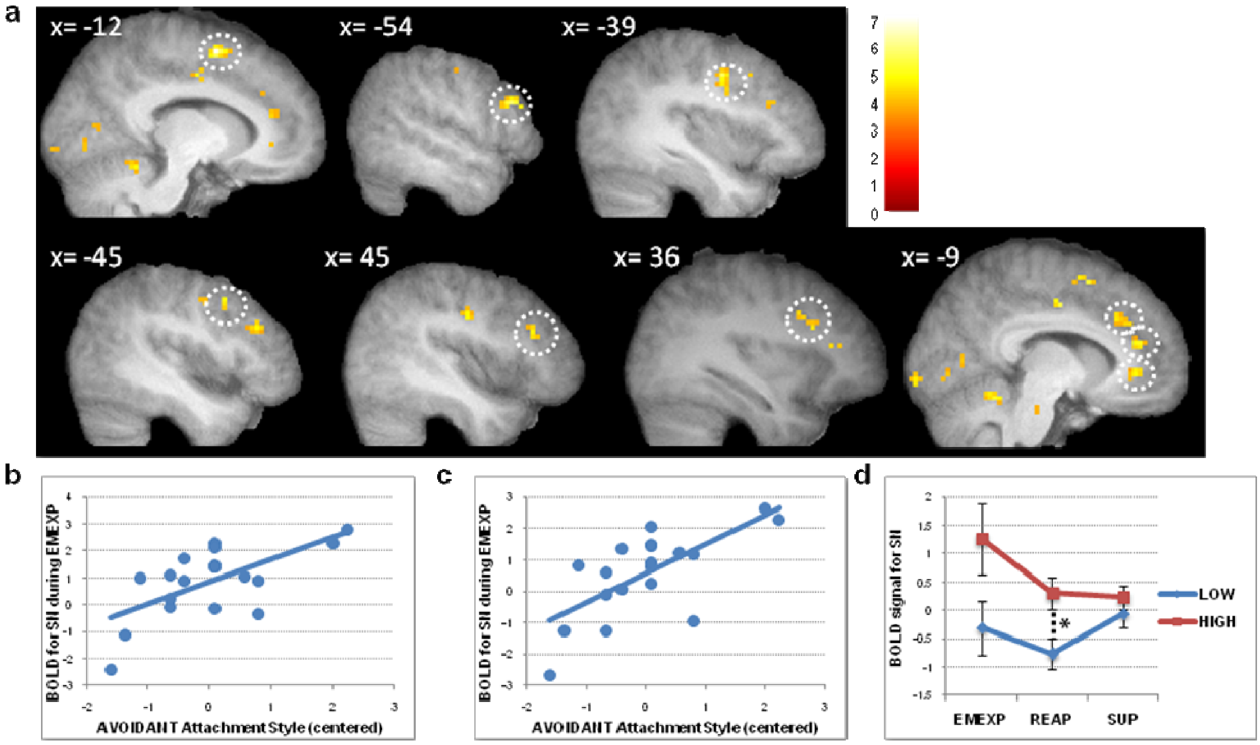


Figure 3: Brain Activations to Social Negative (SN) Emotions during Emotion Experience (EMEXP) in relation with Avoidant Attachment Style (AV). *a) Statistical parametric map for the whole brain multiple regression analysis using the contrast EMEXP SN versus NSN positively correlated with AV, showing increased activity in several areas of the bilateral lateral and dorso-lateral prefrontal cortex, as well as ventral and dorsal anterior cingulate cortex (see Table 2 for more details). b) Significant correlation between parameter estimates (beta values) extracted from right dorso-lateral prefrontal cortex (xyz= 36 21 27) and AV to SN images during EMEXP, averaged across voxels and participants (see Table 2 for more details). c) Significant correlation between parameter estimates (beta values) extracted from right dorsal anterior cingulate cortex (xyz= 3 27 27) and AV to SN images during EMEXP, averaged across voxels and participants (see Table 2 for more details). d) Median split analysis for high (red) versus low (blue) avoidance in dACC (xyz= 3 27 27), showing a significant difference in activity to SN emotions during REAP, and a marginally significant difference during EMEXP (depicted with ± 1 S.E M.). * = p< .01.*

Rating scores indicate that subjects rated positive images as more pleasant and negative images as less pleasant than neutral images, during all three viewing conditions and for all four stimulus categories [ts(18,2)> 6.24, ps< .001], not displaying any SOC *versus* NSOC effects [SP *versus* NSP: ts(18,2)< 1.84, ps> .083; SN *versus* NSN: ts(18,2)< 1.47, ps> .16]. Regulation success was significantly stronger during REAP than during SUP (as compared to EMEXP) for all four stimulus categories [ts(18,2)> 2.72, ps< .014]. In addition, regulation success was found for all four stimulus categories during REAP [ts(18,2)> 4.35, p< .001], but only for POS stimuli during SUP [POS: ts(18,2)> 2.56, p< .020; NEG: ts(18,2)< 1.45, ps> .17] (all comparisons are paired t-tests).

We did not find any effects of AV on behavioral rating scores [all Pearson-rs< .347; ps> .146], but a positive correlation between AX and rating scores for NSP images during SUP [t(18,2)= .555; p= .014].

fMRI Data

To assess effects of AAS on brain activity during experience and regulation of social emotions, we carried out fMRI data analysis in two steps. First, we correlated averaged extracted beta values from regions of interest (ROIs) functionally defined by the contrast social *versus* non-social with AAS scores. Second, we computed whole brain multiple regression analysis by entering AV and AX scores as additional columns in the SPM contrast matrices and applied exclusive masking procedures to ensure specificity of observed effects related to AV and AX, respectively (see Methods). The second step was carried out by VCON (EMEXP, REAP, and SUP) for positive and negative emotional scenes separately to avoid any valence / arousal confounds, and always by contrasting social *versus* non-social emotional scenes (i.e. EMEXP, SP *versus* NSP). Moreover, correlation coefficients were computed from extracted beta values for each stimulus category separately (i.e. EMEXP SP) using SPSS, and thus not only for contrast differences alone (see Methods). This additional step further increased data specificity. Finally, we split our data at the median into high and low AV/AX and computed independent sample two-tailed t-test analysis to test for any additional differences in activity during all three VCONs.

Avoidant Attachment Style and Emotion Experience

Positive Valence: The computation of correlations between averaged extracted beta values from ROIs defined by the contrast SOC *versus* NSOC revealed only a selective positive relation between AV and activity to SP emotions during EMEXP in medial orbito-frontal cortex (mOFC; see Figure 2ab and Table 1). Moreover, when splitting the extracted beta values at the median between low (n= 8) and high (n= 5) avoidance, there was only a significant difference between activity to SP emotions during EMEXP [t(11,1)= 3.43, p= .006], but not REAP and SUP [ts(11,1)< .90, ps> .39] (two-tailed independent samples t-test; see Figure 2c). The whole brain multiple regression analysis (contrast SP *versus* NSP during EMEXP correlated positively with AV)

SP>NSP AV POS

Region	*BA*	*Voxel*	*T-Value*	*p-Value*	*xyz*	*Correlation with AV*
SMA	6	37	6.15	p< .001	0 0 57	EMEXP_SP; r= .484; p= .036 *
						EMEXP_NSP; r= -.635; p= .003 **

SN>NSN AV POS

Region	*BA*	*Voxel*	*T-Value*	*p-Value*	*xyz*	*Correlation with AV*
dlPFC left		59	5.6	p< .001	-54 24 24	EMEXP_SN; r= .539; p= .017 *
dlPFC left		51	5.48	p< .001	-39 -3 42	EMEXP_SN; r= .630; p= .004 **
dlPFC left		12	4.86	p< .001	-45 9 45	EMEXP_SN; r= .592; p= .008 **
dlPFC right		10	4.48	p< .001	45 30 24	EMEXP_SN; r= .474; p= .040 *
dlPFC right		17	4.43	p< .001	36 21 27	EMEXP_SN; r= .646; p= .003 **
dACC right		26	5.54	p< .001	15 30 30	EMEXP_SN; r= .526; p= .021 *
dACC left		12	4.94	p< .001	-9 42 18	EMEXP_SN; r= .606; p= .006 **
dACC right		34	4.78	p< .001	3 27 27	EMEXP_SN; r= .689; p= .001 **
vACC left		12	4.71	p< .001	-9 39 -3	EMEXP_SN; r= .414; p= .078 (*)

TABLE 2: Whole Brain Correlations with Avoidant Attachment Style during Emotion Experience for Social (versus Non-Social) Emotions
*Peak coordinates are given in MNI space and listed with best estimates of anatomical location. BA = Brodmann's area, AV = avoidant attachment style, SMA = supplemental motor area, dlPFC = dorso-lateral prefrontal cortex, dACC = dorsal anterior cingulate cortex, vACC = ventral ACC. * = p< .01, ** = p< .001*

revealed another area of the brain overlapping with the supplementary motor area (SMA; BA 6; see Figure 2d and Table 2), where average BOLD signal change was related positively with AV for activity to SP, but negatively with AV for activity to NSP images (see Figure 2e and Table 2). In addition, when computing the median split analysis as a function of AV, there was a significant difference between activity to SP emotions during EMEXP and SUP [ts(11,1)> 2.75, ps<.034], but not REAP [t(11,1)= 1.07, p= .32], and a trend towards a difference during EMEXP exclusively regarding NSP emotions [t(11,1)= 1.93, p= .093] (see Figure 2f).

Negative Valence: There were no significant effects for the ROI analysis regarding SN images during EMEXP and AV. When computing the whole brain multiple regression analysis (contrast SN *versus* NSN during EMEXP correlated positively with AV), we discovered selective positive relations between AV and activity to SN images in several brain areas, including bilateral lateral and dorso-lateral pre-frontal cortex (lPFC and dlPFC), and bilateral dorsal and ventral ACC (see Figures 3abc and Table 2). For lPFC and dlPFC activity associated with AV, the median split analysis did not reveal any additional significant differences regarding SN emotions during all VCON, except for one dlPFC cluster (xyz= -54 24 24), where we found a significant difference between low (n= 8) and high (n= 5) avoidance during REAP [t(11,1)= 3.11, p= .018] (data not shown). These results thus indicate that lPFC and dlPFC was almost exclusively more strongly activated to SN emotions as a function of AV during EMEXP, as revealed by the significant correlation data (see above). In two out of the three dACC clusters, the median split analysis revealed an additional significant difference between activity to SN emotions during REAP [*xyz= 15 30 30*: t(11,1)= 2.86, p= .015; *xyz= 3 27 27:* t(11,1)= 2.67, p= .023] (see Figure 3d). No such effects were present in vACC.

<u>Avoidant Attachment Style and Re-Appraisal</u>

Positive Valence: There were no significant effects, neither for the ROI, nor for the whole brain correlation analysis regarding SP images during REAP.

Negative Valence: When computing correlations between extracted beta values from ROIs, we found positive relations between activity to SN images and AV in bilateral amygdala (see Figures 4ab and Table 1). When splitting the beta values at the median into high (n=8) and low (n=5) avoidant subjects, we found that bilaterally, activity to SN emotions during REAP was significantly (in left amygdala marginally) higher for high as compared to low avoidant subjects [*left:* t(11,2)= 6.34, p< .001; right: t(18,1)= 1.94, p= .079], but not during EMEXP or SUP [*left:* ts(18,1)< 1.94, ps> .079; *right:* t(18,2)< .69, p> .51] (see Figure 4c).
The whole brain multiple regression analysis (contrast SN *versus* SNS during REAP correlated positively with AV) revealed positive relations between AV and activity to SN, but negative relations between AV and activity to NSN in two clusters in the right putamen (see Figures 4de and Table 3). The median split analysis of beta values in one right putamen cluster (xyz= 30 9 -3) as a function of AV revealed that activity to SN emotions was lower during REAP for low as compared to high avoidant subjects (even though not significant, data not shown), and that activity to NSN emotions was lower for high as compared to low avoidant subjects exclusively during REAP [t(11,2)= 2.25, p= .051] (see Figure 4f).

<u>Avoidant Attachment Style and Suppression</u>

Positive Valence: No significant effects were found for the ROI analysis.
The whole brain multiple regression analysis (contrast SP *versus* NSP during SUP correlated positively with AV) revealed positive relations between averaged BOLD signal change to SP images and AV in left caudate and left SMA (see Figures 5abde and Table 4). In addition, in the two clusters overlapping with the right putamen described above (see Figure 4d and Table 3),

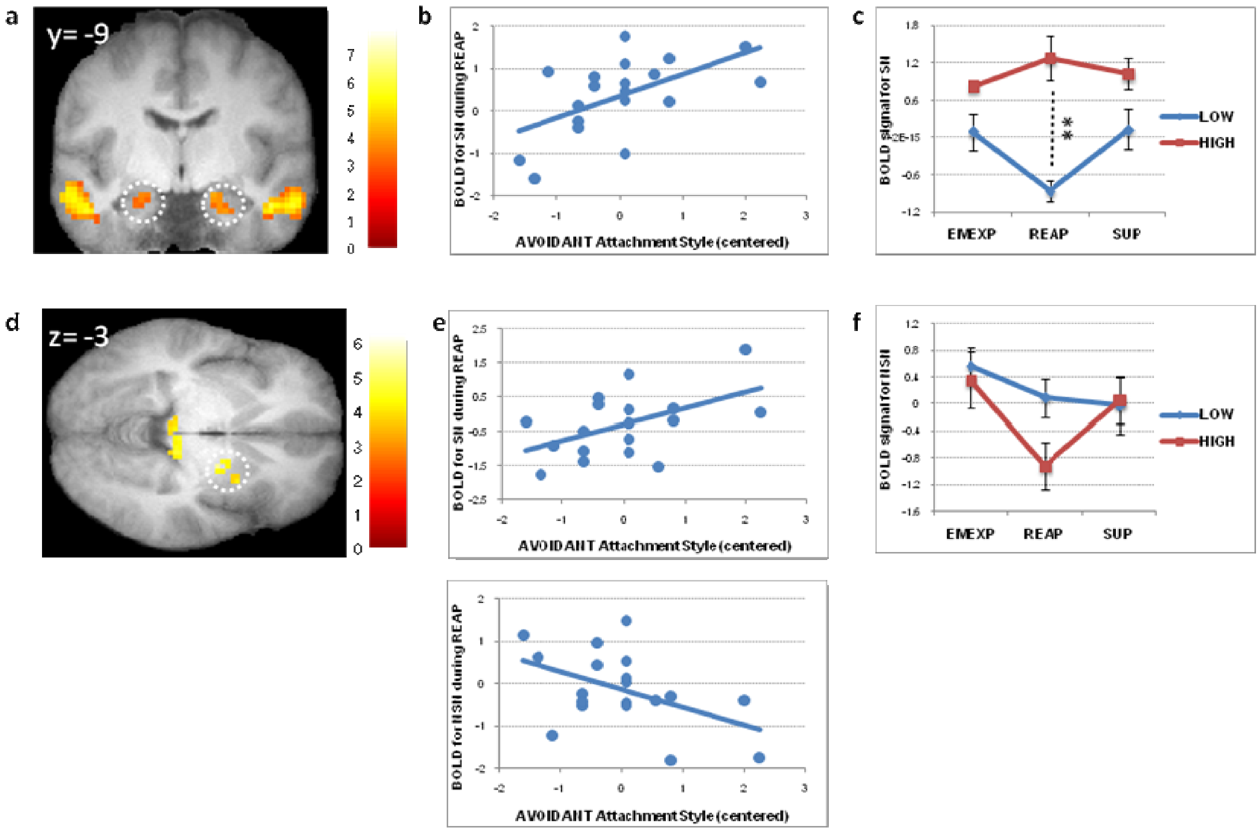


Figure 4: Brain Activations to Social Negative (SN) Emotions during Re-Appraisal (REAP) in relation with Avoidant Attachment Style (AV). *a) Statistical parametric map for the initial contrast SOC versus NSOC, showing increased activity in bilateral amygdala (left: xyz= -21 -9 -18; right: xyz= 21 -6 -18). b) Significant correlation between parameter estimates (beta values) extracted from the left amygdala and AV to SN images during REAP, averaged across voxels and participants (see Table 1). c) Median split analysis for high (red) versus low (blue) avoidance in left amygdala, showing a significant difference between activity to SN emotions during REAP only. d) Statistical parametric map for the whole brain multiple regression analysis using the contrast REAP SN versus NSN positively correlated with AV, depicting increased activity in right putamen (xyz= 21 -3 -6 and xyz= 30 9 -3) e) Significant correlations between parameter estimates (beta values) extracted from the more lateral right putamen cluster (xyz= 30 9 -3) and AV to SN (above) and NSN (below) images during REAP, averaged across voxels and participants (see Table 3 for more details). F) Median split analysis for high (red) versus low (blue) avoidance in right putamen (xyz= 30 9 -3), showing a marginally significant difference in activity to NSN emotions during REAP. All median split analysis graphs are depicted with ± 1 S.E M. ** = p< .001.*

we also found positive relations between AV and activity to SP images during SUP [Pearson-rs> .493, ps< .032]. In the caudate nucleus, the median split analysis revealed an additional significant difference between activity to SP emotions during REAP [t(11,2)= 2.34, p= .041] (see Figure 5c). The same analysis showed a significant difference between activity to SP emotions during SUP [t(11,2)= 2.83, p= .032], and a marginally significant effect during EMEXP [t(11,2)= 2.14, p= .056] in SMA (see Figure 5e). Finally, the median split analysis in relation to AV also revealed that the observed effect in right putamen was selective for SUP, even though not significant (see Figure 5g).

Negative Valence: There were no significant effects, neither for the ROI, nor for the whole brain correlation analysis regarding SN images during SUP.

All effects reported above regarding AV did not show any relations to AX, and were thus specific for the avoidance dimension of AAS. Moreover, there were no relations to general anxiety measures (STAI-T; see below)

There were only very few brain areas displaying increased activity to non-social emotions in correspondence to AV, especially for POS emotions during SUP, including dACC and left operculum, where we found positive relations between AV and activity to NSP images (data not shown).

Anxious Attachment Style

The ROI analysis on the initial SOC *versus* NSOC contrast did not reveal any significant effects of AX on social emotion experience and regulation.

When computing the six whole brain multiple regression analysis (contrasts SP *versus* NSP and SN *versus* NSN correlated positively with AX during EMEXP, REAP, and SUP), we only found effects of AX on brain activity in very few ROIs and only during EMEXP and REAP, but not SUP (see Table 5). Moreover, after extracting beta values, computing correlations with AX, and performing median spit analysis (low [n= 9] and high [n=8] anxiety) for separate stimulus categories, only activity in left parahippocampus (see Figure 6a) during EMEXP was found to be related positively with activity to social emotions, namely SP (see Figure 6b and Table 5), independent of general anxiety (STAI-T; see below). Also, only the difference for SP activity during EMEXP between low and high anxious participants was significant [t(15,2)= 4.64, p< .001] (see Figure 6c).

SN>NSN AV POS

Region	*BA*	*Voxel*	*T-Value*	*p-Value*	*xyz*	*Correlation with AV*
Putamen right		6	4.21	p< .001	30 9 -3	REAP_SN; r= .508; p= .027 *
						REAP_NSN; r= -.479; p= .038 *
Putamen right		15	5.32	p< .001	21 -3 -6	REAP_SN; r= .436; p= .062 (*)
						REAP_NSN; r= -.408; p= .083 (*)

TABLE 3: Whole Brain Correlation with Avoidant Attachment Style during Re-Appraisal for Social (versus Non-Social) Emotions. *Peak coordinates are given in MNI space and listed with best estimates of anatomical location. BA = Brodmann's area, AV = avoidant attachment style,. * = p< .01, (*) = marginally significant.*

General Anxiety, ERQ and RQ Self and Other

General Anxiety: All effects in ROIs showing relations between activity to social *versus* non-social emotions and AAS described here were independent of general anxiety (STAI-T) measures.

ERQ and RQ Self and Other: Besides the correlations between questionnaire measures reported above (see Questionnaire Data), we did not find any significant effects of general ER strategy employment (ERQ) and Self or Other models (RQ) on fMRI data.

DISCUSSION

This fMRI study was aimed at elucidating the influence of adult attachment style (AAS) on brain activity during experience and regulation of social emotions. By including both cognitive re-evaluation (re-appraisal, REAP) and behavioral inhibition of emotion expression (suppression, SUP) as emotion regulation (ER) strategies, and both positive as well as negative social emotions, we wanted to more thoroughly investigate the neural correlates of individual differences in emotion processing related to AAS.

Our questionnaire data revealed that the two different AAS measures (avoidance [AV] and anxiety [AX]) derived by the relationships scales questionnaire (RSQ) were not correlated with each other, as intended. We also found that participants scoring high on AV indicated to more often use suppression SUP as general ER strategy. This finding is well in line with adult attachment theory (AAT) associating AV with deactivating strategies to keep the attachment system in a low activity state, mainly through SUP [3].

AX was correlated negatively with general use of REAP as ER strategy to *down*-regulate emotions in our subjects. Such an association is not exactly predicted in the same way by AAT, the latter mostly associating AX with intensification of particularly negative emotions through *up*-regulation by REAP [3]. However, one could imply that, because our anxiously attached subjects should normally use REAP to *up*-regulate emotions, they are less likely to apply the same strategy for the opposite intention, namely to *down*-regulate affect. In addition, we found that AX was negatively associated with the Self model derived from the Relationships Questionnaire (RQ). Such a relation is well documented in adult attachment literature, linking hyperactivating strategies with intensification of doubts about self-worth and self-efficacy, and intensification of a person's sense of vulnerability to rejection or abandonment [3]. Finally, our questionnaire data revealed a positive association between AX and general anxiety, as measured by the Spielberger Trait Anxiety Inventory (STAI-T). Moreover, we also found negative correlations between general anxiety, general use of REAP, and the RQ Self model. These data suggest that, in our sample, AX and general anxiety were rather tightly linked. Yet, such correlated AX, general anxiety and sometimes also neuroticism scores in the same population are quite common [6, 29, 30]. It is therefore even more important to carefully separate effects of AX from general anxiety / neuroticism when analyzing and interpreting brain activations in relation with AAS.

The behavioral data showing stronger subjective emotion experience decrease during REAP as compared to SUP replicates previous findings regarding the impact of ER on negative emotion processing [12]. Regarding AAS, we did not find any effects on behavioral ratings in our experiment, except for a positive correlation between AX and rating scores for non-social positive (NSP) images during SUP. This means that, the higher subjects scored on AX, the less efficient they indicated SUP to be in reducing subjective experience of positive affect induced particularly by non-social images. Such a relation between AX and subjective experience of NSP emotions during SUP would not be predicted by AAT, but is mirrored in our fMRI data, mostly revealing modulation of brain activation by AX of non-social emotions (see below).

Attachment Style and Negative Social Emotion

According to adult attachment theory (AAT), AV is normally associated with the employment of deactivating strategies to prevent any conscious experience and expression of especially social negative (SN) emotions. This should mainly be achieved through SUP as emotion regulation (ER) strategy, because REAP still requires the activation of the attachment system to a certain degree [3]. We therefore hypothesized that high avoidant subjects should display increased limbic and frontal cortical activation to SN emotions during emotion experience (EMEXP), being asked explicitly to not use any ER to down-regulate or suppress any negative effect induced during this viewing condition (VCON). At the same time, we also predicted that high avoidant participants should display sustained increased limbic activation during REAP, this particular ER strategy being rather inefficient in reducing the impact of SN emotions.

Our fMRI data revealed that high scores on AV were associated with increased activation to SN emotions during EMEXP in several lateral and dorso-lateral areas of the prefrontal cortex (lPFC and dlPFC) not found for the initial contrast comparing general social (SOC) *versus* non-social (NSOC) emotion processing. Such an activation pattern is of particular interest, because previous studies on ER have repeatedly described

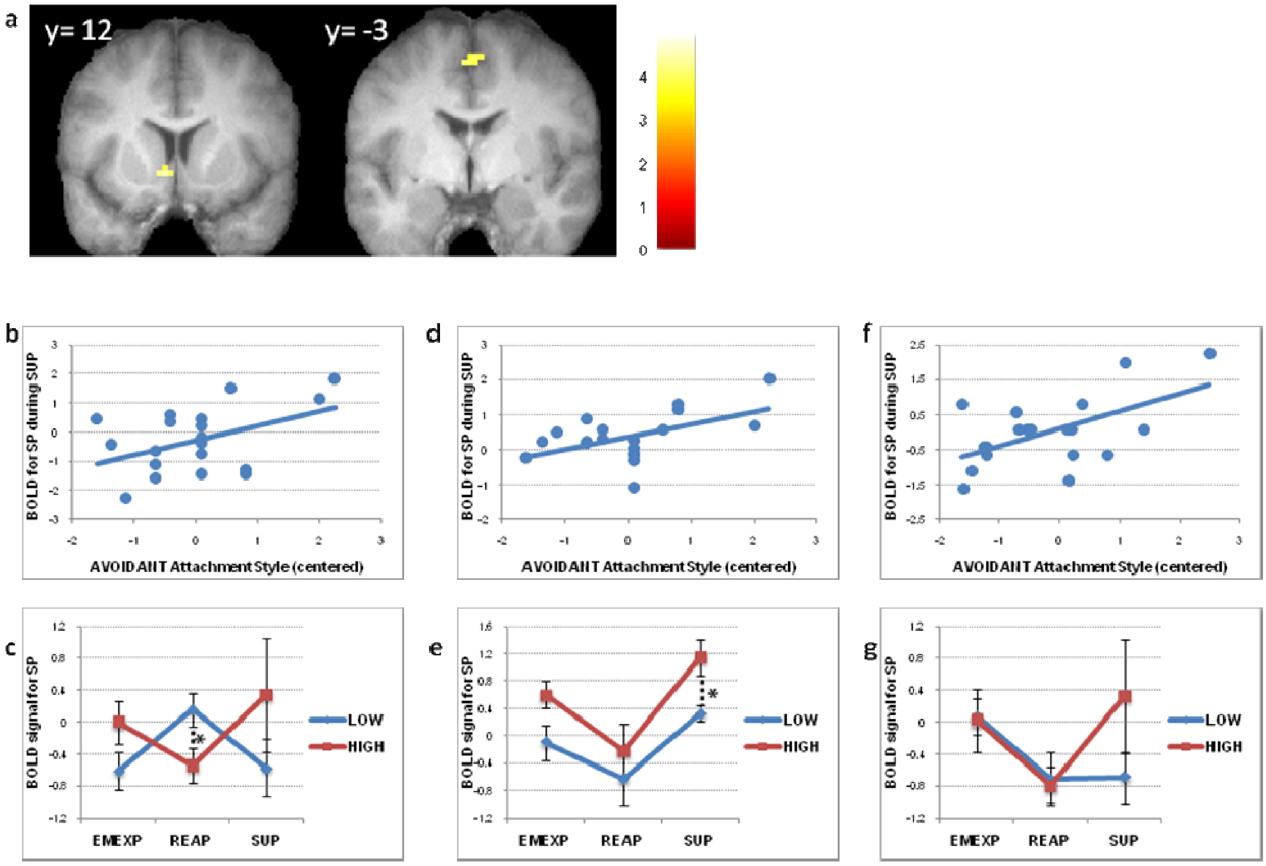


Figure 5: Brain Activations to Social Negative (SN) Emotions during Suppression (SUP) in relation with Avoidant Attachment Style (AV). *a) Statistical parametric map for the whole brain multiple regression analysis using the contrast SUP SP versus NSP positively correlated with AV, showing increased activity in left caudate nucleus (xyz= -6 12 -6) and SMA (xyz= -3 -3 54; see Table 4 for more details). b) Significant correlation between parameter estimates (beta values) extracted from left caudate nucleus and AV to SP images during SUP, averaged across voxels and participants (see Table 4 for more details). c) Median split analysis for high (red) versus low (blue) avoidance in left caudate nucleus, showing a significant difference between activity to SP emotions during REAP. d) Significant correlation between parameter estimates (beta values) extracted from SMA and AV to SP images during SUP, averaged across voxels and participants (see Table 4 for more details). e) Median split analysis for high (red) versus low (blue) avoidance in SMA, showing a significant difference between activity to SP emotions during SUP, and a marginally significant difference during EMEXP. f) Significant correlation between parameter estimates (beta values) extracted from right putamen (xyz= 30 9 -3) and AV to SP images during SUP, averaged across voxels and participants (see text for more details). g) Median split analysis for high (red) versus low (blue) avoidance in right putamen, showing a trend for a difference between activity to SP emotions during SUP only. All median split analysis graphs are depicted with ± 1 S.E M. * = p< .01.*

lPFC and dlPFC to be importantly implicated in REAP (*down*- and *up*-regulation) as well as SUP of emotions, both negative and positive [7-10, 12]. Our new fMRI results therefore suggest that high avoidant subjects, if asked to naturally process SN emotions during EMEXP, employ much more cognitive regulation mechanisms through the recruitment of lPFC and dlPFC than securely or anxiously attached participants. This in turn might be explained by the fact that high avoidant individuals normally try to prevent any conscious experience and expression of social emotions through ER – especially through SUP [3] –, and are therefore much more affected by SN emotions if asked to process them naturally, being instructed not to employ any ER attempts.

Interestingly, we found the same positive relation between activity to SN emotions during EMEXP and AV in three areas of the dorsal, and one region of the ventral anterior cingulate cortex. Such activations were not found for the initial contrast SOC *versus* NSOC either. Usually, ACC activations are interpreted by applying a dorsal-ventral anatomical dissociation, relating more dorsal activity to cognitive and more ventral activity to emotional processes [31]. In addition, because ACC activity is known to represent neural correlates of conflict, dACC activation is normally associated with cognitive, and vACC activation with emotional conflict [32-34]. Finally, recent fMRI studies investigating the processing of social emotions have shown that vACC was specifically implicated during the evaluation [35] or implicit memory (Vrtička et al., in press) of situations involving social rejection. Our new results therefore suggest that high avoidant individuals experience much more cognitive as well as emotional conflict when asked to naturally process SN emotions, mirrored by increased dACC and vACC activity. Such findings can again be explained by the tendency of high avoidant individuals to normally suppress any conscious emotion experience (see above), therefore displaying a relatively high involvement in – particularly negative – social emotions if asked to attend to them naturally. In addition, as we found sustained increased activity during REAP in two dACC clusters for SN emotions by comparing high versus low AV, our data suggests that high avoidant subjects experienced more cognitive conflict associated with social negative scenarios not only during EMEXP, but also during cognitive re-evaluation (REAP). This nicely illustrates

SP>NSP AV POS

Region	*BA*	*Voxel*	*T-Value*	*p-Value*	*xyz*	*Correlation with AV*
Caudate left		*7*	*4.44*	*p< .001*	*-6 12 -6*	*SUP_SP; r= .449; p= .054 (*)*
SMA left		*7*	*3.95*	*p= .001*	*-3 -3 54*	*SUP_SP; r= .524; p= .021 **
Precentral cortex right		*21*	*4.8*	*p< .001*	*33 -12 66*	
Occipital Cortex left		*7*	*4.91*	*p< .001*	*-21 -84 -3*	

TABLE 4: Whole Brain Correlation with Avoidant Attachment Style during Suppression for Social (versus Non-Social) Emotions. *Peak coordinates are given in MNI space and listed with best estimates of anatomical location. BA = Brodmann's area, AV = avoidant attachment style, SMA = supplemental motor area. * = p< .01, (*) = marginally significant.*

the notion in AAT that REAP is not a very efficient ER strategy in relation with AV (see above). Also of interest is the fact that all abovementioned effects of AV on the processing of SN emotions during EMEXP (and REAP) were specific to AV, as we did not find any correlations of activity in PFC and ACC with AX for the same stimulus conditions. These activation patterns might be explained by the general tendency of high avoidant individuals to view SN emotions as goal-incongruent and undesirable – similarly to securely attached people –, quite in contrary to high anxious persons [3]. It is therefore logical to expect that AV, but not AX, will be associated with increased ER efforts and cognitive as well as emotional conflict during natural processing of SN emotions.

Regarding AV and its effect on brain activity to SN emotions during re-appraisal (REAP), we only found a modulatory role in bilateral amygdala (initial contrast SOC *versus* NSOC) and right putamen (whole brain multiple regression analysis).

The amygdala is known to play a central role in social as well as non-social emotion processing, most probably by detecting information that is appraised as self-relevant, based on one's needs, values, goals, or concerns [36, 37]. Because social stimuli are thought to have a stronger intrinsic importance in directing behavior and modulating interactions, they should activate the amygdala more strongly than comparable non-social stimuli. This has already been shown before [38, 39], Vrtička et al., submitted). In the present study, activity in bilateral amygdala to SN emotions during REAP was increased as a function of AV, mirrored by a significant difference in activation levels to SN emotions between low and high avoidant subjects during REAP. Interestingly, such an activation pattern emerged from the fact that amygdala activation to SN emotions during EMEXP was only *down*-regulated by low but not high avoidant subjects during REAP. Our data thus implies a relative inability of high avoidant subjects to *down*-regulate any effect of SN emotions on amygdala activity through cognitive re-evaluation. Such an activation pattern is consistent with AAT in two ways. Firstly, AV has been particularly linked with SUP as the ER strategy of choice, and at the same time with the non-use of REAP (see above). Second, SA (in our study the low AV group) is described in AAT to make successful use of REAP to perform ER [3]. Taken together, our new data provides for the first time direct evidence that negative affect induced by SN emotions activating bilateral amygdala during EMEXP is only reduced by REAP in low, but not in high avoidant individuals. Moreover, as we did not find any significant correlations of activity with AX and activation differences between low and high anxiety in bilateral amygdala, the reported effects are specific for the avoidance dimension of AAS.

In right putamen, we found positive relations between AV and activity to SN, and negative relations between activity to NSN emotions during REAP, but no such effects during EMEXP and SUP. These results thus reveal a selective effect of AV on activity in right putamen during cognitive re-evaluation of negative emotional scenes. Such an activation pattern is quite intriguing, because ventral striatum activation is normally thought to represent reward-related activity, thereby coding for subjective positive experience in relation with regulation success (see Discussion part in relation with caudate nucleus and putamen activation to SP emotions during SUP below).

Avoidant Attachment Style and Positive Social Emotions

Predictions regarding the influence of AV on the neural correlates of experience and regulation of social positive (SP) emotions were rather difficult, because AAT mostly focuses on attachment system activation by potential threats, which are mostly thought to be negative attachment-related scenarios. However, as mentioned above (see Introduction), AV has already been shown to also affect the processing of positive stimuli in general [13], and particularly SP emotions, representing social reward [15]. Moreover, as AV is generally characterized by a more negative model of others [14] and increased distrust of affective social signals [3], high avoidant individuals should regard particularly SP emotions as more negative and less genuine. Along these lines, we mainly expected to see less reward-related activity, together with increased activity in cortical areas representing mentalizing/theory of mind to SP emotions during EMEXP, due to more difficulties in natural processing and evaluation of positive social scenarios.

Our fMRI data showed that AV modulated activity to social positive (SP) emotions in medial orbito-frontal cortex (mOFC), supplementary motor area (SMA), left caudate nucleus, and right putamen, during both emotion experience (EMEXP) and SUP, but not REAP. In addition, we found a negative association between AV and BOLD signal change to non-social positive (NSP) emotions in SMA during EMEXP exclusively.

Regarding mOFC, previous studies have found associations between increases in activity in this brain area and reinforcement-guided decision making, especially in terms of context-sensitive evaluation of outcomes [40]. In addition, and specifically related to AAS, two other experiments report OFC involvement in regulating and encoding particularly the positive affect in attachment system functioning [41, 42]. According to the abovementioned characteristics of AV, such increased mOFC activation to SP emotions during EMEXP in high avoidant subjects probably means that

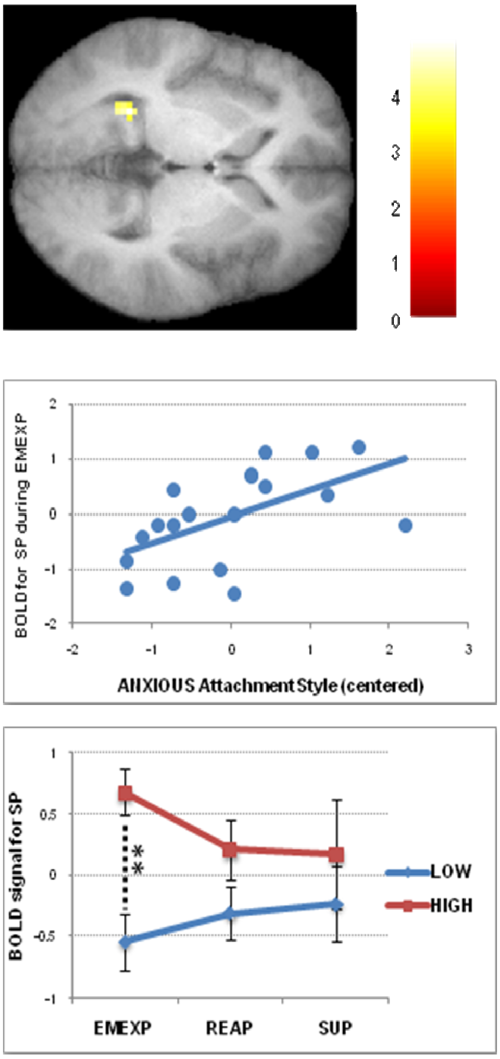


Figure 6: Brain Activations in relation with Anxious Attachment Style (AX). *a) Statistical parametric map for the whole brain multiple regression analysis using the contrast EMEXP SP versus NSP positively correlated with AX, showing increased activity in left parahippocampus (xyz=-27 -45 0; see Table 5 for more details). b) Significant correlation between parameter estimates (beta values) extracted from left parahippocampus and AV to SP images during EMEXP, averaged across voxels and participants (see Table 5 for more details). c) Median split analysis for high (red) versus low (blue) anxiety in left parahippocampus, showing a significant difference between activity to SP emotions during EMEXP only (depicted with ± 1 S.E M.). ** = p< .001.*

avoidant participants needed to attend to SP emotions more strongly, because for them, they were less intrinsically positive and less genuine, requiring more efforts to evaluate their “real” nature. Interestingly, such an effect was only found during EMEXP, but not during REAP and SUP. This might be explained by the fact that REAP was the condition with the least social implication (“pretend unreal”), and subjects probably directed their attention towards behavioral inhibition of emotion expression during SUP, thereby limiting cognitive processes related to outcome evaluation in the latter two viewing conditions VCON.

For SMA, our findings are particularly interesting, because a previous positron emission tomography (PET) study investigating underlying neural differences between voluntary facial movement generated by mimicking a laughing/smiling *versus* spontaneous laughter/smile induced by visual comics found a significant correlation between regional cerebral blood flow (rCBF) and the magnitude of electromyographic (EMG) activity in the SMA for voluntary, but not spontaneous facial displays of happiness [43]. Our data thus suggests that high avoidant subjects were more strongly inhibiting any facial muscle activity during EMEXP as well as SUP for SP, but less for NSP emotions. According to theory and recent findings on AV (see above), high avoidant participants were probably expressing their positive emotions in social *versus* non-social contexts during EMEXP to a lesser degree, because they prefer not getting involved in positive social interactions with others [44]. Moreover, because SUP corresponds to the “natural” ER strategy employed by high avoidant individuals, they might have more strongly inhibited their positive emotion expression related particularly to SP emotions during SUP, because this was the only experimental condition where they could freely do so. Caudate nucleus activation has previously been reported to be increased during successful inhibition of responses to happy expressions [45]. Because happy faces are normally associated with positive affect and reward [46] and increased activity in ventral striatum (including caudate nucleus) with reward processing [47], such activation was interpreted as overcoming of the increased difficulty to avoid positively valenced facial expressions. Along those lines, in our study, high avoidant subjects displayed increased caudate activity to SP emotions as a function of AV during SUP as well as EMEXP (even though the latter relation was not significant), suggesting reward-related representation of successful avoidance of SP emotions. Moreover, this effect was reversed during REAP, where low avoidant subjects had stronger caudate activity for SP emotions. As already mentioned above, such results are well corroborated by AAT, describing AV to be associated with the preferential use of SUP, but SA with the employment of REAP as ER strategy.

Finally, we found increased activity as a function of AV to SP emotions during SUP in right putamen. This effect was specific to SP emotions during SUP, as revealed by the median split analysis. Such an activation pattern can be interpreted similarly to that found in the left caudate (see above), because the putamen is also a part of the ventral striatum, which has previously been associated with reward processing [47]. Consequently, our data suggests that behavioral inhibition of SP emotions during SUP was associated with positive affect, and thus increased right putamen activity in high avoidant subjects. Implications of such a finding in relation with AAT have already been discussed above.

No such associations between mOFC, SMA, caudate and putamen activity were found for AX, implying that AV more strongly affects the neural correlates of evaluation and regulation of positive – and particularly SP – emotions, as shown previously [13, 15].

Anxious Attachment Style

AX is associated with hyperactivating strategies maintaining and even intensifying negative affect, especially concerning attachment-related scenarios [3]. In addition, AX is characterized by increased vigilance to signs of rejection or support and their related emotional states, and eased access to memories of and rumination about actual and potential threats [3]. We therefore hypothesized high anxious persons to display increased limbic and cortical activity to SN emotions during EMEXP.

Emotion Experience						
SP>NSP AX POS						
Region	*BA*	*Voxel*	*T-Value*	*p-Value*	*xyz*	*Correlation with AX*
Parahippocampus left		*8*	*4.92*	*p< .001*	*-27 -45 0*	*Pearson-r= .558; p= .013 **

TABLE 5: Whole Brain Correlations with Anxious Attachment Style for Social (versus Non-Social) Emotions. *Peak coordinates are given in MNI space and listed with best estimates of anatomical location. BA = Brodmann's area, AX = anxious attachment style. * = p< .01.*

Our fMRI data only revealed one brain region where BOLD signal change to social emotions was positively associated with AX scores, namely the parahippocampus, and this only during EMEXP for positive scenarios. The parahippocampus is thought to be involved in the more perceptual aspects of memory [48]. It might therefore be possible that this region was activated when anxious participants recalled situations in which they experienced SP emotions similar to those shown during scanning in our study. Because AX is normally associated with eased access to negative memories of threats through hyperactivating strategies [3], anxious subjects might have had to recruit additional resources to recall contexts where they previously experienced positive memories, in case of our study through activation of the left parahippocampus.

The absence of any effects of AX on experience and regulation of SN emotions was rather intriguing. However, there are several possible explanations.

On one hand, as already mentioned above, anxious individuals are described as perceiving SN emotions as goal-congruent, in contrary to avoidant and secure persons. Therefore, experience of SN emotions during EMEXP would not necessarily have to be related with increased activations in brain areas including lPFC, dlPFC, ACC, representing heightened cognitive control and cognitive as well as emotional conflict, as seen for AV. On the other hand, even though AX is associated with hyperactivating strategies *up*-regulating (particularly SN) emotions through REAP [3], there is no clear indication of a preferred ER strategy to *down*-regulate affect. This implies that anxious individuals could still follow instructions during REAP as well as SUP, thereby successfully reducing the impact of SN emotions. If so, it would be difficult to find any activation differences in high versus low anxious subjects during ER. This should however be investigated further, particularly regarding SUP, because there is behavioral evidence that attachment-anxious individuals actually have difficulties with suppressing negative thoughts and emotions, even when asked to do so [5]. Finally, there might be the possibility that high anxious individuals could not be very responsive to displays of SN emotions from the very fist, because they are too used to such scenarios. In this sense, it would be likely not to find any signal differences between high and low anxious individuals, because the activation levels should be comparable.

In any case, more research is needed to better describe and specify the neural correlates of emotion experience and regulation in association with AX.

CONCLUSION

This fMRI study investigated the influence of adult attachment style (AAS) on the experience and regulation of social emotions. Our new data reveals that, if asked to process social emotions naturally, high avoidant participants recruited several prefrontal brain areas implicated in emotion regulation (PFC), evaluation of outcomes (mOFC), and cognitive as well as emotional conflict (ACC). We also found sustained high ACC and amygdala activation to particularly social negative (SN) emotions during REAP in high avoidant subjects, suggesting a relative failure to *down*-regulate the experience of SN affect through cognitive re-evaluation. Finally, our data revealed increased activity as a function of AV during SUP and/or EMEXP in supplementary motor area (SMA) and ventral striatum (putamen and caudate) during behavioral inhibition of emotion expression to especially social positive (SP) emotions. All these findings are in accordance with AAT, providing for the first time direct evidence of underlying neural activity during ER associated with individual differences in AAS.

Concerning AX, we only found modulation of activity in left parahippocampus to SP during EMEXP. This might be interpreted by stronger efforts to recall situations/places where SP emotions comparable to the ones shown during the experiment have been experienced before. However, because no areas in the brain were found where AX was positively associated with activity representing the processing of SN emotions, more research is needed to further elucidate brain mechanisms in relation with the anxiety dimension of AAS.

Our findings importantly extend the knowledge on underlying neural correlates of ER associated with AAS and could ultimately help designing more specific approaches for therapeutic interventions in social dysfunctions.

AKNOWLEDGMENT

This research was supported by the National Center of Competence in Research (NCCR) Affective Sciences financed by the Swiss National Science Foundation (n° 51NF40-104897) and hosted by the University of Geneva, and grants of the Swiss National Science Foundation to DS and PV.

REFERENCES

1. Bowlby, J., *Attachment and loss*. 1969, New York,: Basic Books. v.
2. Bowlby, J., *Attachment and loss*. 2nd ed. 1982, New York: Basic Books. v.
3. Mikulincer, M. and P.R. Shaver, *Attachment in Adulthood: Structure, Dynamics, and Change*. 2007, New York: The Guilford Press.
4. Gross, J.J., *Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology.* Journal of Personality and Social Psychology, 1998. 74(1): p. 224-237.
5. Fraley, R.C. and P.R. Shaver, *Adult attachment and the suppression of unwanted thoughts.* Journal of Personality and Social Psychology, 1997. 73(5): p. 1080-1091.

6. Gillath, O., et al., *Attachment-style differences in the ability to suppress negative thoughts: Exploring the neural correlates.* Neuroimage, 2005. 28(4): p. 835-847.
7. Levesque, J., et al., *Neural circuitry underlying voluntary suppression of sadness.* Biological Psychiatry, 2003. 53(6): p. 502-510.
8. Ochsner, K.N., et al., *Rethinking feelings: An fMRI study of the cognitive regulation of emotion.* Journal of Cognitive Neuroscience, 2002. 14(8): p. 1215-1229.
9. Ochsner, K.N., et al., *For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion.* Neuroimage, 2004. 23(2): p. 483-499.
10. Kim, S.H. and S. Hamann, *Neural correlates of positive and negative emotion regulation.* Journal of Cognitive Neuroscience, 2007. 19(5): p. 776-798.
11. Gillath, O., et al., *Individual differences in thought suppression: An fMRI exploration.* Journal of Cognitive Neuroscience, 2005: p. 199-199.
12. Goldin, P.R., et al., *The neural bases of emotion regulation: Reappraisal and suppression of negative emotion.* Biological Psychiatry, 2008. 63(6): p. 577-586.
13. Niedenthal, P.M., et al., *Adult attachment and the perception of facial expression of emotion.* Journal of Personality and Social Psychology, 2002. 82(3): p. 419-433.
14. Bartholomew, K. and L.M. Horowitz, *Attachment Styles among Young-Adults - a Test of a 4-Category Model.* Journal of Personality and Social Psychology, 1991. 61(2): p. 226-244.
15. Vrtička, P., et al., *Individual attachment style modulates human amygdala and striatum activation during social appraisal.* PLoS ONE, 2008. 3(8): p. e2868.
16. Fujita, F., E. Diener, and E. Sandvik, *Gender Differences in Negative Affect and Well-Being - the Case for Emotional Intensity.* Journal of Personality and Social Psychology, 1991. 61(3): p. 427-434.
17. Phan, K.L., et al., *Neural correlates of individual ratings of emotional salience: a trial-related fMRI study.* Neuroimage, 2004. 21(2): p. 768-780.
18. Ainsworth, M.D.S., *Bowlby-Ainsworth Attachment Theory.* Behavioral and Brain Sciences, 1978. 1(3): p. 436-438.
19. Chotai, J., et al., *Adolescent attachment styles and their relation to the temperament and character traits of personality in a general population.* European Psychiatry, 2005. 20(3): p. 251-259.
20. Feeney, J.A. and S.M. Ryan, *Attachment Style and Affect Regulation - Relationships with Health Behavior and Family Experiences of Illness in a Student Sample.* Health Psychology, 1994. 13(4): p. 334-345.
21. Brennan, K.A., C.L. Clark, and P.R. Shaver, *Self-report measurement of adult attachment: An integrative overview.*, in *Attachment theory and close relationships*, J.A. Simpson and W.S. Rholes, Editors. 1998, Guilford Press: New York. p. 46-76.
22. Griffin, D. and K. Bartholomew, *Models of the Self and Other - Fundamental Dimensions Underlying Measures of Adult Attachment.* Journal of Personality and Social Psychology, 1994. 67(3): p. 430-445.
23. Kurdek, L.A., *On being insecure about the assessment of attachment styles.* Journal of Social and Personal Relationships, 2002. 19(6): p. 811-834.
24. Simpson, J.A., W.S. Rholes, and J.S. Nelligan, *Support Seeking and Support Giving within Couples in an Anxiety-Provoking Situation - the Role of Attachment Styles.* Journal of Personality and Social Psychology, 1992. 62(3): p. 434-446.
25. Aiken, L.S. and S.G. West, *Multiple regression: Testing and interpreting interactions*. 1991, Newbury Park, CA: Sage.
26. Gross, J.J. and O.P. John, *Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being.* Journal of Personality and Social Psychology, 2003. 85(2): p. 348-362.
27. Spielberger, C.D., *Manual for the State-Trait Anxiety Inventory*
Palo-Alto: Consulting Psychologists Press, 1983.
28. Spielberger, C.D., Sydeman, S. J., Owen, A. E., & Marsh, B. J., *Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI).* In M. D. Maruish (Ed.), The use of psychological testing for treatment planning and outcomes assessment (2nd ed.) (pp. 993-1021). Mahwah: Laurence Erlbaum, 1999.
29. Mikulincer, M. and E. Sheffi, *Adult attachment style and cognitive reactions to positive affect: A test of mental categorization and creative problem solving.* Motivation and Emotion, 2000. 24(3): p. 149-174.
30. Crawford, T.N., P.R. Shaver, and H.H. Goldsmith, *How affect regulation moderates the association between anxious attachment and neuroticism.* Attachment & Human Development, 2007. 9(2): p. 95-109.
31. Amodio, D.M. and C.D. Frith, *Meeting of minds: the medial frontal cortex and social cognition.* Nature Reviews Neuroscience, 2006. 7(4): p. 268-277.
32. Carter, C.S. and V. van Veen, *Anterior cingulate cortex and conflict detection: An update of theory and data.* Cognitive Affective & Behavioral Neuroscience, 2007. 7(4): p. 367-379.
33. Egner, T., et al., *Dissociable neural systems resolve conflict from emotional versus nonemotional distracters.* Cerebral Cortex, 2008. 18(6): p. 1475-1484.
34. Mohanty, A., et al., *Differential engagement of anterior cingulate cortex subdivisions for cognitive and emotional function.* Psychophysiology, 2007. 44(3): p. 343-351.
35. Somerville, L.H., T.F. Heatherton, and W.M. Kelley, *Anterior cingulate cortex responds differentially to expectancy violation and social rejection.* Nature Neuroscience, 2006. 9(8): p. 1007-1008.
36. Sander, D., J. Grafman, and T. Zalla, *The human amygdala: an evolved system for relevance detection.* Rev Neurosci, 2003. 14(4): p. 303-16.
37. Vuilleumier, P. and D. Sander, *Trust and valence processing in the amygdala.* Social Cognitive and Affective Neuroscience, 2008. 3(4): p. 299-302.
38. Hariri, A.R., et al., *The amygdala response to emotional stimuli: A comparison of faces and scenes.* Neuroimage, 2002. 17(1): p. 317-323.
39. Britton, J.C., et al., *Neural correlates of social and nonsocial emotions: An fMRI study.* Neuroimage, 2006. 31(1): p. 397-409.
40. Rushworth, M.F.S., et al., *Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour.* Trends in Cognitive Sciences, 2007. 11(4): p. 168-176.
41. Nitschke, J.B., et al., *Orbitofrontal cortex tracks positive mood in mothers viewing pictures of their newborn infants.* Neuroimage, 2004. 21(2): p. 583-592.
42. Minagawa-Kawai, Y., et al., *Prefrontal Activation Associated with Social Attachment: Facial-Emotion Recognition in Mothers and Infants.* Cerebral Cortex, 2009. 19(2): p. 284-292.
43. Iwase, M., et al., *Neural substrates of human facial expression of pleasant emotion induced by comic films: a PET Study.* Neuroimage, 2002. 17(2): p. 758-68.
44. Dozier, M. and R.R. Kobak, *Psychophysiology in Attachment Interviews - Converging Evidence for Deactivating Strategies.* Child Development, 1992. 63(6): p. 1473-1480.
45. Hare, T.A., et al., *Contributions of amygdala and striatal activity in emotion regulation.* Biological Psychiatry, 2005. 57(6): p. 624-632.
46. O'Doherty, J., et al., *Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness.* Neuropsychologia, 2003. 41(2): p. 147-155.
47. Montague, P.R. and G.S. Berns, *Neural economics and the biological substrates of valuation.* Neuron, 2002. 36(2): p. 265-284.
48. Burgess, N., E.A. Maguire, and J. O'Keefe, *The human hippocampus and spatial and episodic memory.* Neuron, 2002. 35(4): p. 625-641.

VI.6. fMRI Study N°3

In Preparation for Publication

One central aspect of adult attachment theory (AAT) is the notion of self-dependence and other-avoidance (see Chapter IV.4.6.3. and [183, 184]), leading to different self- and other-perceptions as a function of avoidance (AV) and anxiety (AX). Even though such accounts have recently not been assessed using a categorical model of adult attachment style (AAS) [183, 184] anymore, they still remain important hallmarks of the respective attachment styles [185]. Consequently, fMRI study N°3 had the following main question:

- **How does AAS influence the neural correlates of self- *versus* other-perception?**

According to recent accounts that individual differences in adult attachment style are more reliably assessed by dimensional models [188, 199, 200], AV and AX were used as continuous measures to compare their impact on the ascription of positive and negative trait attributes to either oneself or a close friend.

The behavioral results showed that, in general, positive traits were more readily attributed to the self as well as to close friends (as opposed to negative ones). However, correlations with AAS revealed that AV was marginally significantly positively related to the ascription of positive traits to the self.

The fMRI data revealed that AV, as already suggested by behavioral data, was associated with different brain activation patterns implying preferential processing of positive self-traits, including superior temporal sulcus, fusiform cortex, and ventro-medial prefrontal cortex. In turn, AX was linked with more pronounced processing of other-traits in general (medial prefrontal cortex), increased representation of negative other- (orbito-frontal cortex), as well as negative self- (temporo-parietal junction and superior frontal cortex) traits.

The Neural Correlates underlying Self- and Other-Perception are modulated by Individual Differences in Adult Attachment Style

Karim N' Diaye (1,2), Pascal Vrtička (1,2), Ulrich Wagner (2), David Sander (1,3), & Patrik Vuilleumier (1,2)

(1) Swiss Center for Affective Sciences, University of Geneva
(2) Laboratory for Neurology & Imaging of Cognition, Department of Neurology and Department of Neuroscience,
University Hospital and Medical School, University of Geneva
(3) Department of Psychology, FPSE, University of Geneva

ABSTRACT

Attachment theory suggests that the distinct secondary attachment strategies involving either hyper- (anxiety; AX) or de- (avoidance; AV) activation of the attachment system originate from important differences in the underlying attachment working models. The latter were initially characterized with a four-category model describing high other-avoidance and low self-dependence as hallmarks of AV, but low other-avoidance and high self-dependence as key features of AX. Even though more recent theoretical accounts favor a bi-dimensional attachment-space organization and challenge some self- and other-representations linked with AX and AV, it is still common to use different models of the self and others to investigate individual differences in attachment style.

The present functional magnetic resonance imaging (fMRI) study for the first time reveals that AV is not only behaviorally, but also functional-anatomically linked with increased positive self-representations in superior temporal sulcus, fusiform cortex, as well as ventro-medial prefrontal cortex. In turn, AX was found associated with generally increased other-representations in medial prefrontal cortex, and more pronounced activation to negative other-evaluation in orbito-frontal cortex, temporo-parietal junction and superior frontal cortex. These findings are discussed according to the general attachment literature as well as previous evidence from fMRI experiments in social neuroscience.

INTRODUCTION

First described by Mary Ainsworth and John Bowlby [1-3] about four decades ago, attachment theory provides an important means for investigating various processes regarding emotional bond formation between people. The theory was initially developed on the basis of mother-child relationships, where the main function of the attachment system was described as to maintain proximity to significant others (i.e. the mother as main attachment figure) in times of stress, or in other words, to regulate support seeking behavior. Through repeated interactions with attachment figures in times of need, such proximity seeking tendencies are thought to become more and more cognitively encoded and finally stored as mental representations of self and others, referred as to *attachment working models* (AWMs). These *"… allow for mental simulation and prediction of likely outcomes of various attachment behaviors …"* [4] and form the basis of a person's individual adult attachment style (AAS).

The first study aimed at investigating such cognitive self / other underpinnings of AAS was conducted by Hazan and Shaver in the context of (adolescent and adult) romantic love [5, 6]. Even though the authors only used a preliminary self-report questionnaire to assess the AWMs of self and others in this experiment, they could already show that different AAS were predictably related to beliefs about self and relationships. Elaborating on these findings, Bartholomew and Horowitz [7, 8] subsequently introduced a new measure to assess AAS by means of positive and negative AWMs of self and others with their *relationships questionnaire* (RQ), including four independent AAS categories as a function of self-dependence and other-avoidance. Thereby, the main hallmarks of secure attachment style (SAS) were characterized with both positive self- and other-models. In turn, avoidant attachment style (AV) was described with a negative other- but a positive self-model, whereas anxious attachment style (AX) was associated with a negative self- but a positive other-model. Such distinct mental representations of the self and others were linked with different patterns of repeated interactions with attachment figures during development [9].

In the case of available and responding attachment figures providing a "secure base" for restoring emotional balance in times of stress – important pre-requisites for the emergence of a SAS –, people are able to not only develop a positive model of others associated with supportiveness and trustworthiness, but also to attribute positive traits such as worthy, competent, lovable and special to themselves, entailing the establishment of a positive self-model [4]. In contrast, if attachment figures are repeatedly experienced as unavailable and unresponsive in times of need, the emergence of two different kinds of psychological pain [10] leads to the establishment of either *de*-activating secondary attachment strategies linked with AV, or *hyper*-activating strategies associated with AX. In the case of AV, proximity seeking is viewed as futile or even dangerous because of the distress felt by failing to achieve or maintaining proximity to an attachment figure.

CORRESPONDING AUTHOR:
Pascal Vrtička
Swiss Center for Affective Sciences
7, rue des Battoirs, 1205 Geneva, Switzerland
Tel: +41 – (0)22 379 9824 Fax: +41 – (0)22 379 9844
Email: pascal.vrticka@unige.ch

Consequently, avoidant people develop a critical approach to and a negative model of others, a perception which is further strengthened by the dismissal of any positive other traits. Concomitantly, avoidant people tend to suppress any negative self-aspects and to boost their positive features related to independence with the goal to overcome the felt rejection by others, leading to the emergence of a positive self-model. Regarding AX, the perceived failure to handle threats autonomously encourages people to intensify their support-seeking attempts despite the fact that attachment figures are experienced as unresponsive. In this constellation, others are still viewed as (partly – see below) positive due to the desire for attention and protection. However, constant social rejection leads to an increased sense of helplessness and vulnerability, paired with doubts about self-worth and -efficacy, all hallmarks of a negative self-model (see [4] for a more thorough review).

More recently, such a categorical distinction between AV and AX in terms of self-dependence and other-avoidance has been challenged in two ways. Firstly, AAS has been found to be more reliably measurable by continuous attachment dimensions rather than categories [11], favoring the employment of the *relationships scales questionnaire* (RSQ) [9] with a specific model of analysis according to Simpson [12, 13]. Such a dimensional attachment space organization suggests that – like for other personality traits, like general anxiety, neuroticism, etc. – AAS features can be present with varying degrees rather than fitting into fixed categories. Second, whereas the assumed positive self-model of AV has often been found associated with defensive self-enhancement and thus not representing a genuine positive self-perception, the positive other-model of AX has often been found linked with simultaneous negative other-constructs due to constant experiences of social rejection tantamount with a notion of ambivalence [4]. As a consequence, it seems that straight-forward conclusions regarding AAS can only be drawn for other-models (negative) as a function of AV and self-models (negative) in the context of AX, but more caution is advised regarding the respective inverse social perceptions.

Moreover, even though much behavioral evidence regarding such cognitive representations of the self and others in form of AWMs related to AAS is available (see e.g. [4]), no study using neuro-imaging methods has specifically investigated the neural correlates underlying self- and other-perception as a function of AAS up to date. There is, however, some indirect neuro-imaging evidence that AV is linked with decreased experience of positive emotions in relation with rewarding feedback from others as e.g. during social support, whereas AX seems to lead to increased neural representation of social negative emotions during e.g. social punishment [14].

Bearing in mind the above-mentioned issues, we adopted a functional magnetic resonance imaging (fMRI) paradigm widely used for investigating self- and other-representation [15-18], during which participants had to attribute positive or negative trait-adjectives either to themselves or to a close other. As an additional new step, we then computed whole-brain multiple regression analyses to pull out those voxels within the brain where self- and/or other-perception was modulated as a function of AAS, and particularly AV and AX (see Methods for details).

Regarding our predictions, we referred to recent findings showing that mentalizing about self and others employs largely overlapping brain areas, including prefrontal cortex (PFC), posterior cingulate cortex (PCC), as well as temporal parietal (TPC), anterior temporal (ATP) and somatosensory (SC) cortex [19, 20]. Activity in these brain regions has mainly been associated with theory of mind (TOM; [21-23]), the representation of self- and other-knowledge [24, 25], and with social decision making including moral reasoning [26, 27]. We therefore hypothesized that AV would be associated with increased activity in this social cognition network during the attribution of negative trait-adjectives to others (negative other-model), whereas the same should hold true for AX in the context of negative trait-attribution to the self (negative self-model). What is concerning brain activation as a function of AV associated with self- and AX with other-perception, predictions were more difficult (see above). However, we expected to find more modulation of activity in the social cognition network (PFC, PCC, TPC, ATP, SC) by AV during positive self-attribution (positive self-model, or at least eased access to positive self-traits due to defensive self-inflation), and by AX during mentalizing about others in general (employment of *hype*r-activating strategies for all social clues) and more strongly during negative trait-attribution to others (increased vigilance to signs of particularly social rejection).

METHODS

Participants and task

This study involved 33 participants (24 females, mean age: 29.6 years) with no history of neurological or psychiatric illness. All had normal or corrected-to-normal vision. Thirteen participants also participated in another study bearing on memory which will be reported elsewhere. All participants gave informed consent in agreement with the local ethical committee regulation. Before scanning, participants filled the Relationship Style Questionnaire (RSQ; [9]) and Rosenberg’s [28] self-esteem questionnaire.

The task performed in the scanner required participants to rate how much adjectives describing personality traits would apply to themselves or to one of their best friends whom they picked at the beginning of the experiment (without revealing his or her identity to the experimenter). More precisely, we used a block-design scheme alternating conditions of self-focus, other-focus and a control condition (non-semantic syllable counting task). Adjectives were taken from Anderson’s database [29]. An equal number of positive valence and negative valence words was selected and translated into French and German by two independent translators and non-concordant translations were solved by consensus yielding to 55 positive (e.g. sincere, polite, confident) and 55 negative (e.g., hypocrite, cruel, clumsy) traits adjectives.

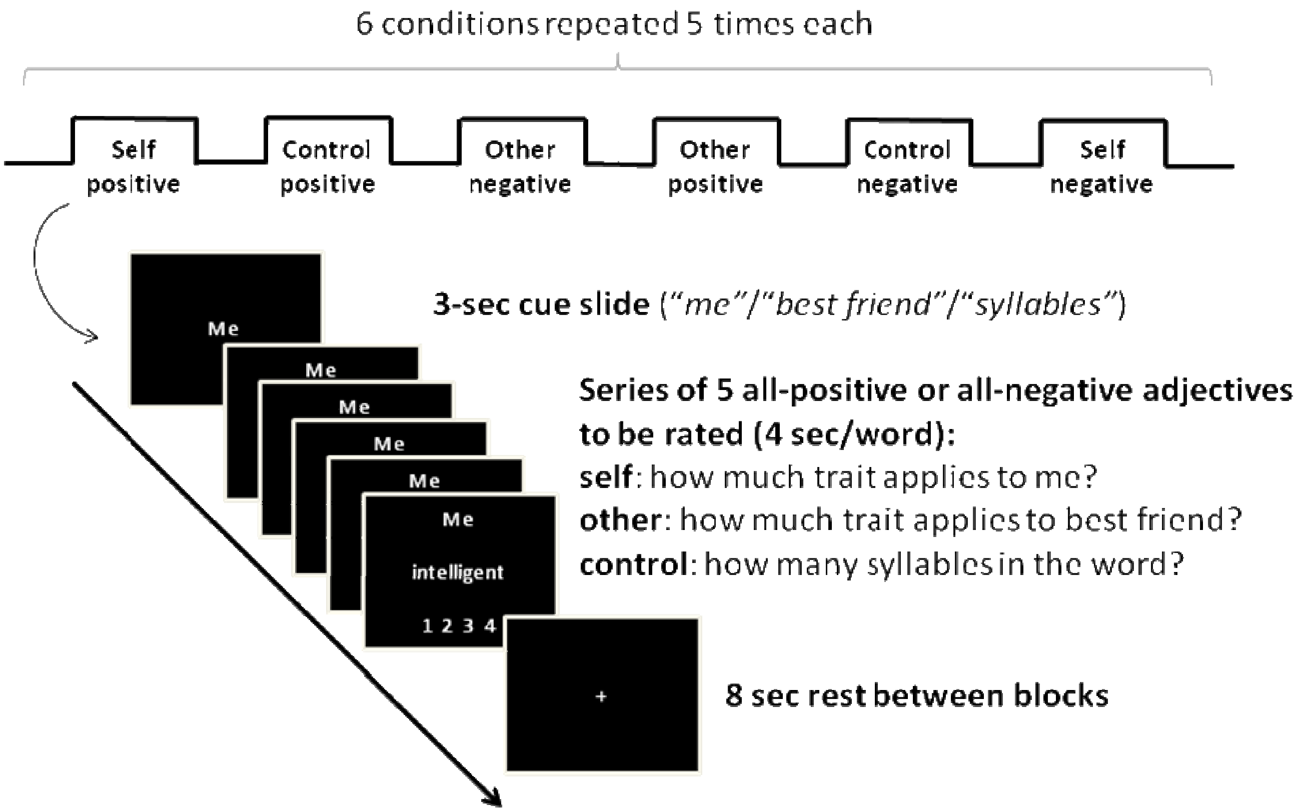


***Figure 1: Illustration of the experimental paradigm**. The four experimental conditions (positive or negative trait adjective attribution either to the self or other [best friend], and two control conditions (syllable counting of positive or negative trait adjectives) were presented in blocks of 5 words each, and every block was repeated 5 times (with different words). For more information, please refer to the text.*

Each block was introduced by a 2-sec cue slide indicating the type of judgment to be performed (self / other / syllables) on the following 5 adjectives, which were all of the same valence, either positive or negative. Each adjective word was presented for 4 sec in the middle of the screen (see Figure 1). In the two personality rating conditions, subjects used a 4-point scale describing how much the adjectives apply to themselves (self condition) or to their best friend (other condition), ranging from 1 (not at all like me / him) to 4 (perfectly like me / him). In the control condition, participants used the same 4-point scale but indicated the number of syllables present in the words (from 1 to 4 or more). Responses were given by pressing with the finger of the right hand one of the 4 matching buttons on a MR-compatible respond pad. Blocks were separated by an 8-sec blank screen. The whole experiment consisted of two runs. Each run comprised 30 blocks, 10 per condition (5 using positive and 5 using negative traits). Orders of blocks were pseudo-randomized within and across subjects so as to avoid two blocks of the same conditions to follow one after the other.

Data acquisition & imaging

Scanning was performed at Geneva University Hospital,Center for Bio-Medical Imaging, on a 3T Trio MRI scanner manufactured by Siemens (Erlangen, Germany). Standard functional EPI T2*-weighted volumes were collected as 36 3-mm contiguous axial slices: TR = 2.2 s, TE = 30 ms, Flip Angle = 85°, FOV : 240mm, in-plane resolution of 3 by 3mm. One structural T1-weighted image using standard MPRAGE sequence was also collected from each subject (TR = 2.2 s, TE = 2.89 ms, TI = 1 s, voxel size: 0.9*0.9*0.9 mm3). Presentation of the stimuli and collection of the behavioral responses from the subjects were made using E-Prime software (PST Software Inc., Pittsburgh, US-PA) running on a PC computer operated under Windows XP (Microsoft, Redmond, US-WA). Due to technical failure, behavioral data from two participants were lost.

Data analysis

Behavioral data were analyzed using Excel 2003 (Microsoft, Redmond, PA) and Matlab (Mathworks, BC). For each participant, mean ratings for positive and negative traits were separately computed for self- and other-focus conditions. Accuracy in the control task was evaluated as the match between the number of syllables reported by the participant and the correct answer decided by two independent native-speakers. Questionnaire data were analyzed according to the recent recommendations of Kurdek [11] based on Simpson [12, 13]. More precisely, using 13 items from the RSQ questionnaire, we derived two attachment style scores, one for the anxious attachment style dimension (AX) and one for the avoidant one (AV).

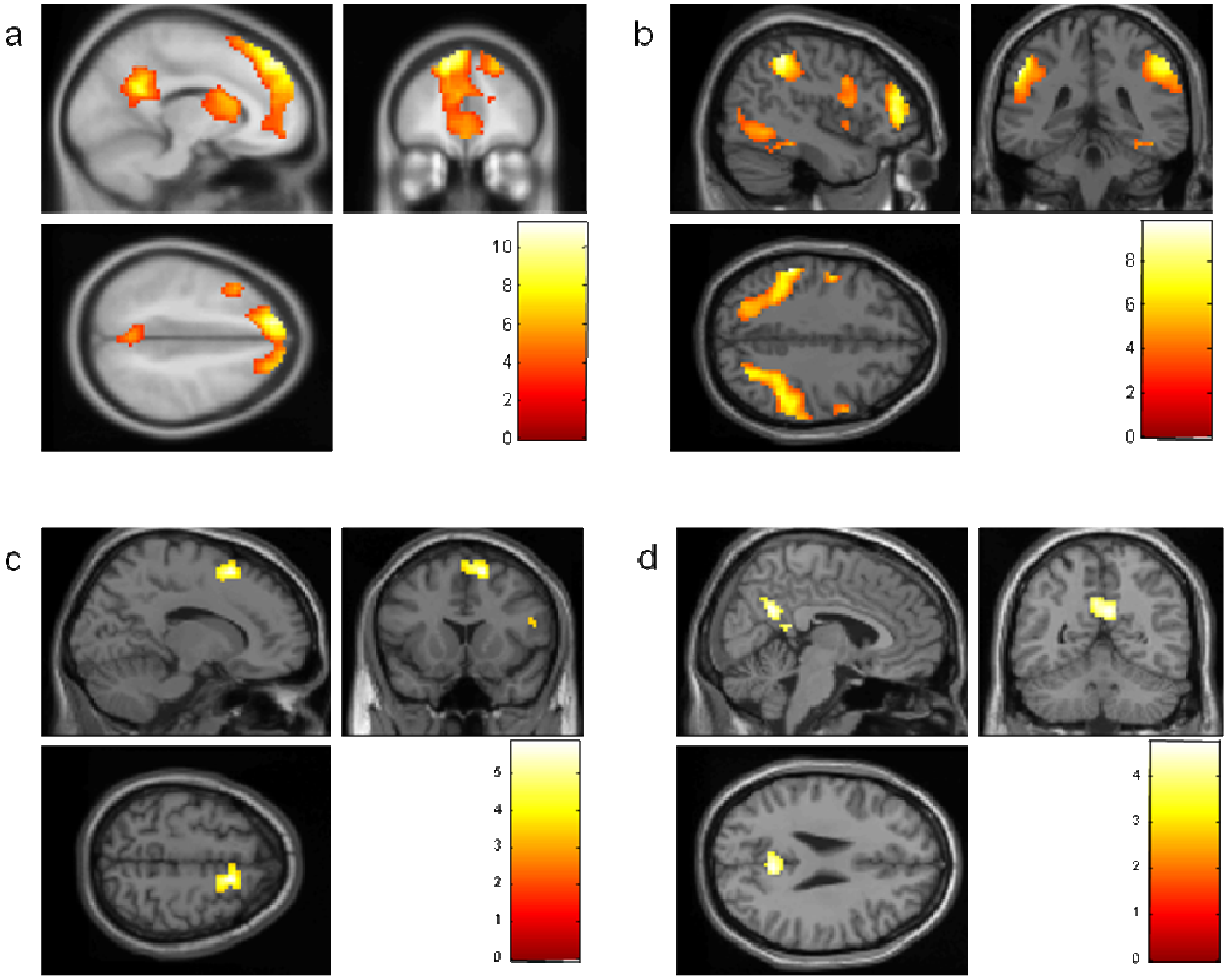


Figure 2: Statistical parametric maps of the regions activated by the main effects. *a) The task of interest (combining self and other ratings) versus the control task (contrast [SELF*$_{(POS + NEG)}$ *+ OTHER*$_{(POS + NEG)}$ *> CONTROL]; b) the inverse contrast [CONTROL > SELF*$_{(POS + NEG)}$ *+ OTHER*$_{(POS + NEG)}$*]; c) the specific comparison between self and other ratings (contrast [SELF*$_{(POS + NEG)}$ *> OTHER*$_{(POS + NEG)}$*]; and d) the inverse contrast [OTHER*$_{(POS + NEG)}$ *> SELF*$_{(POS + NEG)}$*]. All maps are thresholded at p<0.001 uncorrected, cluster size > 50; only maxima being 8-mm apart appear.*

Functional images were analyzed in the framework of a general linear model using the SPM2 Toolbox (Wellcome Dept. of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm) under Matlab environment running on PC computers equipped with Windows XP. EPI volumes were realigned, normalized to the MNI (Montreal Neurological Institute) template, re-sampled to a voxel-size of 3 mm, and spatially smoothed using a 8 mm FWHM Gaussian kernel. Coordinates will thus be given in millimeters in the MNI stereotaxic space.

For each subject, the six different conditions (Positive / negative valence words; self focus / other focus / control task) were modeled as 20-s boxcar functions starting from the first word of five in each block and convolved with the standard hemodynamic response. To capture variance due to movement, realignment parameters were incorporated as six additional regressors of no interest. During the estimation of the model, a high-pass frequency filter (cutoff 128 s) and corrections for auto-correlation between scans were applied to the time series. Random-effects were evaluated by combining contrast images computed from individual analyses. Contrasts maps are reported with a significance threshold of p<0.001 uncorrected for multiple comparisons (equivalent to a T>3.37 with 32 d.o.f.), with a cluster extent threshold of 50 voxels. Whole-brain correlation between contrasts effects and AAS dimensions scores (avoidance, AV and anxiety, AX) were computed using a multiple regression model incorporating both measures; significant correlations are here reported with an uncorrected threshold of p<0.005 and a clusters extent threshold of 25 voxels within the mask derived from the activations seen in the main task (*[SELF*$_{(POS + NEG)}$ *+ OTHER*$_{(POS + NEG)}$ *> CONTROL]*).

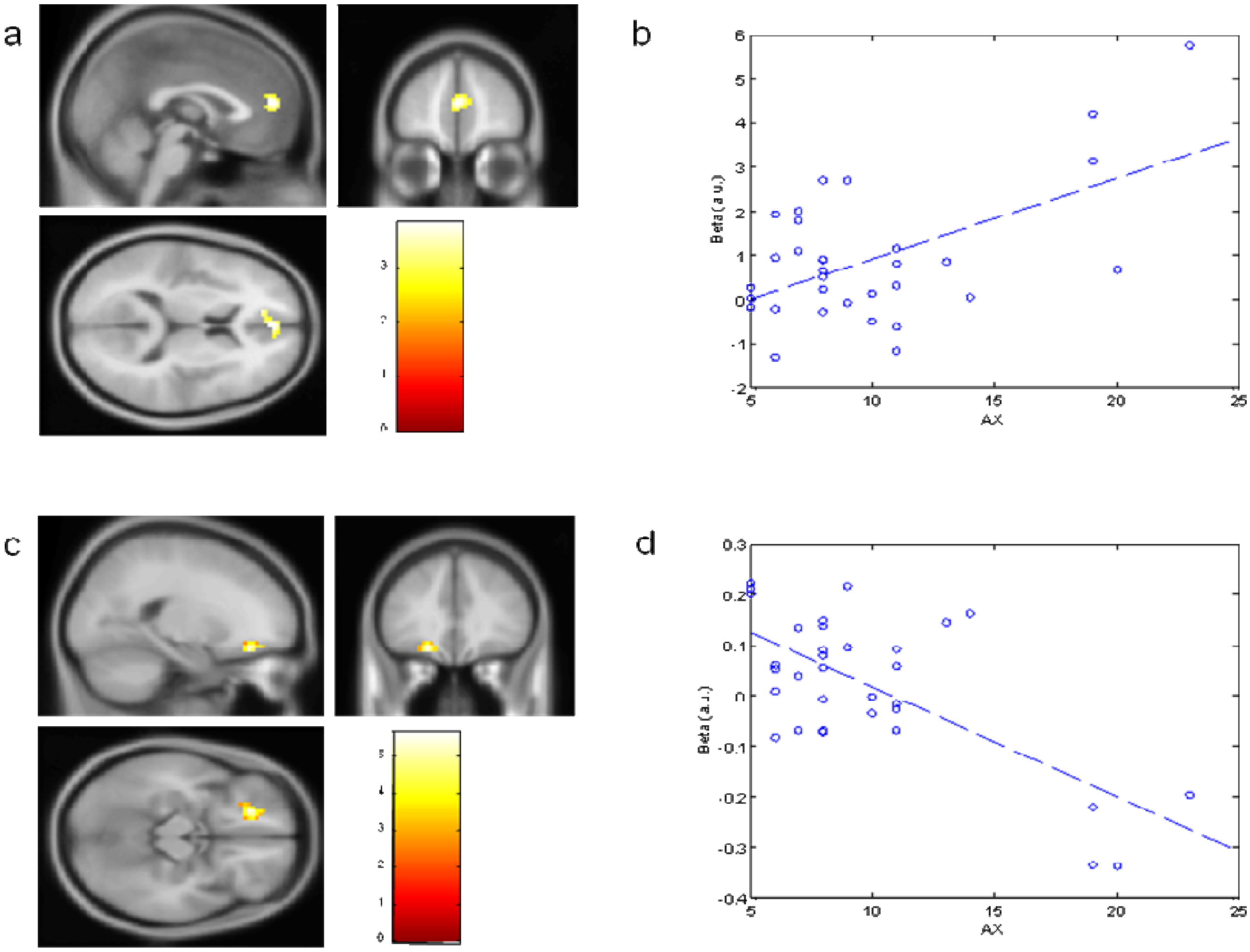


Figure 3: Correlations between anxious attachment and other-evaluations. *a) Statistical parametric map for the whole brain correlation [OTHER$_{(POS + NEG)}$ > CONTROL] x AX, displaying a positive effect in mPFC. b) Correlation graph depicting a positive correlation between AX (x-axis) and the differential response in mPFC (y-Axis). c) Statistical parametric map for the whole brain correlation [OTHER$_{POS}$ > OTHER$_{NEG}$] x AX, displaying a negative effect in OFC. d) Correlation graph depicting a negative correlation between AX (x-axis) and the differential response in OFC (y-Axis).*

RESULTS

Behavioral and questionnaire data

Behavioral ratings revealed a significant main effect of the valence of trait, positive traits being rated to match better than negative ones (mean ratings: 3.08 vs. 1.48; F=224.9; $p<0.05$) but no main effect of the focus (self: 2.45 vs other: 2.40; F=0.2; $p>0.1$) nor interaction between those two factors (F = 2.5; $p>0.1$). Response times revealed no main effect of valence (F<0.1) nor focus (F=1.8) and no significant interaction between the two (F<0.1).

Scoring of the RSQ resulted in AV ranging from 11 to 33 (mean= 19.2 ± 5.1) and AX ranging from 5 to 23 (mean= 9.8 ± 4.5) across participants, with no significant correlation between the two scores ($r2<0.01$; $p>0.1$). There was no significant correlation between the ratings in the main task and the attachment styles (all $r2 < 0.05$; all p's >0.1), except for a marginally significant correlation between positive self ratings and AV (r2= 0.11, p=0.07).

Imaging results

Main effects

The contrast [SELF$_{(POS + NEG)}$ + OTHER$_{(POS + NEG)}$ > CONTROL] showed increased activity in a bilateral network encompassing the dorsal medial and rostral cingulate PFC, the retrosplenial medial parietal / cingulate, left and right temporo-parietal junctions as well as middle temporal gyrus and temporo polar regions (see Table 1 and Figure 2a). Conversely, the control task (CONTROL > SELF$_{(POS + NEG)}$ + OTHER$_{(POS + NEG)}$) involved a bilateral fronto-parietal network (see Table 1 and Figure 2b). Comparing the self versus other focus (SELF$_{(POS + NEG)}$ > OTHER$_{(POS + NEG)}$) showed that self-evaluation was linked to increased activity in dorsal ACC, dorsal mPFC, right premotor and left temporo-parietal regions (see Table 2 and Figure 2c). The reverse contrast (OTHER$_{(POS + NEG)}$ > SELF$_{(POS + NEG)}$) showed that focusing on other increased activity in retrosplenial regions (see Table 2 and Figure 2d).

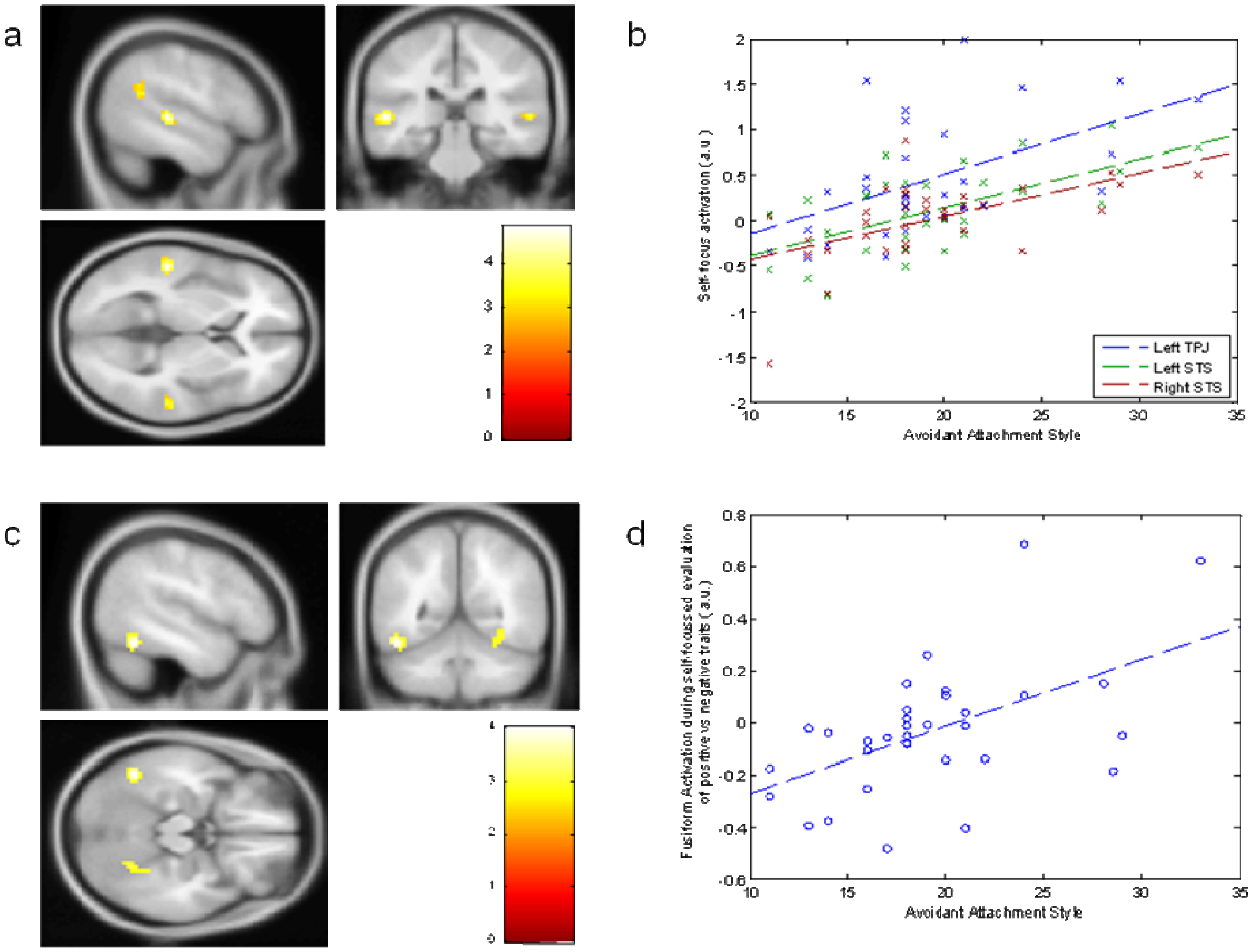


Figure 4): Correlations between avoidant attachment and self-evaluation. *a) Statistical parametric map for the whole brain correlation [SELF$_{(POS + NEG)}$ > CONTROL] x AV, displaying a positive effect in left STS. b) Correlation graph depicting positive correlations between AV (x-axis) and the differential response in bilateral STS and left TPJ (y-Axis). c) Statistical parametric map for the whole brain correlation [SELF$_{POS}$ > SELF$_{NEG}$] x AV, displaying a positive effect in left fusiform cortex. d) Correlation graph depicting a positive correlation between AV (x-axis) and the differential response in left FC (y-Axis).*

Correlation of brain responses with AAS

We first evaluated how attachment style modulated the evaluation of others (trait adjective attribution to a close friend). This was explored by correlating the attachment scores with the contrast between the OTHER focus conditions (positive and negative) and the control task ([OTHER$_{(POS + NEG)}$ > CONTROL] x AAS). By doing so, we only found a positive correlation between AX and differential activity in mPFC (xyz= 0 51 12; r2 = .308; p=0.0008; see Figure 3ab). We then more specifically assessed whether AAS differently modulated the evaluation of positive *versus* negative traits in the other condition ([OTHER$_{POS}$ > OTHER$_{NEG}$] x AAS). This only revealed a significant negative relation between differential activity in OFC (xyz= 21 33 15) and AX (r2 = 0.47; p<0.0001; see Figure 3cd). No effects were found for the inverse contrast ([OTHER$_{NEG}$ > OTHER$_{POS}$] x AAS).

Subsequently, we investigated the modulatory role of AAS on brain activity underlying self-evaluations. In the SELF focus condition in general (compared to the control task; [SELF$_{(POS + NEG)}$ > CONTROL] x AAS), AV was positively correlated with differential activity in the superior temporal sulcus bilaterally (STS; left: xyz= -51 -30 3, r2= .43, p< .001; right: 57 -27 0, r2= .32, p= .007), and in the left supramarginal gyrus (xyz= -48 -51 27, r2= .29, p= .001; see Figure 4ab). Subsequently, by looking at positive *versus* negative self evaluation ([SELF$_{POS}$ > SELF$_{NEG}$] x AAS) more specifically, we found that AV was positively correlated with activity in the ventral temporo-occipital regions including the left fusiform cortex (xyz= -48 -54 -18; r2= 0.28; p= 0.002; see Figure 4cd). Finally, by looking at the condition of SELF evaluation of positive traits exclusively (compared to baseline, ie. Raw betas; [SELF$_{POS}$]), we observed a significant positive correlation between AV and differential activation of the ventral mPFC (xyz= -6 48 -6; r2=0.305; p = 0.0009; see Figure 5ab) as well as other cortical midline structures including the superior frontal gyrus (xyz= 0 36 51) and posterior cingulate (xyz= -6 -48 9). The exclusive inverse contrast (negative traits; [SELF$_{NEG}$]) revealed that differential activation in the left angular gyrus / temporoparietal junction (xyz= -36 -69 21) was positively correlated with AX (r2= 0.47; p<0.001; Figure 5cd). A similar pattern was also observed in superior frontal cortex (xyz= -18 45 30).

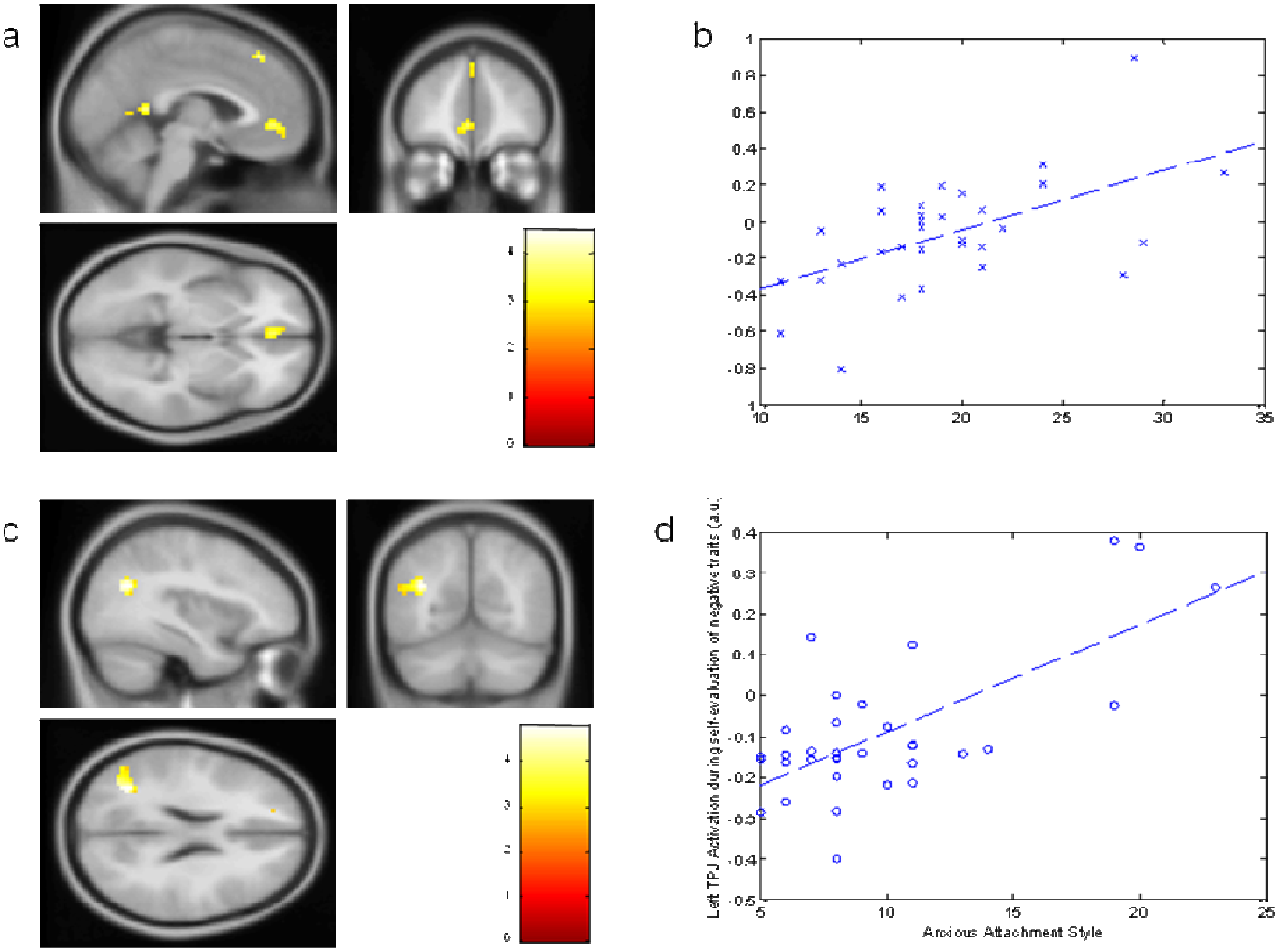


Figure 5: Correlations between raw betas and for positive self-evaluation and avoidant attachment and negative-self-evaluation and anxious attachment. *a) Statistical parametric map for the whole brain correlation [SELF$_{POS}$] x AV, displaying a positive effect in vmPFC. b) Correlation graph depicting a positive correlation between AV (x-axis) and the response in vmPFC (y-Axis). c) Statistical parametric map for the whole brain correlation [SELF$_{NEG}$] x AX, displaying a positive effect in left TPJ. b) Correlation graph depicting a positive correlation between AX (x-axis) and the differential response in left TPJ (y-Axis).*

DISCUSSION

The present fMRI study aimed at investigating the influence of different attachment working models (AWMs) entailing either *de-* (AV) or *hyper*-activating (AX) secondary attachment strategies on the attribution of positive and negative traits to oneself or a close other. The results showed that AV was associated with biases towards positive self-attributions in the behavioral measures as well as brain activity in particularly superior temporal Sulcus (STS), fusiform cortex (FC), and ventro-medial prefrontal cortex (vmPFC). AX did not display and influence on behavioral ratings, but revealed relations between brain activity in medial PFC (mPFC), orbito-frontal and superior frontal cortex (OFC and SFC), as well as temporo-parietal junction (TPJ) and the attribution of negative traits to others. These data represent the first investigation of neural correlates underlying different self- and other-perceptions as a function of adult attachment style (AAS).

Main effects of social evaluation in general and self- versus other-perception in particular

The main contrast of social evaluation *versus* syllable counting and the comparisons between trait-attribution to oneself *versus* others largely reproduced the findings of previous fMRI studies using the same experimental layout [15-17]. Thereby, activity in medial superior frontal, anterior and posterior cingulate, anterior temporal, as well as temporal parietal brain regions has been linked with theory of mind (ToM; [21-23]) and mentalizing about both self and others [19, 20], self- and other-knowledge representation [24, 25], as well as social decision making including moral reasoning [26, 27].

Correlations with avoidant attachment style

In all the experimental measures obtained in the present study (behavioral ratings; [SELF$_{(POS + NEG)}$ > CONTROL], [SELF$_{POS}$ > SELF$_{NEG}$], and [SELF$_{POS}$] fMRI contrasts), we observed a bias towards positive trait-attributions to oneself as a function of avoidant attachment style (AVS).

Contrast	Cluster peak coordinates (xyz; mm)	Peak T-value	Anatomical location
(Self+Other) > Word	+3 -57 +30	11.27	Precuneus (R)
	-12 +45 +45	10.16	Superior Frontal Gyrus (L)
	-48 +3 -39	9.94	Anterior Middle Temporal Gyrus (L)
	-45 -57 +21	8.05	Posterior Superior Temporal Gyrus (L)
	-45 +30 -18	7.09	Inferior Frontal Gyrus (L)
	-9 +15 +15	6.86	Caudate Nucleus (L)
	+51 +6 -33	6.50	Anterior Middle Temporal Gyrus (R)
	-39 +18 +45	5.78	Middle Frontal Gyrus (L)
	0 -15 +39	5.72	Cingulate Gyrus (B)
	+51 -60 +27	5.61	Posterior Superior Temporal Gyrus (R)
Word > (Self+Other)	-51 -57 -15	9.84	Inferior Temporal Gyrus (L)
	-54 -36 +39	9.74	Inferior Parietal Lobule (L)
	+45 -42 +48	9.12	Inferior Parietal Lobule (R)
	+48 +45 +6	7.81	Inferior Frontal Gyrus (R)
	-48 -6 +42	7.26	Precentral Gyrus (L)
	+51 +9 +9	5.93	Precentral Gyrus (R)

Contrast	Cluster peak coordinates (xyz; mm)	Peak T-value	Anatomical location
Self > Other	+15 +15 +57	5.82	Superior Frontal Gyrus (R)
	+6 +30 +21	5.06	Anterior Cingulate (R)
	45 -42 +39	5.01	Inferior Parietal Lobule (L)
Other > Self	+6 -54 +27	4.73	Retrosplenial Cingulate Gyrus (B)

Table 1: Activations for the main effects *(see Figure 2).*

The behavioral data indicated that, the higher participants scored on the AV subscale of the RSQ, the more likely they were to attribute positive trait-adjectives to themselves, which was reflected in a marginally significant relation between the latter two measures. Three subsequent whole-brain multiple regression analyses revealed that similar relations were present in the STS for the difference in activation between self-evaluations *versus* syllable counting, in the FC if contrasting positive *versus* negative self-attribution, and in the vmPFC when correlating raw betas for the positive self-evaluation condition with AV specifically. These findings suggest that avoidantly attached participants might have employed increased neural resources to 1) process perceived action or intentionality (STS; [24, 30]), 2) perform mental imagery by means of refreshing previously seen stimuli or visual word recognition in general (FC; [31, 32]), as well as 3) monitor possible outcomes in terms of reinforcement-guided decision making by context-sensitive stimulus evaluation (vmPFC; [33, 34]). Altogether, these behavioral and functional-anatomical findings suggest that avoidantly attached people associate positive trait-adjectives more likely with their own intentions and previously experienced events and/or outcomes of particular social situations.

Such a bias of positive trait-attributions to the self as a function of AV is well supported by the very exhaustive literature on this particular AAS (see e.g. [4] for a recent review). However, as already mentioned in the introduction, it is still debated whether this more positive self-perception represents a genuine self-positivity bias, or whether the latter is present as a consequence of positive self-inflation to distract avoidant people from their own weaknesses as well as the impact of negative social interactions with others. Even though the present fMRI study cannot provide any definitive answer to such considerations, the fact that we only found associations between AV and brain activity to positive self- and not negative-other attributions might suggest the observed self-positivity bias could represent a somewhat other-independent trait of avoidantly attached individuals.

However, by taking a second look at the fMRI data, we noticed that the anatomical location of the vmPFC cluster found for the restricted correlation of only the SELF_{POS} condition with AV (xyz= -6 48 -6) almost exactly corresponded to the anatomical location of brain activation found during an experimental task involving the emotional representation of social rejection (xyz= -6 49 -13; [35]). It might thus be hypothesized that the increased activity in vmPFC reflecting more pronounced monitoring of outcomes of positive self-attributions as a function of AV might be influenced by the emotional representation of previous negative interactions with others. Unfortunately, we did not find any such direct associations between AV and negative trait-attributions to others. Because negative other-perception is more often linked with AWMs of avoidantly attached individuals [4, 8, 9], and has already been shown in a previous fMRI study of the same authors to be linked with decreased reward-related activity during social positive interactions [14], similar effects could have been expected for the present study as well. However, the lack of such effects could be explained by the experimental design of this latest fMRI investigation, as the “others” were always close friends, who are normally associated with more positive than negative traits. To obtain any measure of more comprehensive other-representations including negative perceptions, further fMRI experiments including different “others” will be necessary.

Correlations with anxious attachment style

The present fMRI study revealed effects of AX on brain activity in mPFC (contrast [OTHER$_{(POS + NEG)}$ > CONTROL]), OFC (contrast [OTHER$_{POS}$ > OTHER$_{NEG}$]), as well as TPJ and SFG (contrast [SELF$_{NEG}$]). These data suggest that anxiously attached individuals recruited increased brain resources for mentalizing about others in general (mPFC; [36, 37]), tried to decrease the emotional involvement in negative other-evaluations by voluntary suppression or re-appraisal (OFC; [38, 39]), and had a stronger representation of agency for words representing negative self-attributes (TPJ; [40]).

The rather ventral location of the mPFC activation representing increased mentalizing about others as a function of AX (see above) is worth further consideration, because there is some evidence that there could be a ventral-dorsal dissociation of PFC activation during mentalizing about similar and dissimilar others [41], the more ventral part (xyz= 0 51 12 [our study xyz= 18 57 9]) being more activated for similar as compared to dissimilar other-representation. This is particularly interesting regarding AX, because the latter attachment orientation has been associated with an over-estimation of self-other similarity [42]. Consequently, anxiously attached participants in our study might have had stronger mPFC activity during trait-attribution (*versus* syllable counting) due to their tendency to evaluate themselves as more similar to others than secure or avoidantly attached people.

What is concerning the found OFC activation probably associated with increased emotion regulation efforts for negative (*versus* positive) other-evaluations (see above), such an interpretation is well in line with the attachment literature, characterizing AX with *hyper*-activating strategies entailing increased sensitivity to – particularly negative – external social clues [4]. Accordingly, when anxiously attached subjects had to attribute negative traits to a close friend (OTHER) in our study, they most likely had to regulate the emotional impact of the displayed adjectives associated with negative memories of events involving such concepts, because the close friend did not meet such representations. In other words, subjects most likely had to down-regulate their subjectively experienced negative emotions induced by reading negative trait-adjectives to more objectively attribute such traits to their close friends, who should have been associated with positive qualities.

Finally, the only areas of the present study displaying activation differences during self-attribution associated with AX were observed in TPJ and SFC, and more precisely during negative self-evaluations (contrast [SELF$_{NEG}$]). As AX is generally linked with a negative self-model entailing low self-esteem [4], but the participants in our study showed a general bias to attribute more positive than negative traits to the self, the SFC might have been more activated in anxiously attached subjects due to increased emotion regulation attempts with the aim of decreasing the negative impact of such trait-descriptions. Previous fMRI results have demonstrated the involvement of SFC in both the *in*- and *de*-crease of positive as well as negative emotions [43], and thus emotion regulation in general. What is concerning the lateral temporal activations (TPJ), it might be speculated that AX was associated with a stronger sense of agency linked with negative trait-attributes, for the same reasons as mentioned just above (negative self-model).

CONCLUSIONS

The present fMRI study represents the first functional-anatomical investigation of the distinct self- and other-representations underlying avoidant (AV) and anxious (AX) adult attachment styles.

Our data indicates that AV was generally associated with a bias towards positive trait attributions to the self, as well as with increased brain activity in regions involved in the representation of intentionality (STS), visual imagery (FC), as well as outcome monitoring (vmPFC). Whereas the behavioral data and the first two above-mentioned brain activations point towards a rather self-specific positive evaluation of the own person (positive self-model), vmPFC recruitment could also be evidence of a link to negative other-perceptions due to previous accounts relating activity in this particular brain area with emotional conflict experienced during social rejection. However, more data is needed on the negative other-model as a hallmark of AV, requiring a different experimental design (other should not be a close friend) to tackle this issue.

Our findings regarding AX suggest that this attachment orientation might indeed entail a more pronounced subjective similarity between oneself and others, because of the rather ventral location of the observed mPFC activity during social-emotional trait attribution overall. In addition, the general bias to attribute more positive traits to oneself and others in combination with increased sensitivity to negative emotional clues as a hallmark of AX, heightened activity in OFC and SFC during negative trait-attribution to others and oneself might represent more pronounced emotion regulation attempts to diminish the impact of negative emotions. Finally, TPJ activity during negative self-evaluation could imply a more prominent sense of agency, associated with the generally described negative self-model as a function of AX.

Altogether, these data indicate that adult attachment style does have an effect on brain activity underlying self- and other-evaluation, even though the uncovering of the exact relations still need further investigation.

REFERENCES

1. Ainsworth, M.D.S., *Bowlby-Ainsworth Attachment Theory.* Behavioral and Brain Sciences, 1978. 1(3): p. 436-438.
2. Bowlby, J., *Attachment and loss*. 1969, New York,: Basic Books. v.
3. Bowlby, J., *Attachment and loss*. 2nd ed. 1982, New York: Basic Books. v.
4. Mikulincer, M. and P.R. Shaver, *Attachment in Adulthood: Structure, Dynamics, and Change*. 2007, New York: The Guilford Press.
5. Hazan, C. and P. Shaver, *Romantic Love Conceptualized as an Attachment Process.* Journal of Personality and Social Psychology, 1987. 52(3): p. 511-524.
6. Hazan, C. and P.R. Shaver, *Love and Work - an Attachment-Theoretical Perspective.* Journal of Personality and Social Psychology, 1990. 59(2): p. 270-280.
7. Bartholomew, K., *Avoidance of Intimacy - an Attachment Perspective.* Journal of Social and Personal Relationships, 1990. 7(2): p. 147-178.
8. Bartholomew, K. and L.M. Horowitz, *Attachment Styles among Young-Adults - a Test of a 4-Category Model.* Journal of Personality and Social Psychology, 1991. 61(2): p. 226-244.
9. Griffin, D. and K. Bartholomew, *Models of the Self and Other - Fundamental Dimensions Underlying Measures of*

Adult Attachment. Journal of Personality and Social Psychology, 1994. 67(3): p. 430-445.
10. Mikulincer, M., P.R. Shaver, and D. Pereg, *Attachment theory and affect regulation: The dynamics, development, and cognitive consequences of attachment-related strategies.* Motivation and Emotion, 2003. 27(2): p. 77-102.
11. Kurdek, L.A., *On being insecure about the assessment of attachment styles.* Journal of Social and Personal Relationships, 2002. 19(6): p. 811-834.
12. Simpson, J.A., *Influence of Attachment Styles on Romantic Relationships.* Journal of Personality and Social Psychology, 1990. 59(5): p. 971-980.
13. Simpson, J.A., W.S. Rholes, and J.S. Nelligan, *Support Seeking and Support Giving within Couples in an Anxiety-Provoking Situation - the Role of Attachment Styles.* Journal of Personality and Social Psychology, 1992. 62(3): p. 434-446.
14. Vrtička, P., et al., *Individual attachment style modulates human amygdala and striatum activation during social appraisal.* PLoS ONE, 2008. 3(8): p. e2868.
15. Heatherton, T.F., et al., *Medial prefrontal activity differentiates self from others.* Soc Cogn Affect Neurosci, 2006. I: p. 18-25.
16. Seger, C.A., M. Stone, and J.P. Keenan, *Cortical activations during judgments about the self and an other person.* Neuropsychologia, 2004. 42(9): p. 1168-1177.
17. Schmitz, T.W., T.N. Kawahara-Baccus, and S.C. Johnson, *Metacognitive evaluation, self-relevance, and the right prefrontal cortex.* Neuroimage, 2004. 22(2): p. 941-947.
18. Ochsner, K.N., et al., *The neural correlates of direct and reflected self-knowledge.* Neuroimage, 2005. 28(4): p. 797-814.
19. Lombardo, M.V., et al., *Shared Neural Circuits for Mentalizing about the Self and Others.* J Cogn Neurosci, 2009.
20. Yaoi, K., N. Osaka, and M. Osaka, *Is the self special in the dorsomedial prefrontal cortex? An fMRI study.* Soc Neurosci, 2009. 8: p. 1-9.
21. Frith, C.D. and U. Frith, *Cognitive psychology - Interacting minds - A biological basis.* Science, 1999. 286(5445): p. 1692-1695.
22. Frith, U. and C.D. Frith, *Development and neurophysiology of mentalizing.* Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 2003. 358(1431): p. 459-473.
23. Gallagher, H.L. and C.D. Frith, *Functional imaging of 'theory of mind'.* Trends in Cognitive Sciences, 2003. 7(2): p. 77-83.
24. Gobbini, M.I. and J.V. Haxby, *Neural systems for recognition of familiar faces.* Neuropsychologia, 2007. 45(1): p. 32-41.
25. Lieberman, M.D., *Social cognitive neuroscience: A review of core processes.* Annual Review of Psychology, 2007. 58: p. 259-289.
26. Greene, J.D., et al., *The neural bases of cognitive conflict and control in moral judgment.* Neuron, 2004. 44(2): p. 389-400.
27. Greene, J.D., et al., *An fMRI investigation of emotional engagement in moral judgment.* Science, 2001. 293(5537): p. 2105-2108.
28. Rosenberg, M.J., *Society and the adolescent self-image*. 1965, Princeton: Princeton University Press.
29. Anderson, N.H., *Likableness Ratings of 555 Personality-Trait Words.* Journal of Personality and Social Psychology, 1968. 9(3): p. 272-&.
30. Allison, T., A. Puce, and G. McCarthy, *Social perception from visual cues: role of the STS region.* Trends in Cognitive Sciences, 2000. 4(7): p. 267-278.
31. Johnson, M.R., et al., *A brief thought can modulate activity in extrastriate visual areas: Top-down effects of refreshing just-seen visual stimuli.* Neuroimage, 2007. 37(1): p. 290-299.
32. Hauk, O., M.H. Davis, and F. Pulvermuller, *Modulation of brain activity by multiple lexical and word form variables in visual word recognition: A parametric fMRI study.* Neuroimage, 2008. 42(3): p. 1185-1195.
33. Amodio, D.M. and C.D. Frith, *Meeting of minds: the medial frontal cortex and social cognition.* Nature Reviews Neuroscience, 2006. 7(4): p. 268-277.
34. Rushworth, M.F.S., et al., *Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour.* Trends in Cognitive Sciences, 2007. 11(4): p. 168-176.
35. Somerville, L.H., T.F. Heatherton, and W.M. Kelley, *Anterior cingulate cortex responds differentially to expectancy violation and social rejection.* Nature Neuroscience, 2006. 9(8): p. 1007-1008.
36. Mitchell, J.P., M.R. Banaji, and C.N. Macrae, *General and specific contributions of the medial prefrontal cortex to knowledge about mental states.* Neuroimage, 2005. 28(4): p. 757-762.
37. Mitchell, J.P., C.N. Macrae, and M.R. Banaji, *Forming impressions of people versus inanimate objects: Social-cognitive processing in the medial prefrontal cortex.* Neuroimage, 2005. 26(1): p. 251-257.
38. Levesque, J., et al., *Neural circuitry underlying voluntary suppression of sadness.* Biological Psychiatry, 2003. 53(6): p. 502-510.
39. Ochsner, K.N., et al., *Rethinking feelings: An fMRI study of the cognitive regulation of emotion.* Journal of Cognitive Neuroscience, 2002. 14(8): p. 1215-1229.
40. Spengler, S., D.Y. von Cramon, and M. Brass, *Was it me or was it you? How the sense of agency originates from ideomotor learning revealed by fMRI.* Neuroimage, 2009. 46(1): p. 290-298.
41. Mitchell, J.P., C.N. Macrae, and M.R. Banaji, *Dissociable medial prefrontal contributions to judgments of similar and dissimilar others.* Neuron, 2006. 50(4): p. 655-663.
42. Mikulincer, M., I. Orbach, and D. Iavnieli, *Adult attachment style and affect regulation: Strategic variations in subjective self-other similarity.* Journal of Personality and Social Psychology, 1998. 75(2): p. 436-448.
43. Kim, S.H. and S. Hamann, *Neural correlates of positive and negative emotion regulation.* Journal of Cognitive Neuroscience, 2007. 19(5): p. 776-798.

VI.7. Selectivity of Results regarding Adult Attachment Style

As adult attachment style (AAS) has not yet been widely used in fMRI research regarding social perception, the general validity of applying self-report questionnaires in adults is still debated [201], and particularly the anxiety (AX) subscale is often related to other, more general measures of anxiety or neuroticism (see e.g. [194]), the employment of AAS in three different fMRI paradigms withinhe present thesis was also intended to probe the strength of this psychological concept in social neuroscience. Consequently, a final question arose:

- **How specifically does AAS influence the neural correlates of social perception?**

Because this question can only be answered in a comprehensive way by considering the results of all three fMRI studies within the present thesis, it will be addressed in the synthesis part of the general discussion (see Chapter VII.3).

CHAPTER VII

General Discussion & Conclusions

The new research domain of social neuroscience aims at elucidating the biological mechanisms underlying human social behavior by integrating biological, neuro-physiological, as well as social psychological concepts (see Chapters I.1 and I.2). Accordingly, the experiments conducted in the scope of the present thesis took into consideration both functional anatomical methods – by means of fMRI – as well as social psychology accounts – reflecting the influence of social context and individual differences – to reveal so far unknown relations between personality, context, behavior, emotion and cognition.

For that purpose, two core processes of social neuroscience were chosen as central research areas of interest, namely *emotion perception* (see Chapter I.3.1.) and *emotion regulation* (see Chapter I.3.2.). Within these research areas, the modulation of brain activity by two different methods of *social context manipulation* (see Chapter IV. 3) and *individual differences* (see Chapter IV.4.) was probed.

Below, the findings of the three fMRI and the behavioral study are discussed in a more comprehensive manner, first assessing the role of social context (Chapter VII.1.) and then the influence of individual differences – and in particular adult attachment style (AAS; Chapter VII.2.) – on emotion perception and regulation, respectively. These two sections are then followed by a general synthesis chapter, aiming at integrating all the accounts made beforehand (Chapter VII.3.).

Importantly, it should be noted here that, throughout the present doctoral thesis and also the first part of the general discussion section, the distinction between external (social context) *versus* internal (individual differences) context on one hand, and emotion perception *versus* regulation on the other hand, is used for an illustrative purpose. This is, however, not tantamount with a complete separation or lack of interaction between the two respective factors. For example, individual differences might not only modulate subjective experience of and corresponding brain activity during different social situations (as shown in fMRI study N° 1b), but already influence the decision whether to engage in a certain social situation at a first place – if such an option is available. Similarly, emotion regulation – and particularly re-appraisal – might already influence the way how emotions are perceived, again on both the level of subjective experience and brain activity. This notion of interconnectedness will be picked up later in the general discussion.

VII.1.Social Context

The first main goal of the present doctoral thesis was to further extend the knowledge about the influence of social contextual mechanisms on emotion perception and regulation and their corresponding neural correlates. For this purpose, context manipulation was carried out in two different ways. On one hand, the social meaning of happy (smiling) and angry faces was directly modulated by pairing the latter with different positive or negative personally relevant game-feedback, thereby creating a social connotation of either “friends” or “foes”. These notions of affiliation or hostility were specifically employed to study emotion perception (fMRI studies N° 1ab). On the other hand, social context was manipulated by including either social or non-social images into the experimental design, thereby directly comparing basic stimulus properties. This second social context manipulation approach was employed in both the behavioral and fMRI studies N° 2ab particularly designed to address emotion regulation, but some conclusions can nonetheless be drawn regarding emotion perception as well. Because fMRI study N° 3 did not include any social context manipulation, it will not be further considered here. The influence of social context on neural correlates of social perception will first be discussed for emotion perception (Chapter VI.1.1.), and then for emotion regulation (Chapter VI.1.2.), before being integrated as a whole (Chapter VI.1.3.).

VII.1.1. Emotion Perception

1.1.1. Obtained Results

fMRI studies N° 1ab revealed that the encoding (study N° 1a) and recognition (study N° 1b) of emotional facial expressions – in this case smiling (happy) and angry faces – did not only depend on their intrinsic positive or negative value, but was crucially modulated by their concomitant personally relevant social meaning. These findings therefore importantly extend previous results describing only *im*personal social contextual effects on 1) ambiguous facial expressions (i.e., surprise) perceived in a negative social context (i.e., “*he just lost $500*”) [41], 2) neutral faces seen in a positive social context (i.e., intentional co-operators) [154], or 3) neutral faces associated with a negative (i.e., disgust or anger) social context [198]. Moreover, the new results show that once established, such a specific social meaning of a given face, or, in other words, a first impression of a particular person, 1) leaves an unconscious emotional memory trace in the human brain, which is re-activated when this person is encountered again at a later stage, even if with a different emotional facial expression and/or intention, and 2) that first impressions of “foes” leave stronger marks than first impressions of “friends”.

The **behavioral study** showed effects of social context – manipulated by the distinction between basic stimulus properties by means of social *versus* non-social emotional image content – regarding the subjective experience of intensity (arousal) and controllability, but not pleasantness (valence). More precisely, whereas social and non-social scenes were subjectively experienced as equally (un-)pleasant, social images were judged as more arousing than their non-social counterparts – but only for positive valence –, and social images were rated as less controllable in general. These findings, even though not revealing particularly new relationships between intrinsic stimulus properties and evoked corresponding subjective experience, provided important information concerning the experimental setup of fMRI studies N° 2a & b in terms of stimulus selection.

Finally, **fMRI studies N° 2ab** very nicely illustrated in general that different brain areas are implicated in the encoding of either valence (related to arousal), or, more crucially, social *versus* non-social image content representation (see Figures 1 and 2 in supplements of fMRI study N° 2a).

Whereas positive (*versus* negative) stimulus encoding only showed differential activity in subcallosal ACC and primary somatosensory cortex, the inverse contrast revealed differential processing of negative (and highly arousing) information in dorso-medial and dorso-lateral PFC, ventro-lateral OFC, bilateral insula, and more medial visual areas. Such an activation pattern might therefore correspond to the employment of cognitive control as well as the representation of bodily arousal in relation to highly negative stimulus content in general (see Chapter III).

More interestingly, the encoding of non-social *versus* social stimuli also revealed activity in lateral and ventro-lateral PFC probably linked to cognitive control, and additional activations in middle ACC most likely related to the encoding of pain in particular (the middle ACC is part of the “pain matrix”; see Chapter III). This might originate from the fact that non-social (as compared to social) pictures included scenes directly inflicting pain or scenes of natural / man-made disasters easily associated with pain. Lastly, the social *versus* non-social contrast revealed increased activations in medial PFC and OFC, inferior temporal lobes, posterior cingulate cortex, posterior superior temporal sulcus, the fusiform gyrus, as well as bilateral amygdala. Many of these areas are thought to be primarily involved in mentalizing and theory of mind (see Chapter III), functions that are apparently more strongly required when appraising social (as opposed to non-social) stimuli. Of particular interest was the finding that activity in bilateral amygdala did not encode for valence (or arousal) in this experiment, but for the difference between social *versus* non-social stimulus content, even though these two stimulus categories did not display any intrinsic stimulus property differences (matched on pleasantness [valence], arousal [intensity], and controllability). In addition, because the subject’s task was to naturally experience, cognitively down-regulate, or behaviourally inhibit emotion experience induced by viewing both social as well as non-social emotional pictures, no explicit task instructions were given to consider social stimuli as in any regard different from non-social stimuli. This suggests that such a social context effect in the amygdala had to arise from the computational properties of this brain area itself (see Chapter VII.1.1.2. below for more details).

1.1.2. Specific Reconsiderations

By re-considering the activation patterns revealed by the different methods of context manipulation in fMRI studies N° 1ab and N° 2ab more comprehensively, some particularities were observed.

(1) Firstly, in both parts of fMRI study N° 1, the neural substrates mediating the found social context effects were identified in the amygdala, different regions of the basal ganglia, and the anterior cingulate cortex (ACC). However, there were two intriguing differences in activation patterns during encoding (study N° 1a) as compared to recognition (study N° 1b).

On one hand, in the amygdala, there was a selective activation difference between angry *versus* happy faces seen while losing (AF-L *versus* SF-L) during the encoding phase (fMRI study N° 1a, Figure 3ab), but an incongruent *versus* new effect in general during the recognition phase (fMRI study N° 1b, Figure 5ab). Similarly, activity in the anterior cingulate cortex (ACC) showed increased BOLD signal change for the AF-W condition during both encoding (fMRI study N° 1a, mentioned in the text) and recognition, but an additional effect for the SF-L condition during recognition (fMRI study N° 1b, Figure 6cd). These two activation patterns thus suggest that the amygdala and ACC represented the SF-L “foe” condition only during the recognition phase, raising questions regarding its processing in those two areas during the encoding phase.

Yet, such apparent discrepancies can be easily explained in terms of a simple baseline effect. In fMRI study N° 1a, amygdala activation is depicted in Figure 3b only including the four different feedback conditions, implying a (relative) de-activation for faces seen as “foes” while losing (SF-L). Yet, if the same activation is illustrated including the control-condition (dot-display), it is apparent that there was also amygdala activation for the SF-L condition, even though the latter was obviously lower than for the AF-L condition (see Figure M15 below) The same holds true for the ACC activation (contrast AF-W > SF-W) reported in fMRI study N° 1a, where activity to the SF-L condition seems to be “negative”, but only relative to the AF-W and not the baseline (dot-display) condition. This means that both incongruent feedback conditions elicited activity in the amygdala and ACC during encoding, and that these notions were re-activated during the recognition task.

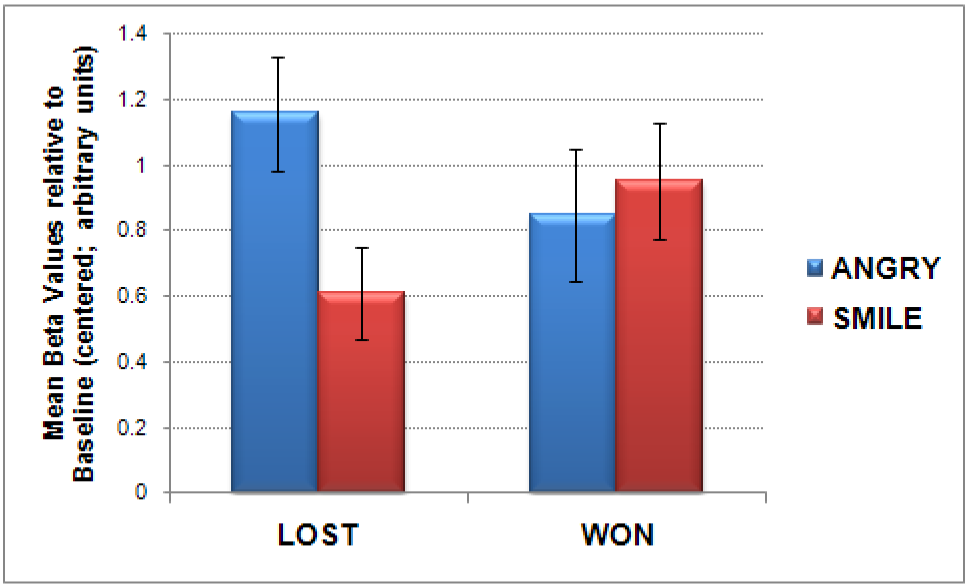


Figure M15: Activation Pattern in left dorso-medial amygdala for the contrast AF-L > SF-L during the encoding phase (fMRI study N°1a) relative to the dot-display (baseline). *See Text for more details.*

On the other hand, and similar to the account made above regarding amygdala activation, different areas of the basal ganglia displayed a general activation difference between wining *versus* losing or a selective activation for smiling (happy) faces seen while winning (SF-W) during encoding (fMRI study N° 1a, Figure 2ab), whereas there was a selective effect for angry faces seen while winning (AF-W; as compared to AF-L) during recognition in the caudate nucleus (fMRI study N° 1b, Figure 6ab). The latter two activation patterns thus imply that activity in the basal ganglia during recognition was more specific or even inversed as compared to encoding. However, the selective encoding effect for SF-W faces was only found in one particular region of the basal ganglia, a result that should be regarded somewhat separately from the others. All other reward-related brain areas displayed a more basic won *versus* lost activation pattern during encoding, which was also present in the caudate nucleus during recognition [t(15)= 2.96, p< .01; paired t-test]. General reward-related activity was thus preserved during both experimental tasks in the basal ganglia.

Taken together, the abovementioned considerations show that, if including the dot-task as baseline condition, all three regions of interest represented the same stimulus properties during encoding as well as during recognition. In other words, no information that was present during encoding was "lost" during recognition, and all information that was represented during recognition has already been present during encoding. However, it still remains striking that all three regions of interest displayed task-specific differences in activity, meaning that some stimulus properties were represented preferentially during either encoding or recognition. For example, in the amygdala, there was a specific representation of social punishment (AF-L) during encoding, whereas during recognition, a general incongruent *versus* new – but no congruent *versus* new – effect could be observed. This was most likely due to the different task instructions and the thereby resulting distinct goals of the subjects during encoding as compared to recognition.

During encoding, the subjects were eager to “win” as many trials as possible. Accordingly, certain emotion expression x feedback associations were more relevant for them than others, and were thus preferentially encoded in the three regions of interest. More precisely, the amygdala – importantly (but not exclusively) involved in the processing of negative emotions and threats (see Chapter III.2.1.) – particularly retained angry faces seen while losing, because this condition represented the most aversive social feedback, namely social punishment (negative responses from “friends” if the subjects already felt bad about a personal failure). In turn, the ventral striatum – being part of the basal ganglia and thus the reward system (see chapter III.2.2.) – encoded for the fact of having “won” as positive experience in general, and one particular anterior region (see Figure 3a in fMRI study N° 1a) the notion of social reward (SF-W condition) in particular, because the latter was the most positive condition in the experiment (positive responses from “friends” if the subjects already felt good about their own success). Finally, the ACC – previously found to be activated by conflict in general, and social rejection more specifically (see Chapters III.2.4 and 3.3.) – retained the most obstructive social context condition to “winning” (AF-W) representing mockery by “foes”.

In turn, during recognition, the three regions of interest retained the most important social context conditions presented during encoding according to their evolutionary function, namely information about “foes” that could be crucial for survival during a second encounter. The amygdala signaled increased vigilance and the presence of a potential threat by selective heightened activity for “foes” (but not “friends”) *versus* new faces. The ACC complemented this notion by retaining information about social rejection by both categories representing “foes”. And lastly, the caudate nucleus also encoded for a previous experience with “foes”, but this time in a positive sense in terms of signaling an increase in social status due to an “outperformance” of somebody else (AF-W).

To sum up the point about differential activation of amygdala, basal ganglia and ACC during encoding as compared to recognition, our data indicates that there were no discrepancies between stimulus properties represented *per se*, but more likely a task effect, rendering some information more important while subjects played the pseudo-interactive game, and other information more relevant while they performed the subsequent surprise memory task. Such findings thus speak in favor of the component process / appraisal model of emotion (see Figure M2), emphasizing goal / need relevance and conduciveness as crucial factors in emotion processing.

(2) Second, and related to the first point mentioned above, activity in the amygdala in particular during fMRI studies N° 1ab and 2ab displayed activation patterns that clearly imply more elaborated computation properties than simple valence or arousal processing. Accordingly, amygdala activity in general can not only be nicely matched with predictions from the component process / appraisal model of emotions [37], but also with the notion of the amygdala being mainly a "relevance detector" [69].

Most important during *fMRI study N° 1a* was the fact that activity in the (left) amygdala did not display any main effect of expression [F(1,15)= .761, p= .397], but a highly significant expression x feedback interaction [F(1,15)= 8.19, p= .012] (see Figure 3ab). This implies that emotional faces *per se* did not lead to increased BOLD signal change in the amygdala, as could have been expected by predictions from basic emotion – "angry > smile" because of preferred negative emotion processing – or bi-dimensional - "angry > smile" due to higher arousal properties of negative stimuli – theories (see Chapter I.3.1.1. and Figure M1). Instead, the amygdala integrated both expression *and* feedback, and thus performed a basic evaluation / appraisal of all the present stimulus conditions. In addition, because the interaction effect was driven by a significant activation difference between angry *versus* smiling faces during losing (AF-L > SF-L), it could be argued that the AF-L condition was more relevant – more goal-obstructive –, since representing negative feedback from a "friend" as opposed to negative feedback from a "foe" (SF-L).

Regarding *fMRI study N° 1b*, it was not only very interesting to observe a general "old" incongruent > new – and no old congruent > new – effect in bilateral amygdala (see Figure 5ab), but even more striking that the amygdalae were activated by neutral faces at the first place. This comes from the fact that neutral faces should not contain any (negative) emotional information (basic emotion theory) or induce the perception of any valence or arousal property (bi-dimensional models), leading to amygdala activation, but that this effect should be attributed to the increased relevance of some neutral faces as compared to others. Even though in our case, the latter argument is not as strong as the finding that the amygdala was activated by neutral faces *as such* during identity matching [117], where no emotional information was processed at all and the relevance of neutral faces came from the task alone, it still points into the same direction. If our data represented a general emotional memory effect, the amygdalae should have been found activated more strongly for all "old" as compared to "new" faces. Similarly, if our findings were due to a more specific negative emotional memory effect, the amygdalae should also have displayed increased activity for the "old" AF-L condition as compared to "new" faces as well, which they did not. Our results therefore speak more in favor of the relevance detector account [69], rather than a simple (emotional) memory effect.

Finally, in *fMRI study N° 2a*, we found bilateral amygdala activation (irrespective of task instructions, averaged over all trials) only for the contrast opposing social *versus* non-social, but not negative *versus* positive scene content. This was even more remarkable when considering that social and non-social stimuli were matched on pleasantness (valence), arousal (intensity), and controllability properties, whereas negative images were in average rated as more intense and less controllable than positive images in a validation study (behavioral study) prior to fMRI study N° 2a(b). This finding thus once more demonstrates that the amygdala does not encode for intrinsic stimulus properties like valence, intensity, or control *per se*, but integrates these factors with other kinds of stimulus evaluation processes. Other authors [157] have already argued that such a preferential response of the amygdala to faces / social stimuli as opposed to scenes / non-social stimuli *"… reflects the intrinsic importance and usefulness of facial expressions in directing our behavior and regulating our social interactions…"*.

In summary, our complete data regarding contextual effects on amygdala activation during emotion perception clearly indicates that the functional anatomical properties of this brain region speak in favor of appraisal theories of emotion and the role of the amygdala as a relevance detector in particular within this theoretical framework.

Regarding the interpretation of general amygdala activation during fMRI study N° 2a(b) – averaged over all trials –, it should additionally be noted here that we found a (marginally) significant scene content x valence interaction [left: F(1,18)= 3.95, .062; right: F(1,18)= 10.24, .005] bilaterally, and that this interaction was driven by a selective activation difference between negative *versus* positive images for non-social scene content (see Figure M16 below). In other words, whereas the level of amygdala activation to social stimuli was roughly identical regardless of valence, there was a significant difference in activity regarding valence for non-social stimuli (negative > positive).

From the point of view of an appraisal theorist arguing in favor of the opinion that the amygdala is a relevance detector, this activation pattern would suggest that, because social stimuli in general are intrinsically more relevant, both negative as well as positive images are processed in the same way, as they both contain significant information for directing behavior and regulating social interactions (see e.g. [157, 202, 203], the latter two studies showing that positive social stimuli [baby faces] also have a high biological relevance and readily capture attention). In turn, because non-social scenes do not exhibit the same intrinsic relevance properties, they are appraised differently, e.g. according to their significance in terms of survival value, which is intrinsically higher for negative (i.e. threatening) as opposed to positive (i.e. appetitive) non-social content.

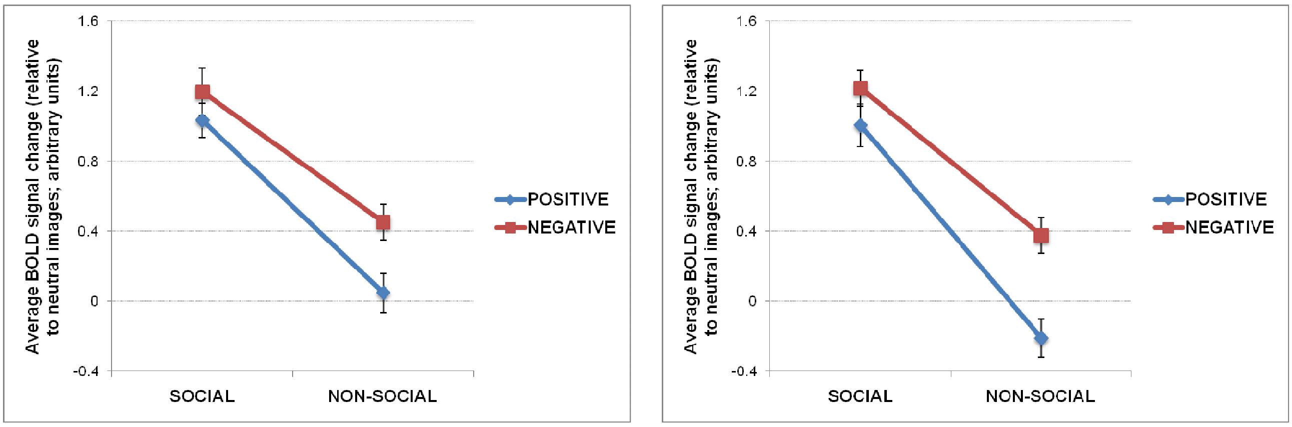


Figure M16: Social content x valence interaction graphs in left (top) and right (bottom) amygdala during fMRI study N°2a(b) . Extracted beta values were averages across all experimental conditions and all voxels in the respective clusters. Paired t-tests show that there is no significant activation difference between social negative *versus* positive images [left: t(1,18)= 1.23, p= .28; right: t(1,18)= 2.38, p= .14], but significant activation differences between non-social negative *versus* positive images [left: t(1,18)= 7.97, p= .001; right: t(1,18)= 18.96, p< .001]. All statistical tests are paired t-tests.

Unfortunately, such an interpretation cannot be proved definitively only based on the presented data. This comes from the fact that there could still be an arousal / intensity effect for social images in the amygdala, as the BOLD signal for social negative scenes could be affected by a signal saturation effect. Therefore, before drawing any final conclusions on this scene content x valence interaction, amygdala activation should be re-considered once more, preferably with the help of a parametric modulation analysis. By adding a parametric modulator representing pleasantness (valence) and intensity (arousal) ratings of each individual image, it should be possible to determine whether the BOLD signal in bilateral amygdala was modulated by pleasantness and/or intensity, and if yes, whether this was only the case for non-social or all image categories in general.

1.1.3. Specific Limitations

Besides the rather specific limitation already mentioned above regarding a possible arousal / intensity effect in bilateral amygdala for activity to SOC images in *fMRI study N°2a* , the second method of context manipulation (dissociation of basic stimulus properties according to the presence or absence of social clues) could also have some limiting factors more generally.

The probably most prominent factor refers to possible differences in the *social complexity* of images. Whereas this could be controlled for by including images with the same amount of depicted individuals for social positive *versus* social negative images, the same procedure was apparently not possible for the comparison between social *versus* non-social scenes in general. It could thus be argued that all effects reported for the latter contrast(s) in fMRI study N° 2a were due to basic differences in social complexity representation rather than caused by the differential engagement of more elaborated stimulus processing, involving e.g. mentalizing, theory of mind, or relevance detection. And indeed, there is some evidence for the fact that activity in posterior superior temporal sulcus (pSTS) and lateral fusiform gyrus (lFG) could have been driven by the simple presence of people / faces as compared to objects, and thus higher social complexity *per se*. This evidence comes from previous experiments showing that the lFG – including the fusiform face and body areas (FFA and FBA) – is generally more activated for stimuli depicting faces (FFA) or bodies (FBA) [147, 150, 151] on one hand, and that the pSTS (including the temporo-parietal junction; TPJ) shows increased BOLD signal change to scenarios containing any kind of person (as compared to object) information [204], on the other hand. However, the latter study [204] also showed that medial PFC, anterior temporal pole (ATP) and posterior cingulate cortex (PCC) activation emerged by contrasting mental state (intentional) with person-space or object-space (un-intentional) attribution, and was thus representing mentalizing and theory of mind. Finally, amygdala activation has already been described above to most likely represent relevance detection of intrinsically more salient social versus non-social stimuli due to their "... *usefulness … in directing our behavior and regulating our social interactions…"* [157]. Taken all together, even though some brain areas (pSTS / TPJ and lFG) might have been more responsive to basic differences in social complexity when contrasting social *versus* non-social stimuli, activity detected in other brain areas more crucially involved in social perception (mPFC, ATP, PCC, amygdala) most likely represented more elaborated social perception processes.

Some limiting factors could also have been present what is concerning the first method of context manipulation, namely the direct modulation of the social meaning of emotional facial expressions by personally relevant game feedback (fMRI studies N° 1ab).

One could for example argue that the mere absence of any main effects of expression (smiling or angry) in *fMRI study N° 1a* – supporting the conclusion that the brain does not encode expression *per se*, but integrate the latter with concomitant game feedback to derive a socially meaningful representation – could have been a result of the fMRI data analysis, which always directly compared the respective two conditions, but never contrasted the latter to a "neutral" baseline. Along these lines, it might have been possible that e.g. some brain areas responded more to angry faces than neutral faces (or even a fixation cross), but that this effect was cancelled out by contrasting them to smiling faces. Yet, even though such an argumentation might be justified, it raises important questions about the specificity of thereby possibly observed activation patterns.

More precisely, if activity in a certain brain area to e.g. smiling faces were not detectable through a direct contrast “smiling > angry”, this would most likely be due to the fact that the same brain area would also be activated to angry faces. One would thus find the respective region for the contrast “smiling > neutral”, but also for the contrast “angry > neutral”. As a result, it could not be concluded, either, that this particular brain area shows a specific BOLD signal change increase for any of the two expressions, but only that it responds to emotional content in general. Therefore, if a main effect of emotion *versus* “neutral” (or perception on general) is not crucial for interpretational purposes, a direct comparison of two emotional conditions is perfectly feasible.

Regarding *fMRI study N° 1b* , the claim could be made that with the employed data analysis procedure, we did not really measure neural correlates of familiar face recognition, as we based our arguments on the contrast “old correct” *versus* “new correct” and not the contrast “hits” *versus* “misses” usually applied in memory paradigms. This, however, did not represent a conceptual error in the study design, but an intentional specification in data analysis. On one hand, by computing the contrast “hits” *versus* “misses”, only explicit memory formation would have been assessed, which was not the main focus of fMRI study N° 1b, nor a procedure supported by the behavioral findings, implying the absence of any explicit memory for the specific feedback x expression interactions. In contrast, by telling subjects to perform an “old correct” *versus* “new correct” differentiation, we also accounted for implicit person recognition mechanisms, but still limited our analysis to correct answers and thus not random assignments to the respective categories. On the other hand, we even think that we increased the specificity of data analysis by applying the contrast “old correct” *versus* “new correct”. This is due to the fact that both “old hits” and “old misses” could have included some familiarity information which would have cancelled each other out (see also above for the same argumentation), because each face identity (“new” and “old” faces) was shown repeatedly (six times) during recognition. Consequently, even “new” faces could have acquired some sense of familiarity throughout the memory task. This would not have been accounted for by using the contrast “hits” *versus* “misses” – as “hits” also include “new” faces –, but was perfectly controlled for by our procedure, only assessing familiarity acquired during the preceding game (“old correct”). Taken all together, we can thus say that our memory task was specifically aimed at revealing neural correlates of 1) both explicit *and* implicit face recognition, and 2) acquired familiarity through preceding social interactions and not only mere exposure effects. Finally, the employment of the contrast “old correct” *versus* “new correct” can also explain the absence of any hippocampus activation as a standard marker of memory, because “new” faces could also have activated the latter region due to their repeated, yet unspecific (no association with specific social scenarios) presentation during recognition.

VII.1.2. Emotion Regulation

1.2.1. Obtained Results

The only fMRI study within this present doctoral thesis specifically designed to investigate the impact of emotion regulation (ER) on brain activity modulated by the exposure to either social or non-social positive or negative scenes (second method of context manipulation) – **fMRI study N°2a** – revealed several interesting and new relations.

Firstly, the direct comparison between re-appraisal (REAP) and suppression (SUP) led to the discovery of three clusters in prefrontal cortex where activity was specifically increased for either of the two ER strategies, independent of any scene content (SC; social [SOC] *versus* non-social [NSOC]) or valence (VAL; positive [POS] *versus* negative [NEG]) factors. This implies that those brain areas were involved in ER more generally, such as in the engagement of working memory to keep track of the re-appraisal strategy used or the ultimate goal of the ER attempts, and in volitional inhibition of emotion expression trough cognitive control [86, 87, 90, 140].

Second, the paired t-test models specifically traced down such brain areas where the differential response to NEG *versus* POS or SOC *versus* NSOC (and vice versa) emotional scenes was distinctively modulated by either REAP or SUP (as compared to “natural” emotion experience [EMEXP]). For negative scenes, REAP was found to selectively reduce activity in dorso-medial PFC, whereas both REAP and SUP had such an effect in anterior insula and medial fusiform gyrus (FG). In turn, activity to NSOC scenarios was specifically reduced during SUP in middle insula, and BOLD signal change increase to SOC images was distinctively decreased during REAP in medial PFC, posterior cingulate (PCC) and lateral FG. Finally, and most interestingly, increased activity to SOC (*versus* NSOC) scenes in bilateral amygdala displayed an intriguing lateralization effect, being reduced more during REAP in the left and more during SUP in the right amygdala.

Altogether, these findings importantly extend the already present literature on ER by not only showing for the first time distinct effects of REAP and SUP on the modulation of emotional responses to images as a function of both VAL and SC, but also direct evidence for a long-debated but not yet well characterized lateralization effect in the amygdala [86, 87, 205, 206].

Because the aim of this chapter is to discuss results regarding the manipulation of social context, only effects of ER on the comparison between SOC and NSOC images (SC), but not POS and NEG scenes (VAL) will be addressed below.

1.2.2. Specific Reconsiderations

By re-considering the differential impact of REAP and SUP on emotions induced by either SOC or NSOC emotional scenes, the following issues merit a closer look.

(1) Firstly, as reported in fMRI study N° 2a (see Chapter VI.4), there was only one region – the right middle insula (see Figure 5) – that displayed a selective modulation of activity to NSOC (*versus* SOC) scenes as a function of the two different ER strategies. This was rather unexpected, because many brain areas were found for the initial contrast NSOC > SOC (see Table 1 in fMRI study N° 2a). However, as already shortly mentioned in the corresponding manuscript, this somewhat surprising finding was due to the fact that the modulatory role of ER in other brain areas than the right middle insula was actually specific to SOC and not NSOC images, despite a significant inverse main effect of SC (NSOC > SOC). Moreover, in most of these regions (including the right middle insula), BOLD signal change to SOC images (all or some VCON) was decreased, whereas activity to NSOC images (all or some VCON) was increased or at least not different from neutral images (BASE).

For example, in the mid cingulate (xyz= 3 6 39, see Figure M17 below), there was a main effect of SC [NSOC > SOC: F(1,18)= 10.33, p= .005] and a VCON x SC interaction [F(1,18)= 5.48, p= .008], but these effects were driven by specific decreases in BOLD signal change to SOC [EMEXP > REAP: t(18)= 2.24, p= .038; SUP > REAP: t(18)= 3.43, p= .003] but not NSOC [all t(18)< .541, ps> .60] images. In addition, whereas activity to NSOC images was not significantly different from BASE [all t(18)< 1.23, ps> .23], activity to SOC scenes was marginally significantly decreased from BASE during EMEXP [t(18)= 2.05, p= .055] and REAP [t(18)= 2.02, p= .058].

Without going into too many details regarding the specific activation patterns of the respective brain areas – e.g. in the mid cingulate (see above), the NSOC *versus* SOC activation difference was only significant during EMEXP and REAP [all t(18)> 2.32, ps< .032] but not SUP [t(18)= .49, p= .63] – these data therefore suggest that such brain areas (except for the right middle insula) found for the t-test model comparing differential activation to NSOC *versus* SOC were *not* characterized by *in*creased processing of NSOC scenes, but instead by *de*creased processing of SOC images, even below the activation level of neutral images (BASE). As a result, no conclusions can be drawn regarding the influence of REAP and/or SUP on activity related to NSOC images in those brain regions, which accounts for the fact that the latter were not included into the manuscript of fMRI study N° 2a. Importantly, such r esults also demonstrate that the simple fact to find a significant activation by means of the t-test model and masking procedure does *not* automatically imply a differential modulation regarding SC (or VAL) with the predicted specificity.

Consequently, a careful secondary analysis of the precise activation patterns in the found clusters was required.

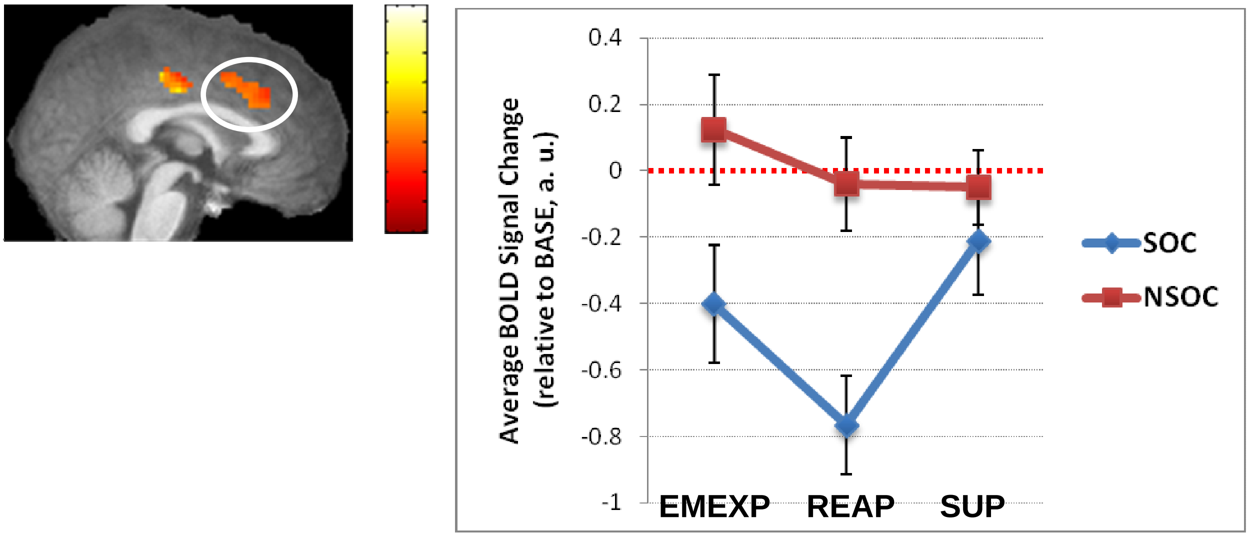


Figure M17: Mid cingulate activation for the paired t-test model ([SUP$_{NSOC}$ > SUP$_{SOC}$] > [REAP$_{NSOC}$ > REAP$_{SOC}$]) masked with the initial contrast (SOC > NSOC), *displayed at a threshold of p= .025 for illustrative purpose. BOLD signal change is depicted ± 1 S.E.M. For more information, please refer to text.*

(2) Second, in such areas where there indeed was a specific effect for the computed t-test model and masking procedure, our new data revealed that REAP and SUP differentially affected brain activity to SOC *versus* NSOC (and vice versa) images – the NSOC > SOC effect in right middle insula only decreased during SUP, whereas the SOC > NSOC effect in mOFC, mPFC, PCC, and FG only decreased during REAP, and the SOC > NSOC effect in addition displayed a lateralization effect in amygdala. Crucially, such activation patterns could not be attributed to simple arousal / intensity effects at the first place, because SOC and NSOC images were carefully matched on the latter properties. Consequently, the found dissociations had to be caused by the different computational properties of the respective brain areas displaying a VCON x SC interaction.

Most interesting in this regard was the lateralization effect discovered in the amygdala: the left amygdala seemed more affected by REAP, but right amygdala more by SUP. This finding implies at least two fundamental issues.

On one hand, bilateral amygdala activation was characterized by a main effect of SC (SOC > NSOC). As already discussed in Chapter VII.1.1.2., this suggests that the amygdala most likely preferentially processed SOC stimuli depicting faces due to the intrinsically higher *"... importance* *and usefulness of facial expressions in directing our behavior and regulating our social interactions…”* [157] and high biological relevance irrespective of valence [202, 203]. Even though it cannot be definitively excluded that there was no VAL and thus arousal / intensity effect for SOC images in bilateral amygdala (see above), it is still striking that the main effect of SC was much more pronounced and determinant than the VAL comparison.

On the other hand, and despite the fact that the just mentioned main effect of SC was present in the amygdala bilaterally, we still found a differential influence of REAP and SUP specifically on activity to SOC images as a function of hemisphere (left: REAP; right: SUP). This in turn implies that the left and right amygdala processed different aspects of SOC images, because their activity was modulated by distinct ER strategies. Regarding the left amygdala, the selective decrease of the SOC *versus* NSOC activation difference during REAP suggests computational properties that are susceptible to cognitive regulation. In turn, the specific decrease of the SOC *versus* NSOC activation difference during SUP in the right amygdala alludes to processes more prone to behavioral inhibition. This is in line with several previous accounts of lateralized amygdala activation during ER and mechanisms related to emotional memory [86, 87, 91, 205, 206], but for the first time demonstrates such effects by directly comparing REAP and SUP.

1.2.3. Specific Limitations

Besides a general limitation of only including female subjects in fMRI study N° 2a (see Chapter VII.3.), it could also be argued specifically regarding the differential impact of REAP or SUP on SOC *versus* NSOC (and vice versa) image processing that the REAP strategy applied by the subjects did not correspond to a usually employed cognitive re-evaluation strategy.

In *fMRI study N° 2ab*, we told our participants to imagine the depicted scenes as real as possible during EMEXP (by pretending as if they were attending to the scenes in real life, and by imagining the involved persons to be persons they know). In turn, during REAP, we instructed the subjects to imagine the pictures seen to be unreal, and more specifically scenes from TV or movies artificially set up to convey a certain message. The latter regulation strategy was thus different from decreasing personal relevance and/or proximity, the opposite strategy as used during EMEXP. Moreover, it may seem a bit strange using the “pretend unreal” strategy in real life.

However, we based our choice of cognitive regulation strategy during REAP on a previous study [87] which showed in a similar fMRI experimental setting that all but one of the subjects employed the “pretend unreal” strategy to down-regulate negative (and to a lesser extent) positive emotions, and that the subjects preferred this strategy over decreasing personal relevance / proximity, all strategies which they could freely choose from. For regulating positive emotions, the results were less consistent, but the study did not reveal any differences in subjective emotion experience ratings between the three distinct regulation strategies. It thus seems like the “pretend unreal” strategy is preferentially employed by participants of similar ER studies that primarily use visual stimuli to induce emotions. As we applied the same experimental setup, the choice of telling our participants to imagine the scenes to be parts of a TV film or movie thus seemed logical. This, however, does not imply that such a cognitive ER strategy would also be the best one in real life. In the latter case, most highly emotional experiences are evoked by scenarios involving loved ones (e.g. the loss of a parent or partner), for which it seems more adequate to apply the personal relevance decrease strategy. Whether this strategy also works in an fMRI environment remains to be seen, and should be tested with images depicting known people, and not strangers as in our case.

Another limitation of *fMRI study N°2a* is the fact that arguments regarding effects of arousal (negative *versus* positive images) are contrasted with differences in the amount of social processing (social *versus* non-social images), and the influence of the two different ER strategies (REAP *versus* SUP), of which REAP is thought to be more cognitive and SUP more behavioral, are expounded without any direct measures of arousal through e.g. physiological measures or subjective emotion experience ratings. This is important especially regarding the lateralization effect in the amygdala as a function of ER strategy. However, it is not the case that such measurements were not performed at the first place, but that there were several technical or other problems preventing their inclusion in the corresponding article (as of its current status).

As mentioned in the article regarding fMRI study N° 2a (see Chapter VI.3.), there were differences in the reduction of subjective emotion experience as a function of ER strategy employment, but the latter did not translate into systematic differential relations with BOLD signal change patterns. In other words, even though there were different behavioral effects of SUP and REAP on the experience of POS or NEG emotions, we found only very few brain areas that displayed correlations between the actual neuro-physiological response and such behavioral effects. As a consequence, such relations were not reported in the corresponding article. However, in the left amygdala, there actually was such a behavior-functional relation (see Figure M18 below): we found a negative correlation between individual ratings of SN images during REAP and fMRI signal change in this condition [r= -.482, p= .037], meaning that the less negative subjects rated SN images during REAP, the less their left amygdala was activated by these pictures. No such correlation was found in right amygdala [r= -.238, p= .33] or any other regions.

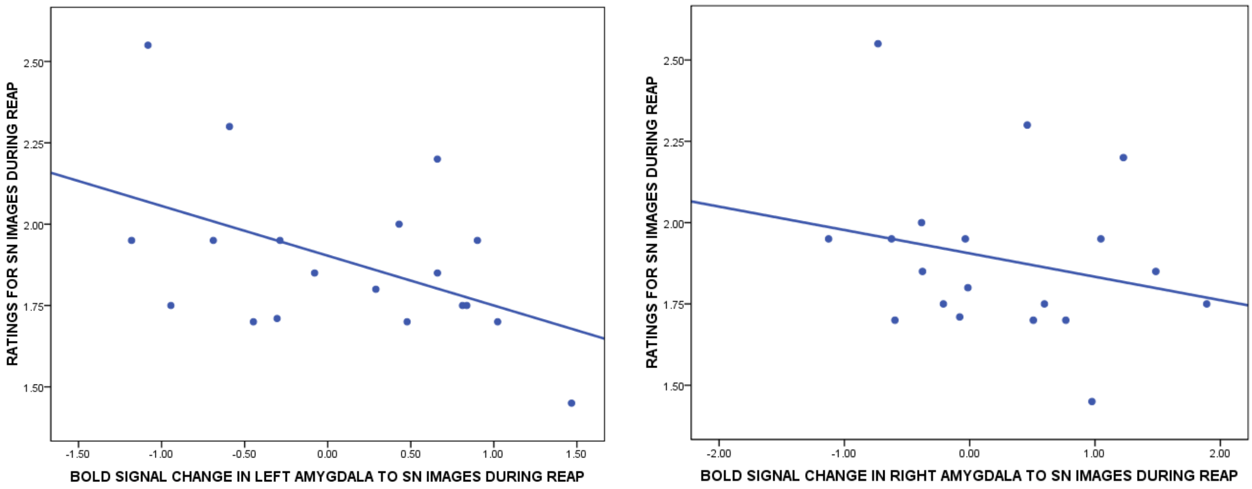


Figure M18: Correlations between behavioral ratings to SN images during REAP and BOLD signal change to the same condition in left and right amygdala. *This finding thus indicates that cognitive re-evaluation reduced activity to negative social displays only in left amygdala, displaying a selective reduction of the differential response to SOC versus NSOC images during REAP. No such effect was found in right amygdala, displaying a selective reduction of the differential response to SOC versus NSOC images during SUP.*

Unfortunately, no relation between ratings of emotional images during SUP and corresponding right amygdala activity could have been revealed, so that the behavioral-functional association was only present for one ER condition, but not the other. Despite the absence of any effects of REAP on BOLD signal change in the right amygdala, it is thus difficult to draw final conclusions. In addition, even though subjective ratings are often used for interpretational purposes in fMRI studies on ER, they are – as their denomination already suggests – subjective, which means that they cannot be controlled for objectively and independently. Different, objective – e.g. physiological – measurements to support findings related to subjective ratings are therefore always a very helpful argument.

Accordingly, we also performed physiological measurements during fMRI study N° 2a, including eye-tracking and pupil diameter measurements, skin conductance, breathing and cardiac frequency registration. At the stage when the article reporting the findings of fMRI study N° 2a (see Chapter VI.4.) was included in the present doctoral thesis, most of these parameters could not have been analyzed yet. This is due to technical problems on one hand, and lack of time on the other hand. However, this will still be pursued after the doctoral qualification. The same also applies for data of fMRI study N° 2b summarized in the corresponding manuscript (see Chapter VI.5.), where the analysis of eye-tracking data would also be important to disentangle decreased brain activity during ER from simple attention effects.

VII.2. Individual Differences

The second main goal of the present thesis was to investigate the modulatory role of internal context, also referred as to social motives and goals, and in particular *individual differences* on emotion perception and regulation and their corresponding neural correlates. For several reasons already explained beforehand (Chapter IV.4.6.), adult attachment style (AAS) was chosen as psychological model of choice to assess the influence of specific personality factors related to avoidance (AV) and anxiety (AX) on social perception.

Again, as already applied in Chapter VII.1., the obtained results on the influence of AAS on emotion perception, including fMRI studies N° 1ab and the behavioral study (Chapter VII.2.1.) will be assessed first, followed by a discussion of found effects on emotion regulation (fMRI study N° 2b; Chapter VII.2.2.). Finally, the interplay between AAS and self- *versus* other-perception will be recapitulated (fMRI study N° 3; Chapter VII.2.3.).

VII.2.1. Emotion Perception

2.1.1. Obtained Results

Even though the investigation of any AAS effects was performed for both **fMRI studies N° 1ab**, only the first study revealed significant results. These showed that AV and AX differentially influence the perception of specific expression x feedback interactions related to either social reward (SF-W) or social punishment (AF-L). More precisely, whereas high AV scores were associated with decreased activity to the SF-W condition in two brain areas of the reward system (ventral tegmental area and ventral striatum; see Figure 2 of fMRI study N° 1a), low attachment security (SAS) and in particular high AX scores were related to increased (left) amygdala activation to the AF-L condition (see Figure 3 of fMRI study N° 1a). Such findings were interpreted according to the different secondary attachment strategies normally associated with the two distinct attachment styles, leading to hyper-vigilance to cues of negative social signals in anxious individuals, but to de-activation of the attachment system in avoidant people [185]. Importantly, fMRI study N° 1a for the first time revealed effects of AAS, and specifically AV, on neural correlates of positive emotion perception.

The **behavioral study** confirmed the findings of fMRI study N° 1a regarding the decreased representation of positive emotions in subjects scoring high on AV by showing a similar relation between subjective pleasantness ratings of selectively social positive emotional scenes and AV measures. In addition, the behavioral study also found effects of AX on subjective emotion intensity and controllability ratings, pointing in the same direction as the functional findings of fMRI study N° 1a. These effects were specific for social negative emotional images, revealing increased intensity and decreased controllability of the latter, the higher the AX scores. However, we also found increased intensity ratings for positive emotional scenes in general as a function of AX, which made the interpretation of the modulatory role of AX on emotion perception somewhat difficult.

2.1.2. Specific Reconsiderations

By re-considering the results obtained from all experiments involving the assessment of AAS on emotion perception, the following issues emerged.

(1) Firstly, what was striking in *fMRI study N° 1a* was the specificity of effects regarding the influence of AV and AX the perception of distinct social emotions. Even though the regions of interest (ventral striatum & VTA, and amygdala) were initially characterized by *contrasts* (SF-W > AF-W and AF-L > SF-L) and later on by whole brain correlations (multiple regressions) on the same contrasts, the correlations with AV and AX only emerged for one *specific stimulus condition* of these comparisons: AV was only negatively related to activity for the SF-W (ventral striatum & VTA), and AX to activity for the AF-L (amygdala) condition. Such relations are thus much more selective than any measures indicating correlations between questionnaire scores and contrast values (see i.e. [194]), where the directionality of the observed effects and their specificity are much harder to assess, if not even undeterminable.

(2) Second, it should of course be mentioned here that, even though it was an objective of *fMRI study N° 1b*, we did not find any effects of AAS on familiar face recognition (behavioral as well as functional data). This was somewhat surprising, because the secondary attachment strategies and attachment working models associated with AV and AX would actually suggest such influences [185]. It could for example be hypothesized that increased vigilance to clues of negative social interactions as a hallmark of AX would entail better memory formation for (certain) faces, whereas deactivating strategies characterizing AV would generally impair the same process. Accordingly, it can only be speculated about the absence of any AAS effects on the memory task in fMRI study N° 1b. This is even more difficult when considering the fact that no other study has investigated this relation up to date.

One argument that could be put forward is concerning the – most likely – unconscious nature of found social context effects in amygdala, ACC, and caudate (see manuscript; Chapter VI.2.).

Particularly regarding AV, there is behavioral evidence [195] that unconscious (subliminal presentation) perception of emotion from a face entails a “normal”, whereas conscious (supraliminal presentation) perception results in an altered (inverted) mimicry response. This would imply that any modulatory influence of AV on mimicry and probably also neural activity during emotion perception required conscious processing of the present emotional information, and would thus not have been detectable in fMRI study N° 1b due to its unconscious nature. However, the influence of AV on unconscious and conscious mimicry reactions has only been shown to angry and not happy faces, and no such effects have been reported regarding AX. Moreover, a very recent fMRI study [207] actually describes some influence of AV on the neural response to masked sad faces and therefore most likely shows spontaneous effects of one AAS dimension on emotion perception . As a consequence, it is difficult to attribute the absence of any AAS effects on the neural correlates of familiar face perception to processes related to consciousness, even though this might represent a possibility.

(3) Third, for both *fMRI studies N° 1ab* , a notable strength of the experimental design was to include eye-tracking to monitor the eye-movements of subjects. This helped to make sure that 1) any effects of AAS on brain activations found during emotion perception were not due to simple attention allocation – e.g. avoiding looking *versus* staring at certain facial expressions, and 2) that emotional memory (even though independent of AAS) was not caused by differences in exposure time to specific faces / stimulus conditions during encoding.

(4) Fourth, as already briefly mentioned in the summary of obtained results (Chapter VII.2.1.1.), the findings of the *behavioral study* were somewhat difficult to interpret, particularly regarding the effects of AX on subjective intensity ratings of emotional images. This was due to the fact that AX was not only positively correlated with intensity ratings of SN, but also with both social positive emotional image categories.

One possible explanation of these findings might lie in the concept of ambivalence associated with AX. Whereas anxiously attached people show increased vigilance to social clues because of their negative model of self entailing low self-esteem and consequently the search for external reconfirmation, they still hold negative representations of others due to aversive memories of previous conflicting interactions. Anxious people might therefore experience the urge to get close to others, but nonetheless feel resentful towards the same people [185]. Such a pattern of ambivalence might thus make anxious people responsive to both positive as well as negative social clues, which could explain some of the data obtained in the behavioral study. However, this explanation should be regarded with suspicion, because very little experimental evidence has been collected up to date regarding AX and ambivalence.

In addition, the abovementioned ambivalence account cannot fully explain why we also found a positive correlation between AX and intensity ratings of non-social positive (NSP) emotional images (the relations between AV and pleasantness ratings, and between AX and control ratings were always specific to social [negative] images). It might, however, be the case that anxiously attached participants are also more responsive to positive (and not only negative) emotional scenarios, because they are generally inclined to searching for the experience of security, be it in an attachment-related or -unrelated context [185]. In any case, more research is needed in the domain of positive emotions (and ambivalence) in combination with AAS.

2.1.3. Specific Limitations

One potential limitation of *fMRI study N° 1a* could be attributed to the applied AAS questionnaire (validated French version of the Attachment Style Measure [ASM] [199]). The latter has been outdated, because it structures AAS into three distinct categories (secure, avoidant, anxious) instead of using a (bi-)dimensional approach, which has recently been found more accurate [188].

To make sure that the ASM still represented a valid means to assess the attachment style of our subjects, we took a closer look at its items and compared them to the questions of more recent questionnaires used to measure AAS. By doing so, we discovered that the items used for AV and AX in the ASM were in fact identical to some of the questions included for the same attachment subscales in the more often used Relationship Scales Questionnaire (RSQ; [184]). Moreover, by applying a specific analysis method of the RSQ that has been shown to be most adequate [188], the exactly same items of the ASM were indicated to serve as a basis to derive the dimensions of AV and AX. All these close relations between the ASM and RSQ were finally proven by directly correlating the AV and AX constructs obtained from the two questionnaires, revealing highly significant relations ($p < .0004$).

Therefore, even though the use of the ASM as AAS measure for fMRI study N° 1a was not ideal and has been changed for the subsequent behavioral as well as fMRI studies N° 2b and 3, it could still be shown to be valid and comparable to other reports of AAS effects on behavior and/or brain activation.

Regarding the *behavioral study*, it could be argued that there was no control for the specificity of effects observed as a function of AAS, particularly regarding AX *versus* general anxiety or neuroticism, which could have helped to disentangle some observations (e.g. related to ambivalence, see above).

Even though this objection is justified, as direct controls are always better than indirect ones, it should be mentioned here that the behavioral study was initially designed as a stimulus-selection paradigm and not as an independent study. Accordingly, attachment-measures were included because of their direct implication at a later stage (fMRI study N° 2b with the stimuli chosen according to the behavioral study), but no other personality traits were systematically measured. In addition, a recent review [188] has shown that the attachment dimensions assessed with the attachment questionnaire used (RSQ) rarely display more than 25% shared variance with other personality constructs derived from different theoretical or descriptive frameworks. This holds true for AX in association with neuroticism and trait anxiety, but also threat appraisal, or rejection sensitivity; as well as for AV in relation with defensiveness, coping by distancing, or cognitive openness. It can thus be assumed that there should not be too many confounds regarding this issue in the behavioral study.

VII.2.2. Emotion Regulation

2.2.1. Obtained Results

fMRI study N° 2b revealed several interesting effects of AV on the regulation of social positive and negative emotions. If asked to attend to social negative (SN) emotional scenes naturally (emotion expression; EMEXP) – without applying any emotion regulation (ER) attempts –, highly avoidant participants showed increased activation in several areas of the pre-frontal and the anterior cingulate cortex (ACC) previously associated with cognitive control and conflict processing. Moreover, such high activation to SN images in highly avoidant participants persisted during re-appraisal (REAP) in ACC and in addition in bilateral amygdala, whereas it decreased for low avoidant subjects in the latter two areas. This points to a failure of REAP to decrease SN emotion processing associated with AV. Finally, AV was also correlated with increased activity to social positive (SP) emotional scenes during EMEXP and suppression (SUP) in supplemental motor area (SMA) and ventral striatum (putamen and caudate). Such effects were attributed to the suppression of positive emotion expression when seeing SP displays (SMA) and the associated positive, rewarding experience of hiding ones positive emotions (ventral striatum). What is concerning the influence of AX on ER of social emotions, our data only revealed an effect in the left para-hippocampus to SP emotions during EMEXP, but no other associations, in particular with the regulation of SN emotions.

2.2.2. Specific Reconsiderations

(1) Firstly, the most striking issue regarding *fMRI study N°2b* and thus the influence of AAS on the regulation of social emotions was the mere absence of any effects of AX. As could have been expected according to AAS theory and literature [185, 194] and also based on the results of fMRI study N°1a (Chapter VI.1), AX should have had a strong impact on the (perception and) regulation of particularly social negative emotions. However, we only found a correlation between activity in the left para-hippocampus to social positive emotional scenes during EMEXP (Figure 6 in fMRI study N°2b). Moreover, our data actually revealed inverse relations between AX and BOLD signal change to non-social negative (NSN) images during both EMEXP as well as SUP in lPFC and dlPFC, SMA and motor cortex, as well as anterior insula (see Figure M18 below for the activations found with the contrast EMEXP$_{SN}$ > EMEXP$_{NSN}$ x AX$_{POS}$). Additional statistical tests using SPSS confirmed the specificity of those correlations to NSN and not SN images. The interpretation of the abovementioned activation patterns is rather difficult, and already partly attempted in the corresponding manuscript.

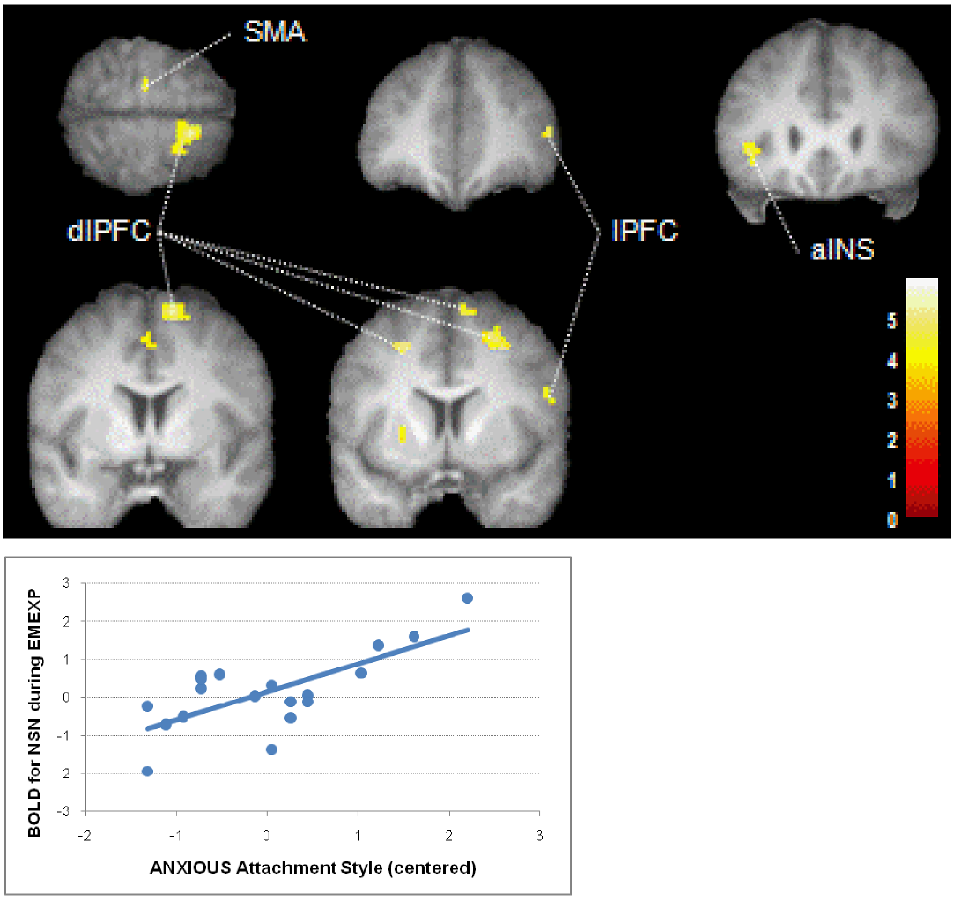


Figure M19: Statistical parametric map for the contrast EMEXP$_{SN}$ > EMEXP$_{NSN}$ correlated negatively with AX at p< .001 (and masked exclusively with AV$_{POS}$ at p= .01) and correlation plot of extracted beta values (dlPFC, xyz= 24 9 54) for the NSN condition with AX (r= .718, p=.001). *For more explanations, please refer to text.*

Such results raise the important question whether the influence of AX is really specific to brain activations encoding for *social* emotional scenes, which is tantamount with the question whether anxious attachment really concerns only relationship-related (as put forward e.g. by [188] on the RSQ) or also more general anxiety mechanisms (as e.g. measured by the STAI-T [163, 164]). This issue will be addressed later (Chapter VII.3.).

(2) Second, as already put forward above, it is remarkable that *fMRI study N° 2b* also revealed effects of AAS, and more precisely AV, on emotion regulation of (social) positive emotional stimuli. This follows data put forward by fMRI study N° 1a and the behavioral study and thus continues to demonstrate that AAS also crucially modulates the perception of positive social clues and not only negative, threatening information.

2.2.3. Specific Limitations

Again, the general limitation of including only female subjects into fMRI studies N° 2ab emerges, which will be addressed later (Chapter VII.3.). In addition, even though physiological measures were collected, the data could not be incorporated into the current stage of the corresponding article, which applies also for fMRI study N° 2a (see Chapter VII.1.1.3.).

Otherwise, the design and analysis of fMRI study N° 2b tried to account for as many limiting factors as possible. Firstly, AAS scores of potential study participants were obtained some days or weeks prior to scanning, and the final group of 19 subjects was selected to display the best possible AV and AX score distribution. Second, general anxiety (STAI-T) was assessed besides the AX score to control for any unspecific responses. And third, all results were obtained with a multiple regression approach, an additional masking procedure, and a final selective correlational analysis using SPSS on single image categories only and not contrast values.

VII.2.3. Self- *versus* Other-Perception

2.3.1. Obtained Results

The last fMRI study (**fMRI study N° 3**) carried out in the context of the present doctoral thesis for the first time tried to disentangle the influence of AAS on the neural correlates of self- versus other-perception. For this purpose, participants were asked to attribute positive and negative trait adjectives either to themselves (self) or a close friend (other).

The behavioral data indicated that there was a general bias for positive adjective attribution, but that AV was associated with a marginally significant increase of this bias specifically regarding self-evaluation. The same notion was also found in the fMRI data regarding AV, revealing increased intentionality, mental imagery, and outcome monitoring during the attribution of positive traits to oneself. Conversely, AX was linked with a more pronounced processing of information regarding others, including mentalizing in general, as well as agency processing and emotion regulation employment for both negative self- and other-evaluation.

2.3.2. Specific Reconsiderations

(1) Firstly, the most remarkable finding regarding AV emerged when specifically investigating positive self-perception, revealing increased activity in ventro-medial PFC (Figure 5ab). As already mentioned in the corresponding article (see Chapter VI.6.), this activation could not only be interpreted as increased resource mobilization to evaluate possible outcomes of positive self-attributions [126, 127] as a function of AV, but also linked with the notion of social rejection previously associated with activity in the same brain region [65]. This finding is particularly interesting, because it implies that avoidantly attached individuals evaluate positive self-attributes not only based on their intrinsic properties and egocentric implications, but also according to their appraisal by others. This is well in line with 1) our findings of fMRI study N°1a, where we observed a dampened ventral striatum response to social reward as a function of AV, but a normal response in reward-related areas to winning in general, and 2) attachment theory in general, linking AV with a negative model of others, but a positive model of the self [185].

(2) Second, the correlational analysis with AX disclosed increased mentalizing about others in general, deduced from higher BOLD signal change in medial pre-frontal cortex (mPFC; see Figure 3ab). The particularity of this finding lies in the anatomical localization of the found effect, as a previous study [72] has shown that this part of the ventral mPFC is not only involved in mentalizing *per se*, but displays a preference for mentalizing about similar (as compared to dissimilar) others. Our finding of ventral mPFC activation as a function of AX could therefore be explained by attachment theory linking AX with an over-estimation of the similarity between the self and others [208].

2.3.3. Specific Limitations

One limitation of (the present stage of the manuscript on) fMRI study N°3 could be seen in the fact that most brain activation results are reported as correlations between attachment style measures and contrast differences, but not specific conditions (e.g. [$SELF_{(POS + NEG)}$ > *CONTROL]*). However, this can be improved rather easily by extracting raw beta values from regions of interest and correlating them with AV and AX separately, and will be incorporated into the final version of the manuscript before submission.

Another limitation could be seen in the fact that no measures to control for any general anxiety effects as opposed to relationship-specific anxiety (AX) were incorporated. Yet, as already discussed before, the use of the RSQ being thought to be rather specific to AX [188] should already have provided some degree of specificity.

VII.3. General Limitations

As already mentioned previously, a general limitation of fMRI studies N° 2ab could be seen in the fact that we only included female subjects. This had two main reasons.

On one hand, stimulus pre-testing (results are described in the behavioral study, Chapter VI.3.) was carried out at the Faculty of Psychology at the University of Geneva, with a total number of n= 57 subjects. Because out of the latter, the prevailing amount were female (n= 55), we decided to keep data only from female subjects for reasons of homogeneity.

On the other hand, as previous fMRI studies on emotion regulation (see e.g. [87]) have also only recruited female subjects, and because of evidence that emotions are typically more intense and more prone to regulation in women [209], we again hoped to obtain more homogenous results if including only female subjects.

One way to address this gender issue in future studies would be to recruit enough participants from each sex to be able to perform a gender effects analysis allowing for a direct comparison of results.

VII.4.Selectivity of Results regarding Adult Attachment Style

In Chapter V.2.5., a more general question about the specificity of adult attachment style (AAS) measures regarding the neural correlates of social perception was asked. This stems mainly from the fact that, despite a large body of behavioral psychological experimental data on AAS is available, only very few studies have used AAS in an fMRI environment [194, 207, 210] up to date.

Such a specificity of AAS regarding fMRI data can be understood in three different ways. On one hand, it can be considered as a measure of how selective the found modulation of brain activity is regarding attachment avoidance and anxiety, as compared to e.g. general anxiety, neuroticism etc. On the other hand, specificity can also concern the effect of AAS on positive or negative emotion perception and regulation, because traditionally, attachment system activation has been conceptualized to come into play in the presence of potential threats, which are mostly regarded as negative. Finally, specificity can also regard the effect of AAS on the perception and regulation of social *versus* non-social emotions, as AAS is usually linked with social processes – such as social interactions between e.g. a mother and her child or between romantic partners – and not so much with scenarios involving only one person or non-social content.

VII.4.1. Attachment Style *versus* other Personality Measures

In all of the experiments of the present thesis including AAS as individual difference variables, it was carefully controlled for whether the found effects were specific for attachment avoidance or anxiety, particularly regarding possible confounds with of the latter attachment orientation with general anxiety / neuroticism. This was done by using an AAS questionnaire already previously shown to derive attachment anxiety measures independent from general anxiety [188] (except for fMRI study N° 1a, but the questionnaire applied there was subsequently probed and found not to differ significantly from the questionnaire referred to above [188]), but also by including a different measure of general anxiety (STAI-T; [163, 164]; except for the behavioral study – the reasons for this are explained in Chapter VII.2.1.3) and other measures of personality.

Whereas the results of fMRI study N° 1a were specific for attachment avoidance and anxiety, not showing any other correlations with different personality measures, this was not the case for fMRI study N° 2b, where the attachment anxiety and general anxiety scales were marginally significantly correlated (p= .08), and general anxiety was in addition significantly related with a measure of the degree of negative self perception as a function of anxious attachment style (p< .001). Moreover, many correlations between anxious attachment and brain activity during fMRI study N° 2b were in fact also significant regarding general anxiety (not mentioned in the corresponding article, because this mainly concerned correlations between attachment anxiety and BOLD signal change to non-social images – see Chapter VII.4.2. below).

Such findings clearly demonstrate that one has to be very careful when attempting to separate any effects of relation-specific anxiety (attachment) from general anxiety (STAI-T) or neuroticism (associated with general anxiety) and that a clear distinction between these two factor might in fact be very difficult to obtain. Even though study participants could be selected beforehand according to their attachment anxiety and general anxiety scores to make sure that the latter two measures are not correlated, this would still not guarantee for the absence of any relations present in brain correlation data. Also, anxious attachment and general anxiety could display correlations in the same brain areas, which would make a dissociation of the latter two variables rather difficult.

However, before drawing any definitive conclusions or even proposing that a clear-cut distinction between attachment anxiety and general anxiety is not possible at all, more experimental fMRI data is needed to further elaborate on this issue.

VII.4.2. Attachment Style and Positive *versus* Negative Emotions

Attachment system activation is described as a consequence of felt insecurity by the presence of a potential threat with the goal to re-establish security. If such attempts are repeatedly experienced as unsuccessful, entailing prolonged or permanent activation of the attachment system, secondary attachment strategies come into play, either causing hyper-vigilance to (anxiety) or suppression of (avoidance) attachment-related information [185]. Within this theoretical framework, both the threats that cause initial attachment system activation, and the attachment-related clues that are either more readily attended to or attempted to be suppressed, are predominantly associated with negative events and/or emotional states. As a consequence, it might be assumed that attachment style specifically affects the processing of negative emotions / events.

However, attachment theory also stresses the fact that the appraisal of possible threats is highly subjective, meaning that "… *every event perceived by a person … as threatening tends to activate the attachment system* …" [185]. Therefore, activating triggers would not necessarily have to be equivalent with negative, but could also represent positive emotions / situations.

Despite this theoretical proposition, most behavioral psychological and all fMRI studies investigating the function of the attachment system up to date have only used negative stimuli in their experimental designs. There are only very few reports indicating that, particularly attachment avoidance, might influence the processing and subjective experience of positive emotions [211-

213], but none of those studies has used fMRI to investigate the neural correlates underlying such effects.

Our new data obtained from all studies included in the present doctoral thesis supports the latter notion by showing various effects of – particularly attachment avoidance – on the perception as well as the regulation of positive emotions. Moreover, these effects were predominantly specific for social but not non-social positive emotional scenarios (see also Chapter VII.4.3. below). This implies that, for people who aim at keeping their attachment system in a de-activated state, social positive emotional situations also seem to represent potential threats, entailing attempts to down-regulate their emotional impact. This leads to e.g. reduced responses to social reward (fMRI study N° 1a), reduced pleasantness ratings of social positive scenes (behavioral study), increased motor and reward-related activity to social positive images during emotion expression as well as suppression (fMRI study N° 2b), as well as increased representation of positive traits in ventral medial PFC and other pre-frontal cortical brain areas (fMRI study N° 3).

What is concerning attachment anxiety, our experiments revealed prevailing effects on negative emotion perception and regulation (except for increased intensity ratings for positive images as a function of anxious attachment style in the behavioral study). It thus seems as if this attachment orientation is more specific regarding the appraisal of only negative events as potential threats. This probably comes from the fact that anxious attachment involves low self-esteem and a thereby resulting dependence on external re-confirmation, making anxiously attached persons highly receptive for positive social feedback as security-enhancing – and thus not threatening – information [185].

VII.4.3. Attachment Style and Social *versus* Non-Social Emotions

Finally, similar to the question about the specificity of attachment system activation to negative versus positive emotions (see above), it is also not yet well established whether attachment avoidance and/or anxiety specifically influence the perception of social (attachment-related), or whether they also affect the processing of non-social (attachment-unrelated) material. This comes from the fact that attachment theory does not make any clear predictions regarding this point. Whereas the resolution of felt insecurity is always associated with support-seeking behavior and thus social processes, the elicitation of felt insecurity is linked with both social (e.g. the loss of a parent or romantic partner) and non-social (e.g. a threatening medical diagnosis) events [185].

What is concerning pervious experiments on attachment system functioning, only very few have included both social as well as non-social stimuli showing differential effects of avoidance and/or anxiety on the processing of the latter [214], but this has not yet been assessed with neuro-imaging methods. fMRI study N° 2b included in the present d octoral thesis was thus the first study of this kind systematically addressing the selectivity question of attachment avoidance and/or anxiety regarding the perception and regulation of social *versus* non-social emotions.

Our results showed that, whereas attachment avoidance seemed to rather specifically affect neural activity during *social* emotion perception and regulation (some brain areas even displayed positive correlations for social, but negative correlations for non-social emotion regulation with attachment avoidance), anxious attachment style affected both stimulus categories, and in fact had a preferential effect on the regulation of *non-social* emotions.

As clear and plausible our findings were on avoidant attachment style, as puzzling the effects observed were for attachment anxiety. There are several different interpretations put forward in the corresponding manuscript(s), but these are far away from providing a final conclusion. More experiments are thus clearly needed to further clarify the exact relations between anxious attachment style and the processing of social versus non-social emotional material.

VII.4.4. Summary

The data collected on the specificity of AAS on emotion perception and regulation in the context of the present thesis suggest the following.

Avoidant attachment style was found to represent a rather specific personality trait measure, to affect both positive and negative emotion processing, but to predominantly modulate social and not non-social stimulus perception and regulation. This attachment orientation thus displayed prototypical subjective experience and brain activation patterns throughout all six experiments included in the present thesis, making it a reliable tool for investigating the modulatory role of individual differences on the neuro-physiology of human social functioning. Particularly of interest is the fact that avoidant attachment style also had strong effects on positive emotion perception and regulation, a field in the domain of social neuroscience not yet exhaustively explored.

Anxious attachment style, in turn, displayed rather strong connections with general anxiety, was predominantly affecting negative emotion processing, but modulated both social as well as non-social emotion perception and regulation. Such a pattern of effects thus makes it difficult to draw any sound conclusions on this attachment orientation. More experimental – fMRI as well as behavioral – data is clearly needed for further clarification.

VII.5.Conclusions

The six experiments included in the present thesis revealed several new relations between emotion perception and regulation as two core processes in social neuroscience; and context as well as individual differences – in particular adult attachment style – as modulatory external and internal variables. Besides the very specific results of each fMRI and behavioral study allowing the derivation of particular conclusions (see previous chapters), at least two issues emerged from a more comprehensive point of view.

On one hand, the obtained experimental data demonstrate in a very descriptive manner that emotion research is a tremendously complex scientific domain. Unlike in many other research fields where there is a rather straight-forward “one-to-one” stimulus-response relationship, most processes in the affective sciences are much more prone to modulatory influences. This already emerges when referring to some emotion theories, such as the appraisal or component process model of emotions (see Figure M2) or the process model of emotion regulation (see Figure M4 and Figure M20 below), but becomes even more clear through experimental work included in the present thesis.

The behavioral and fMRI studies obviously demonstrated that a given emotional stimulus can lead to very different subjective experience-based as well as neuro-physiological responses, rather than to a prototypical, always similar or even uniform reaction. For example, the exposure to a smiling face does not automatically lead to a positive subjective feeling state accompanied by reward-related brain activity – as suggested by e.g. basic emotion or low bi-dimensional space theories (see Chapter I and Figure M1). Quite in contrary, the perception of a smiling face does crucially depend on the external (social) context within which this face is seen – a smile could signify genuine cheerfulness, but also mockery or malicious joy –, as well as on internal psychological variables determining whether social interactions are desirable in general (cf. adult attachment theory), and whether particularly social positive emotional displays are to be trusted or not (cf. avoidant attachment style).

Such considerations give rise to two main considerations. Firstly, any experimental investigation of emotional processes should carefully take into account the more complex modulating factors influencing stimulus perception as well as regulation, like context and individual differences, besides controlling for other basic stimulus properties such as luminosity, spatial frequency, valence, arousal/intensity, social complexity, etc. And second, emotion research should probably refer to more recent theoretical accounts of general emotion organization or structure, as proposed by e.g. the appraisal / component process model of emotion, which allow for interpretations regarding such contextual and individual difference effects.

On the other hand, the results acquired in the scope of the present thesis could help clarifying some aspects of the process model of emotion regulation ([84, 85] see Figure M20 below), which mainly describes two means of emotion regulation (ER) – re-appraisal (REAP) and suppression (SUP) –, but in fact mentions several other antecedent-focused ER strategies. Moreover, the main ER strategies (REAP and SUP) are described in a rather deliberate, cognitive manner, but could also be affected by other influences, such as contextual and individual difference variables.

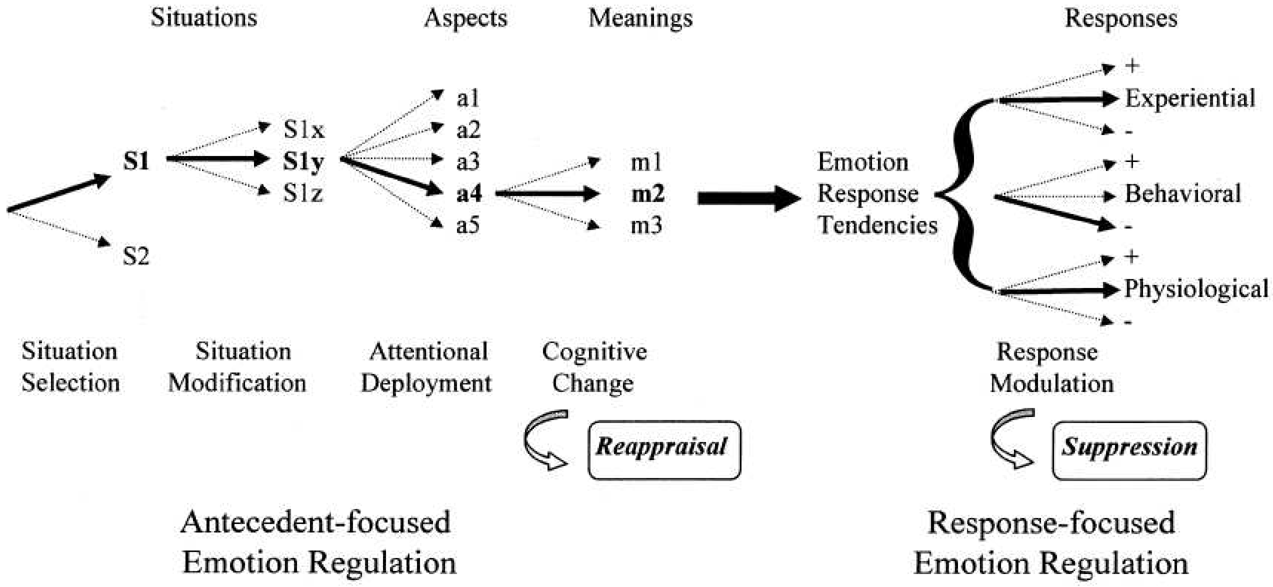


***Figure M20: Process Model of Emotion Regulation** (original in [84]).*

The process model of Emotion Regulation (see Figure M20 above) mentions *situation selection*, *situation modification* and *attentional deployment* as three antecedent-focused ER strategies preceding REAP (and SUP). Without going into too many details regarding the specific underlying theoretical constructs, our new experimental data implies that both social context as well as individual differences would importantly modulate these early ER steps.

The fact whether a situation contains social clues or not should influence its selection, modification, as well as attentional deployment, as social information is thought to be more relevant for humans in general, leading to increased activity in key emotion areas such as the amygdala ([157, 160]). Moreover, adult attachment style should crucially interact with the social scene content factor by not only determining whether a (social or non-social) situation will be selected at the first place, but how its impact will be modified and which of its aspects attention will be deployed to. In the case of attachment avoidance, our results indicate that non-social situations should be preferred over their social counterparts and that within social scenarios not only negative, but also positive emotional information will be avoided. Regarding attachment anxiety, our data is less clear, but still implies that social situations should be chosen more likely and that particularly their negative aspects will be in the focus of attention.

What is concerning REAP and SUP, our new data indicates two things.

Firstly, scene content (social *versus* non-social) not only determines which brain areas are preferentially activated, but also which ER strategy has stronger effects in reducing such localized increased activation. More precisely, REAP seems to be better suited to regulate particularly social, whereas SUP appears to be more efficient to regulate specifically non-social emotions. This might be due to the fact that social emotions are more prone to cognitive influences and thus cognitive re-evaluation effects, whereas non-social emotions are better controllable through behavioral inhibition due to their strong relation with bodily mechanisms related to arousal (cf. [87]).

Second, individual differences in attachment style also seem to entail differences in preferential ER strategy use. This was particularly evident for attachment avoidance in our new fMRI results, suggesting an exclusive employment of SUP to decrease the impact of both positive as well as negative emotions. This implies that it is not only the different stage in which ER becomes active that determines the choice of ER strategy employed (REAP normally earlier than SUP), but that certain personality traits can in fact lead to a more general preference for – in our case – SUP or REAP employment.

Taken together, our new behavioral and fMRI findings suggest that there is a strong interaction between external (social context) as well as internal (individual differences – and in particular adult attachment style [AAS]) variables influencing social perception, meaning that AAS will influence the choice and interpretation of, as well as regulation strategy used to modulate emotions induced by the exposure to different social contexts. In addition, our new data also implies that emotion perception and regulation are tightly linked, as e.g. for avoidantly attached subjects, the use of SUP seems not to come into play only at rather late stages of emotion regulation (when an emotion has already been induced), but already quite early in the emotion perception and appraisal process. It is therefore important to not only further extend the knowledge on the four above-mentioned mechanisms (external *versus* internal modulation & emotion perception *versus* regulation), but also to examine their interactions. Only by doing so, a more complete picture of social perception will eventually emerge.

VII.6. Future Directions

To further extend the knowledge on the influence of context and individual differences on brain activity during social perception, the following investigations could be performed in the future.

On one hand, as the findings of the present thesis were obtained by means of fMRI – allowing for a precise functional-anatomical localization of brain activations but not providing any information about the time-course of such effects –, the same or similar experiments could be performed by using electro-encephalography (EEG) to derive time-plots of social emotion perception and memory (fMRI studies N° 1ab), emotion regulation processes (fMRI studies N° 2ab), or self- versus other-perception mechanisms (fMRI study N° 3), and their modulation by context and individual differences, respectively. Even though EEG would probably not allow for any insights into the temporal aspect of activation patterns in sub-cortical areas such as the amygdala, it could still help clarifying some effects found in pre-frontal and other cortical areas.

Of particular interest would thereby be the question whether the influence of attachment style is already present from very early processing steps, suggesting a rather automatic implementation of individual attachment style difference variables (as found e.g. here [207]), or whether such effects would only be detectable after some delay, proposing more voluntary, cognitive mechanisms (as proposed e.g. here [195]). Similarly, it would be nice to see whether the expression x feedback integration observed in fMRI study N° 1a influences brain activity already very early, or whether there is an emotion or feedback effect first, which is subsequently “over-written” by the integration of the two variables, tantamount to the social context representation; and how these processing steps are modulated by attachment style.

On the other hand, it could be useful to use trans-cranial magnetic stimulation (TMS) to induce transient “functional lesions” in those brain areas found to be differentially implicated in either REAP or SUP (fMRI study N° 2a). This would allow for directly testing the absence of pre-frontal cortical regulatory control on subjective emotion ratings and peripheral as central physiological responses. The same procedure could also be applied on regions of interest found activated during emotion regulation as a function of attachment style – and particularly attachment avoidance (fMRI study N° 2b) – to investigate e.g. the absence of supplementary motor area activity during suppression on the same measures already mentioned above.

Eventually, it would also be promising to investigate whether any context and/or attachment effects on social perception could be attributed to distinct molecular or genetic mechanisms.

What is concerning the molecular aspects, there is emerging evidence that the two neuro-peptides oxytocin and vasopressin could play a crucial role in human social functioning in general, and attachment-related processes more specifically [215-218]. Studies applying physiological measures, such as fMRI and EEG, could thus more systematically be combined with either the administration or measurement of oxytocin and vasopressin to better understand the molecular mechanisms of social context perception and attachment – a prime example for a multi-level approach as suggested by the founders of social neuroscience (see Chapter I). These findings could then be contrasted to other results implying different neuro-peptides such as testosterone, which has also been found to modulate brain activity to e.g. social threat [219], or have an impact on mind reading and trust in human interactions [220].

Regarding genetic mechanisms, there is some indication of an association between different dopamine and/or serotonin receptor gene polymorphisms and individual differences in attachment style [221]. In addition, the same receptor gene polymorphisms have previously been linked with different human love styles [222]. Yet, no such associations have been found for oxytocin or vasopressin (receptor) genes. In any case, genetic sampling of participants before any brain measurements could also offer some interesting new insights, particularly into the question of attachment style heritability *versus* acquisition.

FIGURES and TABLES

FIGURES

Page

TABLES

REFERENCES

REFERENCES

1. James, W., *Principles of Psychology*. Vol. 1. 1950, New York: Dover.
2. Cacioppo, J.T. and G.G. Berntson, *Social Psychological Contributions to the Decade of the Brain - Doctrine of Multilevel Analysis.* American Psychologist, 1992. **47**(8): p. 1019-1028.
3. Harlow, J.M., *Recovery from the passage of an iron bar through the head.* Publications of the Massachusetts Medical Society, 1868. **2**: p. 327-347.
4. Macmillan, M., *An old kind of fame*. 2000, Cambridge, MA: MIT Press.
5. Ochsner, K.N. and M.D. Lieberman, *The emergence of social cognitive neuroscience.* American Psychologist, 2001. **56**(9): p. 717-734.
6. Lieberman, M.D., *Social cognitive neuroscience: A review of core processes.* Annual Review of Psychology, 2007. **58**: p. 259-289.
7. Todorov, A., L.T. Harris, and S.T. Fiske, *Toward socially inspired social neuroscience.* Brain Research, 2006. **1079**: p. 76-85.
8. Fletcher, P.C., et al., *Other Minds in the Brain - a Functional Imaging Study of Theory of Mind in Story Comprehension.* Cognition, 1995. **57**(2): p. 109-128.
9. Frith, C.D. and U. Frith, *Cognitive psychology - Interacting minds - A biological basis.* Science, 1999. **286**(5445): p. 1692-1695.
10. Perner, J. and H. Wimmer, *John Thinks That Mary Thinks That - Attribution of 2nd-Order Beliefs by 5-Year-Old to 10-Year-Old Children.* Journal of Experimental Child Psychology, 1985. **39**(3): p. 437-471.
11. Baron-Cohen, S., *Autism: The Empathizing-Systemizing (E-S) Theory.* Year in Cognitive Neuroscience 2009, 2009. **1156**: p. 68-80.
12. Gallagher, H.L. and C.D. Frith, *Functional imaging of 'theory of mind'.* Trends in Cognitive Sciences, 2003. **7**(2): p. 77-83.
13. Saxe, R., *Uniquely human social cognition.* Current Opinion in Neurobiology, 2006. **16**(2): p. 235-239.
14. Beer, J.S. and K.N. Ochsner, *Social cognition: A multi level analysis.* Brain Research, 2006. **1079**: p. 98-105.
15. Gross, C.G., Rochamir.Ce, and D.B. Bender, *Visual Properties of Neurons in Inferotemporal Cortex of Macaque.* Journal of Neurophysiology, 1972. **35**(1): p. 96-&.
16. Perrett, D.I., E.T. Rolls, and W. Caan, *Visual Neurones Responsive to Faces in the Monkey Temporal Cortex.* Experimental Brain Research, 1982. **47**(3): p. 329-342.
17. Dunbar, R.I.M., *The social brain hypothesis.* Evolutionary Anthropology, 1998. **6**(5): p. 178-190.
18. Adolphs, R., *The Social Brain: Neural Basis of Social Knowledge.* Annual Review of Psychology, 2009. **60**: p. 693-716.
19. Allport, G.W., *The historical background of social psychology*, in *The handbook of social psychology*, G. Lindzey and E. Aronson, Editors. 1985, McGraw Hill: New York.
20. Lewin, K., *The conceptual representation and measurement of psychological forces.* Contrib. Psychol. Theory, 1938. **1**: p. 247.
21. Lewin, K., *Field Theory in Social Science*. 1951, New York: Harper and Brothers.
22. Trope, Y., *Identification and Inferential Processes in Dispositional Attribution.* Psychological Review, 1986. **93**(3): p. 239-257.
23. Chaiken, S. and Y. Trope, *Dual-Process Theories in Social Psychology*. 1999, New York: Guilford Press.
24. Satpute, A.B. and M.D. Lieberman, *Integrating automatic and controlled processes into neurocognitive models of social cognition.* Brain Research, 2006. **1079**: p. 86-97.
25. Adolphs, R., *Recognizing Emotion From Facial Expressions: Psychological and Neurological Mechanisms.* Behavioral and Cognitive Neuroscience Reviews, 2002. **1**(1): p. 21-62.
26. Scherer, K.R., *Psychological Models of Emotion*, in *The neuropsychology of emotion*, J.C. Borod, Editor. 2000, Oxford University Press: New York. p. 137-162.
27. Ekman, P., *Basic Emotions*, in *Handbook of Cognitive Emotion*, T. Dalgleish and M. Power, Editors. 1999, John Wiley & Sons, Ltd.: Sussex, U.K.
28. Russell, J.A., *A Circumplex Model of Affect.* Journal of Personality and Social Psychology, 1980. **39**(6): p. 1161-1178.
29. Rolls, E.T., *The brain and emotion*. 1999, New York: Oxford University Press.
30. Panksepp, J., *Affective Neuroscience*. 1998, New York: Oxford University Press.
31. Scherer, K.R., *On the nature and function of emotion: A component process approach*, in *Approaches to Emotion*, K.R. Scherer and P. Ekman, Editors. 1984, Lawrence Erlbaum: Hillsdale, NJ.
32. Bruce, V. and A. Young, *Understanding Face Recognition.* British Journal of Psychology, 1986. **77**: p. 305-327.
33. Gobbini, M.I. and J.V. Haxby, *Neural systems for recognition of familiar faces.* Neuropsychologia, 2007. **45**(1): p. 32-41.
34. Adolphs, R., *Neural systems for recognizing emotion.* Current Opinion in Neurobiology, 2002. **12**(2): p. 169-177.
35. Ekman, P., *Facial Expressions*, in *Handbook of Cognition and Emotion*, T. Dalgleish and M. Power, Editors. 1999, John Wiley & Sons Ltd.: New York.
36. Rolls, E.T., *Emotion Explained*. 2005, New York: Oxford University Press.
37. Sander, D., D. Grandjean, and K.R. Scherer, *A systems approach to appraisal mechanisms in emotion.* Neural Networks, 2005. **18**(4): p. 317-352.
38. Scherer, K.R., *Appraisal considered as a process of multi-level sequential checking*, in *Appraisal Processes in Emotion*, K.R. Scherer,

A. Schorr, and T. Johnstone, Editors. 2001, Oxford University Press: New York.

39. Sander, D., et al., *Interaction effects of perceived gaze direction and dynamic facial expression: Evidence for appraisal theories of emotion.* European Journal of Cognitive Psychology, 2007. **19**(3): p. 470-480.
40. Kim, H., et al., *Inverse amygdala and medial prefrontal cortex responses to surprised faces.* Neuroreport, 2003. **14**(18): p. 2317-2322.
41. Kim, H., et al., *Contextual modulation of amygdala responsivity to surprised faces.* Journal of Cognitive Neuroscience, 2004. **16**(10): p. 1730-1745.
42. Conty, L., et al., *When eye creates the contact! ERP evidence for early dissociation between direct and averted gaze motion processing.* Neuropsychologia, 2007. **45**: p. 3024-3037.
43. Hadjikhani, N., et al., *Pointing with the eyes: The role of gaze in communicating danger.* Brain and Cognition, 2008. **68**(1): p. 1-8.
44. Adolphs, R., et al., *A mechanism for impaired fear recognition after amygdala damage.* Nature, 2005. **433**(7021): p. 68-72.
45. Itier, R.J. and M. Batty, *Neural bases of eye and gaze processing: The core of social cognition.* Neuroscience and Biobehavioral Reviews, 2009. **33**(6): p. 843-863.
46. van Reekum, C.M., et al., *Gaze fixations predict brain activation during the voluntary regulation of picture-induced negative affect.* Neuroimage, 2007. **36**(3): p. 1041-1055.
47. Cristinzio, C., et al., *Integration of gaze direction and facial expression in patients with unilateral amygdala damage.* Brain, 2009.
48. N'Diaye, K., D. Sander, and P. Vuilleumier, *Interaction between facial emotion and gaze direction in the human amygdala: the role of self-relevance and expression intensity.* Emotion, 2009.
49. Dipellegrino, G., et al., *Understanding Motor Events - a Neurophysiological Study.* Experimental Brain Research, 1992. **91**(1): p. 176-180.
50. Achaibou, A., et al., *Simultaneous recording of EEG and facial muscle reactions during spontaneous emotional mimicry.* Neuropsychologia, 2008. **46**(4): p. 1104-1113.
51. Carr, L., et al., *Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(9): p. 5497-5502.
52. Singer, T. and C. Lamm, *The Social Neuroscience of Empathy.* Year in Cognitive Neuroscience 2009, 2009. **1156**: p. 81-96.
53. Lamm, C., C.D. Batson, and J. Decety, *The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal.* Journal of Cognitive Neuroscience, 2007. **19**(1): p. 42-58.
54. Danziger, N., I. Faillenot, and R. Peyron, *Can We Share a Pain We Never Felt? Neural Correlates of Empathy in Patients with Congenital Insensitivity to Pain.* Neuron, 2009. **61**(2): p. 203-212.
55. Phelps, E.A., et al., *Performance on indirect measures of race evaluation predicts amygdala activation.* Journal of Cognitive Neuroscience, 2000. **12**(5): p. 729-738.
56. Phelps, E.A., C.J. Cannistraci, and W.A. Cunningham, *Intact performance on an indirect measure of race bias following amygdala damage.* Neuropsychologia, 2003. **41**(2): p. 203-208.
57. Cunningham, W.A. and P.D. Zelazo, *Attitudes and evaluations: a social cognitive neuroscience perspective.* Trends in Cognitive Sciences, 2007. **11**(3): p. 97-104.
58. Krendl, A.C., et al., *The good, the bad, and the ugly: An fMRI investigation of the functional anatomic correlates of stigma.* Social Neuroscience, 2006. **1**(1): p. 5-15.
59. Harris, L.T. and S.T. Fiske, *Dehumanizing the lowest of the low - Neuroimaging responses to extreme out-groups.* Psychological Science, 2006. **17**(10): p. 847-853.
60. Wheeler, M.E. and S.T. Fiske, *Controlling racial prejudice - Social-cognitive goals affect amygdala and stereotype activation.* Psychological Science, 2005. **16**(1): p. 56-63.
61. Lieberman, M.D., et al., *An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals.* Nature Neuroscience, 2005. **8**(6): p. 720-722.
62. Cunningham, W.A., et al., *Separable neural components in the processing of black and white faces.* Psychological Science, 2004. **15**(12): p. 806-813.
63. Zeki, S., *The neurobiology of love.* Febs Letters, 2007. **581**(14): p. 2575-2579.
64. Eisenberger, N.I. and M.D. Lieberman, *Why rejection hurts: a common neural alarm system for physical and social pain.* Trends in Cognitive Sciences, 2004. **8**(7): p. 294-300.
65. Somerville, L.H., T.F. Heatherton, and W.M. Kelley, *Anterior cingulate cortex responds differentially to expectancy violation and social rejection.* Nature Neuroscience, 2006. **9**(8): p. 1007-1008.
66. Sander, D., et al., *Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody.* Neuroimage, 2005. **28**(4): p. 848-858.
67. Delplanque, S., et al., *Sequential Unfolding of Novelty and Pleasantness Appraisals of Odors.* Chemical Senses, 2009. **34**(3): p. E42-E43.
68. Grandjean, D. and K.R. Scherer, *Unpacking the cognitive architecture of emotion processes.* Emotion, 2008. **8**(3): p. 341-351.
69. Sander, D., J. Grafman, and T. Zalla, *The human amygdala: an evolved system for relevance detection.* Rev Neurosci, 2003. **14**(4): p. 303-16.
70. von Kriegstein, K., et al., *Interaction of face and voice areas during speaker recognition.* Journal of Cognitive Neuroscience, 2005. **17**(3): p. 367-376.
71. Allison, T., A. Puce, and G. McCarthy, *Social perception from visual cues: role of the STS region.* Trends in Cognitive Sciences, 2000. **4**(7): p. 267-278.
72. Mitchell, J.P., C.N. Macrae, and M.R. Banaji, *Dissociable medial prefrontal contributions to judgments of similar and dissimilar others.* Neuron, 2006. **50**(4): p. 655-663.

73. Young, L., et al., *The neural basis of the interaction between theory of mind and moral judgment.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(20): p. 8235-8240.
74. Thomson, J.J., *The Trolley Problem.* Yale Law Journal, 1985. **94**(6): p. 1395-1415.
75. Greene, J.D., et al., *An fMRI investigation of emotional engagement in moral judgment.* Science, 2001. **293**(5537): p. 2105-2108.
76. Greene, J.D., et al., *The neural bases of cognitive conflict and control in moral judgment.* Neuron, 2004. **44**(2): p. 389-400.
77. Koenigs, M., et al., *Damage to the prefrontal cortex increases utilitarian moral judgements.* Nature, 2007. **446**(7138): p. 908-911.
78. Singer, T., et al., *Empathic neural responses are modulated by the perceived fairness of others.* Nature, 2006. **439**(7075): p. 466-469.
79. Tabibnia, G. and M.D. Lieberman, *Fairness and cooperation are rewarding - Evidence from social cognitive neuroscience.* Social Cognitive Neuroscience of Organizations, 2007. **1118**: p. 90-101.
80. Fehr, E. and C.F. Camerer, *Social neuroeconornics: the neural circuitry of social preferences.* Trends in Cognitive Sciences, 2007. **11**(10): p. 419-427.
81. Metcalfe, J. and W. Mischel, *A hot/cool-system analysis of delay of gratification: Dynamics of willpower.* Psychological Review, 1999. **106**(1): p. 3-19.
82. Wittmann, M., D.S. Leland, and M.P. Paulus, *Time and decision making: differential contribution of the posterior insular cortex and the striatum during a delay discounting task.* Experimental Brain Research, 2007. **179**(4): p. 643-653.
83. Jackson, D.C., et al., *Suppression and enhancement of emotional responses to unpleasant pictures.* Psychophysiology, 2000. **37**(4): p. 515-522.
84. Gross, J.J., *Emotion regulation: Affective, cognitive, and social consequences.* Psychophysiology, 2002. **39**(3): p. 281-291.
85. Gross, J.J., *Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology.* Journal of Personality and Social Psychology, 1998. **74**(1): p. 224-237.
86. Ochsner, K.N., et al., *For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion.* Neuroimage, 2004. **23**(2): p. 483-499.
87. Kim, S.H. and S. Hamann, *Neural correlates of positive and negative emotion regulation.* Journal of Cognitive Neuroscience, 2007. **19**(5): p. 776-798.
88. Botvinick, M.M., *Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function.* Cognitive Affective & Behavioral Neuroscience, 2007. **7**(4): p. 356-366.
89. Aron, A.R., T.W. Robbins, and R.A. Poldrack, *Inhibition and the right inferior frontal cortex.* Trends in Cognitive Sciences, 2004. **8**(4): p. 170-177.
90. Ochsner, K.N. and J.J. Gross, *The cognitive control of emotion.* Trends in Cognitive Sciences, 2005. **9**(5): p. 242-249.
91. Levesque, J., et al., *Neural circuitry underlying voluntary suppression of sadness.* Biological Psychiatry, 2003. **53**(6): p. 502-510.
92. Goldin, P.R., et al., *The neural bases of emotion regulation: Reappraisal and suppression of negative emotion.* Biological Psychiatry, 2008. **63**(6): p. 577-586.
93. Lieberman, M.D., et al., *Putting feelings into words - Affect labeling disrupts amygdala activity in response to affective stimuli.* Psychological Science, 2007. **18**(5): p. 421-428.
94. Petrovic, P., et al., *Placebo in emotional processing - Induced expectations of anxiety relief activate a generalized modulatory network.* Neuron, 2005. **46**(6): p. 957-969.
95. Adolphs, R., *Cognitive neuroscience of human social behaviour.* Nature Reviews Neuroscience, 2003. **4**(3): p. 165-178.
96. Darwin, C., *The origin of species*. 1859, London: John Murray.
97. Parker, S.T. and K.R. Gibson, *Object Manipulation, Tool Use and Sensorimotor Intelligence as Feeding Adaptations in Cebus Monkeys and Great Apes.* Journal of Human Evolution, 1977. **6**(7): p. 623-641.
98. Aiello, L.C. and P. Wheeler, *The Expensive-Tissue Hypothesis - the Brain and the Digestive-System in Human and Primate Evolution.* Current Anthropology, 1995. **36**(2): p. 199-221.
99. Stephan, H., H. Frahm, and G. Baron, *New and Revised Data on Volumes of Brain Structures in Insectivores and Primates.* Folia Primatologica, 1981. **35**(1): p. 1-29.
100. Sawaguchi, T. and H. Kudo, *Neocortical Development and Social-Structure in Primates.* Primates, 1990. **31**(2): p. 283-289.
101. Barton, R. and A. Purvis, *Primate brains and ecology: Looking beneath the surface*, in *Current Primatology*, R. Anderson, B. Thierry, and N. Herrenschmidt, Editors. 1994, Unversity of Strasbourg Press: Strasbourg. p. 1-12.
102. Joffe, T.H. and R.I.M. Dunbar, *Visual and socio-cognitive information processing in primate brain evolution.* Proceedings of the Royal Society of London Series B-Biological Sciences, 1997. **264**(1386): p. 1303-1307.
103. Keverne, E.B., F.L. Martel, and C.M. Nevison, *Primate brain evolution: Genetic and functional considerations.* Proceedings of the Royal Society B-Biological Sciences, 1996. **263**(1371): p. 689-696.
104. Byrne, R., *The thinking primate*. 1995: Oxford University Press.
105. Pawlowski, B., C.B. Lowen, and R.I.M. Dunbar, *Neocortex size, social skills and mating success in primates.* Behaviour, 1998. **135**: p. 357-368.
106. Marino, L., *The relationship between gestation length, encephalization, and body weight in odontocetes.* Marine Mammal Science, 1997. **13**(1): p. 133-138.
107. Kluver, H. and P.C. Bucy, *Psychic blindness and other symptoms following bilateral*

temporal lobectomy in rhesus monkeys. American Journal of Physiology, 1937(119): p. 352-3.
108. Phelps, E.A. and J.E. LeDoux, *Contributions of the amygdala to emotion processing: From animal models to human behavior.* Neuron, 2005. **48**(2): p. 175-187.
109. Pinkham, A.E., et al., *Neural bases for impaired social cognition in schizophrenia and autism spectrum disorders.* Schizophrenia Research, 2008. **99**(1-3): p. 164-175.
110. Rilling, J.K., et al., *Neural correlates of social cooperation and non-cooperation as a function of psychopathy.* Biological Psychiatry, 2007. **61**(11): p. 1260-1271.
111. Leppanen, J.M., *Emotional information processing in mood disorders: a review of behavioral and neuroimaging findings.* Current Opinion in Psychiatry, 2006. **19**(1): p. 34-39.
112. Andersson, F., et al., *Impaired activation of face processing networks revealed by functional magnetic resonance imaging in 22q11.2 deletion syndrome.* Biological Psychiatry, 2008. **63**(1): p. 49-57.
113. Phelps, E.A., et al., *Extinction learning in humans: Role of the amygdala and vmPFC.* Neuron, 2004. **43**(6): p. 897-905.
114. Vuilleumier, P., et al., *Effects of attention and emotion on face processing in the human brain: An event-related fMRI study.* Neuron, 2001. **30**(3): p. 829-841.
115. Vuilleumier, P., et al., *Distant influences of amygdala lesion on visual cortical activation during emotional face processing.* Nature Neuroscience, 2004. **7**(11): p. 1271-1278.
116. Canli, T., et al., *Amygdala response to happy faces as a function of extraversion.* Science, 2002. **296**(5576): p. 2191-2191.
117. Wright, P. and Y.J. Liu, *Neutral faces activate the amygdala during identity matching.* Neuroimage, 2006. **29**(2): p. 628-636.
118. Schiller, D., et al., *A neural mechanism of first impressions.* Nature Neuroscience, 2009. **12**(4): p. 508-514.
119. Engell, A.D., J.V. Haxby, and A. Todorov, *Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala.* Journal of Cognitive Neuroscience, 2007. **19**(9): p. 1508-1519.
120. Lieberman, M.D., J.A. Jarcho, and A.B. Satpute, *Evidence-based and intuition-based self-knowledge: An fMRI study.* Journal of Personality and Social Psychology, 2004. **87**(4): p. 421-435.
121. Chakrabarti, B., E. Bullmore, and S. Baron-Cohen, *Empathizing with basic emotions: Common and discrete neural substrates.* Social Neuroscience, 2006. **1**(3-4): p. 364-384.
122. de Quervain, D.J.F., et al., *The neural basis of altruistic punishment.* Science, 2004. **305**(5688): p. 1254-1258.
123. Zink, C.F., et al., *Know your place: Neural processing of social hierarchy in humans.* Neuron, 2008. **85**(2): p. 273-283.
124. Fliessbach, K., et al., *Social comparison affects reward-related brain activity in the human ventral striatum.* Science, 2007. **318**: p. 1305-1308.
125. Zahn, R., et al., *Social concepts are represented in the superior anterior temporal cortex.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(15): p. 6430-6435.
126. Amodio, D.M. and C.D. Frith, *Meeting of minds: the medial frontal cortex and social cognition.* Nature Reviews Neuroscience, 2006. **7**(4): p. 268-277.
127. Rushworth, M.F.S., et al., *Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour.* Trends in Cognitive Sciences, 2007. **11**(4): p. 168-176.
128. Eisenberger, N.I., et al., *An experimental study of shared sensitivity to physical pain and social rejection.* Pain, 2006. **126**(1-3): p. 132-138.
129. Eisenberger, N.I., M.D. Lieberman, and K.D. Williams, *Does rejection hurt? An fMRI study of social exclusion.* Science, 2003. **302**(5643): p. 290-292.
130. Singer, T., et al., *Empathy for pain involves the affective but not sensory components of pain.* Science, 2004. **303**(5661): p. 1157-1162.
131. Rolls, E.T., *The orbitofrontal cortex.* Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 1996. **351**(1346): p. 1433-1443.
132. McClure, S.M., et al., *Neural correlates of behavioral preference for culturally familiar drinks.* Neuron, 2004. **44**(2): p. 379-387.
133. Camille, N., et al., *The involvement of the orbitofrontal cortex in the experience of regret.* Science, 2004. **304**(5674): p. 1167-1170.
134. Shamay-Tsoory, S.G., et al., *Characterization of empathy deficits following prefrontal brain damage: The role of the right ventromedial prefrontal cortex.* Journal of Cognitive Neuroscience, 2003. **15**(3): p. 324-337.
135. Milne, E. and J. Grafman, *Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping.* Journal of Neuroscience, 2001. **21**(12): p. art. no.-RC150.
136. Blanke, O., et al., *Neuropsychology: Stimulating illusory own-body perceptions - The part of the brain that can induce out-of-body experiences has been located.* Nature, 2002. **419**(6904): p. 269-270.
137. Gilboa, A., et al., *Remembering our past: Functional neuroanatomy of recollection of recent and very remote personal events.* Cerebral Cortex, 2004. **14**(11): p. 1214-1225.
138. Bush, G., P. Luu, and M.I. Posner, *Cognitive and emotional influences in anterior cingulate cortex.* Trends in Cognitive Sciences, 2000. **4**(6): p. 215-222.
139. Etkin, A., et al., *Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala.* Neuron, 2006. **51**(6): p. 871-882.
140. Ochsner, K.N., et al., *Rethinking feelings: An fMRI study of the cognitive regulation of emotion.* Journal of Cognitive Neuroscience, 2002. **14**(8): p. 1215-1229.
141. Ochsner, K.N., et al., *Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other.* Journal of Cognitive Neuroscience, 2004. **16**(10): p. 1746-1772.

142. Hariri, A.R., S.Y. Bookheimer, and J.C. Mazziotta, *Modulating emotional responses: effects of a neocortical network on the limbic system.* Neuroreport, 2000. **11**(1): p. 43-48.
143. Ullsperger, M. and D.Y. von Cramon, *Neuroimaging of performance monitoring: Error detection and beyond.* Cortex, 2004. **40**(4-5): p. 593-604.
144. Harris, L.T., A. Todorov, and S.T. Fiske, *Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind.* Neuroimage, 2005. **28**(4): p. 763-769.
145. Mitchell, J.P., C.N. Macrae, and M.R. Banaji, *Encoding-specific effects of social cognition on the neural correlates of subsequent memory.* Journal of Neuroscience, 2004. **24**(21): p. 4912-4917.
146. Craig, A.D., *Interoception: the sense of the physiological condition of the body.* Current Opinion in Neurobiology, 2003. **13**(4): p. 500-505.
147. Kanwisher, N., J. McDermott, and M.M. Chun, *The fusiform face area: A module in human extrastriate cortex specialized for face perception.* Journal of Neuroscience, 1997. **17**(11): p. 4302-4311.
148. Gauthier, I., et al., *Expertise for cars and birds recruits brain areas involved in face recognition.* Nature Neuroscience, 2000. **3**(2): p. 191-197.
149. Reddy, L. and N. Kanwisher, *Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention.* Current Biology, 2007. **17**(23): p. 2067-2072.
150. Schwarzlose, R.F., C.I. Baker, and N. Kanwisher, *Separate face and body selectivity on the fusiform gyrus.* Journal of Neuroscience, 2005. **25**(47): p. 11055-11059.
151. Peelen, M.V. and P.E. Downing, *Within-subject reproducibility of category-specific visual activation with functional MRI.* Human Brain Mapping, 2005. **25**(4): p. 402-408.
152. Vrtička, P., et al., *Individual attachment style modulates human amygdala and striatum activation during social appraisal.* PLoS ONE, 2008. **3**(8): p. e2868.
153. Sanfey, A.G., et al., *The neural basis of economic decision-making in the ultimatum game.* Science, 2003. **300**(5626): p. 1755-1758.
154. Singer, T., et al., *Brain responses to the acquired moral status of faces.* Neuron, 2004. **41**(4): p. 653-662.
155. Finger, E.C., et al., *Caught in the act: The impact of audience on the neural response to morally and socially inappropriate behavior.* Neuroimage, 2006. **33**(1): p. 414-421.
156. Nawa, N.E., et al., *Do you make a difference Social context in a betting task.* Social Cognitive and Affective Neuroscience, 2008. **3**(4): p. 367-376.
157. Hariri, A.R., et al., *The amygdala response to emotional stimuli: A comparison of faces and scenes.* Neuroimage, 2002. **17**(1): p. 317-323.
158. Britton, J.C., et al., *Neural correlates of social and nonsocial emotions: An fMRI study.* Neuroimage, 2006. **31**(1): p. 397-409.
159. Norris, C.J., et al., *The interaction of social and emotional processes in the brain.* Journal of Cognitive Neuroscience, 2004. **16**(10): p. 1818-1829.
160. Goossens, L., et al., *Selective Processing of Social Stimuli in the Superficial Amygdala.* Human Brain Mapping, 2009.
161. Colden, A., M. Bruder, and A. Manstead, *Human content in affect-inducing stimuli: A secondary analysis of the international affective picture system.* Motivation and Emotion, 2008. **32**(4): p. 260-269.
162. Hart, A.J., et al., *Differential response in the human amygdala to racial outgroup vs ingroup face stimuli.* Neuroreport, 2000. **11**(11): p. 2351-2355.
163. Spielberger, C.D., *Manual for the State-Trait Anxiety Inventory*
Palo-Alto: Consulting Psychologists Press, 1983.
164. Spielberger, C.D., Sydeman, S. J., Owen, A. E., & Marsh, B. J., *Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI).* In M. D. Maruish (Ed.), The use of psychological testing for treatment planning and outcomes assessment (2nd ed.) (pp. 993-1021). Mahwah: Laurence Erlbaum, 1999.
165. Etkin, A. and T.D. Wager, *Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia.* American Journal of Psychiatry, 2007. **164**: p. 1476-1488.
166. Bishop, S.J., *Trait anxiety and impoverished prefrontal control of attention.* Nature Neuroscience, 2009. **12**(1): p. 92-98.
167. Goldin, P.R., et al., *Neural Bases of Social Anxiety Disorder Emotional Reactivity and Cognitive Regulation During Social and Physical Threat.* Archives of General Psychiatry, 2009. **66**(2): p. 170-180.
168. Killgore, W.D.S. and D.A. Yurgelun-Todd, *Social anxiety predicts amygdala activation in adolescents viewing fearful faces.* Neuroreport, 2005. **16**(15): p. 1671-1675.
169. Wright, C.I., et al., *Neuroanatomical correlates of extraversion and neuroticism.* Cerebral Cortex, 2006. **16**(12): p. 1809-1819.
170. Kumari, V., et al., *Neuroticism and brain responses to anticipatory fear.* Behavioral Neuroscience, 2007. **121**(4): p. 643-652.
171. Schneider, T.R., *The role of neuroticism on psychological and physiological stress responses.* Journal of Experimental Social Psychology, 2004. **40**(6): p. 795-804.
172. Haas, B.W., et al., *Emotional conflict and neuroticism: Personality-dependent activation in the amygdala and subgenual anterior cingulate.* Behavioral Neuroscience, 2007. **121**(2): p. 249-256.
173. Harenski, C.L., S.H. Kim, and S. Hamann, *Neuroticism and psychopathy predict brain activation during moral and nonmoral emotion regulation.* Cognitive Affective & Behavioral Neuroscience, 2009. **9**(1): p. 1-15.
174. Hooker, C.I., et al., *The influence of personality on neural mechanisms of observational fear and reward learning.* Neuropsychologia, 2008. **46**(11): p. 2709-2724.

175. Perlman, S.B., et al., *Individual differences in personality predict how people look at faces.* PLoS ONE, 2009. **4**(6).

176. Mobbs, D., et al., *Personality predicts activity in reward and emotional regions associated with humor.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(45): p. 16502-16506.

177. Kross, E., et al., *Neural dynamics of rejection sensitivity.* Journal of Cognitive Neuroscience, 2007. **19**(6): p. 945-956.

178. Burklund, L.J., N.I. Eisenberger, and M.D. Lieberman, *The face of rejection: Rejection sensitivity moderates dorsal anterior cingulate activity to disapproving facial expressions.* Social Neuroscience, 2007. **2**(3-4): p. 238-253.

179. Ainsworth, M.D.S., *Bowlby-Ainsworth Attachment Theory.* Behavioral and Brain Sciences, 1978. **1**(3): p. 436-438.

180. Bowlby, J., *Attachment and loss*. 1969, New York,: Basic Books. v.

181. Bowlby, J., *Attachment and loss*. 2nd ed. 1982, New York: Basic Books. v.

182. Hazan, C. and P. Shaver, *Romantic Love Conceptualized as an Attachment Process.* Journal of Personality and Social Psychology, 1987. **52**(3): p. 511-524.

183. Bartholomew, K. and L.M. Horowitz, *Attachment Styles among Young-Adults - a Test of a 4-Category Model.* Journal of Personality and Social Psychology, 1991. **61**(2): p. 226-244.

184. Griffin, D. and K. Bartholomew, *Models of the Self and Other - Fundamental Dimensions Underlying Measures of Adult Attachment.* Journal of Personality and Social Psychology, 1994. **67**(3): p. 430-445.

185. Mikulincer, M. and P.R. Shaver, *Attachment in Adulthood: Structure, Dynamics, and Change*. 2007, New York: The Guilford Press.

186. Feeney, J.A. and S.M. Ryan, *Attachment Style and Affect Regulation - Relationships with Health Behavior and Family Experiences of Illness in a Student Sample.* Health Psychology, 1994. **13**(4): p. 334-345.

187. Chotai, J., et al., *Adolescent attachment styles and their relation to the temperament and character traits of personality in a general population.* European Psychiatry, 2005. **20**(3): p. 251-259.

188. Kurdek, L.A., *On being insecure about the assessment of attachment styles.* Journal of Social and Personal Relationships, 2002. **19**(6): p. 811-834.

189. Mikulincer, M., O. Gillath, and P.R. Shaver, *Activation of the attachment system in adulthood: Threat-related primes increase the accessibility of mental representations of attachment figures.* Journal of Personality and Social Psychology, 2002. **83**(4): p. 881-895.

190. Fraley, R.C., et al., *Adult attachment and the perception of emotional expressions: Probing the hyperactivating strategies underlying anxious attachment.* Journal of Personality, 2006. **74**(4): p. 1163-1190.

191. Dewitte, M., et al., *What's in a name? Attachment-related attentional.* Emotion, 2007. **7**(3): p. 535-545.

192. Dewitte, M., et al., *Attentive processing of threat and adult attachment: A dot-probe study.* Behaviour Research and Therapy, 2007. **45**(6): p. 1307-1317.

193. Rognoni, E., et al., *Relationship between adult attachment patterns, emotional experience and EEG frontal asymmetry.* Personality and Individual Differences, 2008. **44**(4): p. 909-920.

194. Gillath, O., et al., *Attachment-style differences in the ability to suppress negative thoughts: Exploring the neural correlates.* Neuroimage, 2005. **28**(4): p. 835-847.

195. Sonnby-Borgstrom, M. and P. Jonsson, *Dismissing-avoidant pattern of attachment and mimicry reactions at different levels of information processing.* Scandinavian Journal of Psychology, 2004. **45**(2): p. 103-113.

196. Gobbini, M.I., et al., *Social and emotional attachment in the neural representation of faces.* Neuroimage, 2004. **22**(4): p. 1628-1635.

197. Leibenluft, E., et al., *Mothers' Neural Activation in Response to Pictures of Their Children and Other Children.* Biological Psychiatry, 2004. **56**(4): p. 225-232.

198. Todorov, A., et al., *Spontaneous retrieval of affective person knowledge in face perception.* Neuropsychologia, 2007. **45**(1): p. 163-173.

199. Simpson, J.A., *Influence of Attachment Styles on Romantic Relationships.* Journal of Personality and Social Psychology, 1990. **59**(5): p. 971-980.

200. Simpson, J.A., W.S. Rholes, and J.S. Nelligan, *Support Seeking and Support Giving within Couples in an Anxiety-Provoking Situation - the Role of Attachment Styles.* Journal of Personality and Social Psychology, 1992. **62**(3): p. 434-446.

201. Riggs, S.A., et al., *Attachment, personality, and psychopathology among adult inpatients: Self-reported romantic attachment style versus Adult Attachment Interview states of mind.* Development and Psychopathology, 2007. **19**(1): p. 263-291.

202. Brosch, T., et al., *Beyond fear - Rapid spatial orienting toward positive emotional stimuli.* Psychological Science, 2008. **19**(4): p. 362-370.

203. Brosch, T., D. Sander, and K.R. Scherer, *That baby caught my eye... Attention capture by infant faces.* Emotion, 2007. **7**(3): p. 685-689.

204. Abraham, A., et al., *Minds, persons, and space: An fMRI investigation into the relational complexity of higher-order intentionality.* Consciousness and Cognition, 2008. **17**(2): p. 438-450.

205. Sergerie, K., M. Lepage, and J.L. Armony, *A process-specific functional dissociation of the amygdala in emotional memory.* Journal of Cognitive Neuroscience, 2006. **18**(8): p. 1359-1367.

206. Glascher, J. and R. Adolphs, *Processing of the arousal of subliminal and supraliminal emotional stimuli by the human amygdala.* Journal of Neuroscience, 2003. **23**(32): p. 10274-10282.

207. Suslow, T., et al., *Attachment avoidance modulates neural response to masked facial emotion.* Human Brain Mapping, 2009.

208. Mikulincer, M., I. Orbach, and D. Iavnieli, *Adult attachment style and affect regulation: Strategic variations in subjective self-other similarity.* Journal of Personality and Social Psychology, 1998. **75**(2): p. 436-448.
209. Fujita, F., E. Diener, and E. Sandvik, *Gender Differences in Negative Affect and Well-Being - the Case for Emotional Intensity.* Journal of Personality and Social Psychology, 1991. **61**(3): p. 427-434.
210. Lemche, E., et al., *Human attachment security is mediated by the amygdala: Evidence from combined fMRI and psychophysiological measures.* Human Brain Mapping, 2006. **27**(8): p. 623-635.
211. Cohen, M.X. and P.R. Shaver, *Avoidant attachment and hemispheric lateralisation of the processing of attachment- and emotion-related words.* Cognition & Emotion, 2004. **18**(6): p. 799-813.
212. Searle, B. and N.M. Meara, *Affective dimensions of attachment styles: Exploring self-reported attachment style, gender, and emotional experience among college students.* Journal of Counseling Psychology, 1999. **46**(2): p. 147-158.
213. Tidwell, M.C.O., H.T. Reis, and P.R. Shaver, *Attachment, attractiveness, and social interaction: A diary study.* Journal of Personality and Social Psychology, 1996. **71**(4): p. 729-745.
214. Edelstein, R.S., *Attachment and emotional memory: Investigating the source and extent of avoidant memory impairments.* Emotion, 2006. **6**(2): p. 340-345.
215. Domes, G. and M. Heinrichs, *Oxytocin and Social Cognition in Humans Behavior and Functional Imaging.* Biological Psychiatry, 2009. **65**(8): p. 590.
216. Domes, G., et al., *Oxytocin attenuates amygdala responses to emotional faces regardless of valence.* Biological Psychiatry, 2007. **62**: p. 1187-1190.
217. Rimmele, U., et al., *Oxytocin Makes a Face in Memory Familiar.* Journal of Neuroscience, 2009. **29**(1): p. 38-42.
218. Theodoridou, A., et al., *Oxytocin and social perception: Oxytocin increases perceived facial trustworthiness and attractiveness.* Hormones and Behavior, 2009. **56**(1): p. 128-132.
219. Hermans, E.J., N.F. Ramsey, and J. van Honk, *Exogenous testosterone enhances responsiveness to social threat in the neural circuitry of social aggression in humans.* Biological Psychiatry, 2008. **63**(3): p. 263-270.
220. Van Honk, J., *Effects of Testosterone Administration on Mind Reading and Trust in Humans.* Psychophysiology, 2009. **46**: p. S3-S3.
221. Gillath, O., et al., *Genetic correlates of adult attachment style.* Personality and Social Psychology Bulletin, 2008. **34**(10): p. 1396-1405.
222. Emanuele, E., et al., *Genetic loading on human loving styles.* Neuroendocrinology Letters, 2007. **28**(6): p. 815-821.