

Ethical Know-How and the Ethical World

Towards an embodied, situated account of ethical agency

Thijs Heijmeskamp

Master Thesis

Philosophy of Man and Culture

Faculty of Philosophy

Erasmus University Rotterdam

Date: August 14, 2017
Student number: 318651
Supervisor: Prof. dr. M.M.S.K. Sie
Advisor: Prof. dr. F.A. Muller
Number of words [including notes]: 31.774

“Ethics is closer to wisdom than to reason, closer to understanding what is good than to correctly adjudicating particular situations.”

Francisco J. Varela (1992: 3)

“If a man lived alone in the world there might be some sense in the question “Why be moral?” were it not for one thing: No such question would then arise.”

John Dewey (1922: 326)

Contents

- Introduction iv
- Thesis Outline vi
- Abbreviations vii
- List of Figures..... vii
- 1. Moral Emotions 1**
 - 1.1 The Standard View of Moral Agency 1
 - 1.2 Psychopaths 3
 - 1.3 Moral Dumbfounding..... 6
 - 1.4 Moral Mary 7
- 2. Moral Psychology 10**
 - 2.1 Greene’s Dual-Process Theory 12
 - 2.1.1 Dual-Process and Dual-Systems Theory 13
 - 2.1.2 The Neuropsychological Basis of Emotions and Cognition 14
 - 2.2 Social Intuitionist Model 17
 - 2.2.1 SIM as a Dual-Process theory 17
 - 2.2.2 Control and Unconscious Processes 19
 - 2.3 Awareness and Conscious Control 21
- 3. Embodied Cognition, Emotions and the Social Mind 23**
 - 3.1 Enactivism 23
 - 3.2 Embodied emotions 26
 - 3.2.1 Emotions as Self-Organisation 26
 - 3.2.2 Embodied Appraisal 27
 - 3.3 The Socially Extended Mind 28
 - 3.3.1 Mental Institutions 28
 - 3.3.2 Participatory Sense-Making..... 31
- 4. Ethical Know-How 35**
 - 4.1 Two Models of Moral Agency 37
 - 4.1.1 The Linear Model of Moral Agency 38
 - 4.1.2 The Dynamic Model of Ethical Agency 40
 - 4.2 Ethics as a Skill..... 42
 - 4.3 The Ethical World and the Moral Agent 48
- 5. Conclusions 51**
- References..... 53

0. Introduction

Morality pervades the entirety of human life, by structuring our thoughts, feelings and behaviour. It is a subject that has occupied philosophers since the days of classical antiquity. Human society and life seems unthinkable without our moral practices. Ethics come natural to us and we deeply care about ethics. Currently, ethics captures the interests of not only the philosophers, but also scientists working within the cognitive sciences. This has led to a fruitful and interesting dialogue between disciplines.

The subject of this thesis concerns how we should understand moral agency. Despite many different views on moral agency, a moral agent is generally considered to be someone who acts on the basis of moral reasons and is in conscious control of his actions. This I call the *traditional view of moral agency*; it describes an important aspect of our moral life. We sometimes find ourselves reflecting on what a moral course of action would be. However, this conceptualisation of moral agency has come under attack from those working within the field of moral psychology. Experiments on moral agency show fascinating results about conscious and unconscious processes, which according to some undermine this traditional perspective. For instance, Jonathan Haidt (2001) claims that “[p]eople have quick and automatic moral intuitions, and when called on to justify these intuitions they generate post hoc justifications [...]” (823) This means that our moral judgements are caused by unconscious processes and our reasons are after the fact confabulations. Since moral agency requires conscious control and reasons for action, we are therefore no moral agents.

Although I do not share this conclusion, Haidt’s findings are insightful and do seem to capture something that is lacking in the traditional view. Important elements of our ethical life seem to escape the traditional view of moral agency. In our everyday life, when we walk in the street and we see someone fall over in a bad way our immediate reaction is to help. One can point out that you have a good reason for helping someone who fell over, but that is not an accurate description of what is happening. There is an immediacy in the response, which does not necessarily involve discursive thought. Another example may clarify this. You are sitting with friends in the sun having a lively conversation. Suddenly a topic comes up that embarrasses one of your friends. Perceiving the embarrassment, you steer the topic away with a humorous remark.¹ These actions are best described as moments of *immediate coping*. The action immediately follows from the situation, seemingly without the intervention of a rational agent who first has to deliberate as described by the standard view.

In this thesis, I propose an *embodied, situated conception of moral agency* in order to include actions best described as immediate coping, even though the standard model of moral agency captures an important element of the moral domain. This is also called a skill based model of ethical expertise. In this I join a long tradition of philosophers and scientists who think about habits, skills and flow and this model stands in direct relation with the works of John Dewey, Ludwig Wittgenstein, Hubert Dreyfus, and Francisco J. Varela. Although not all of these philosophers are discussed explicitly, their influence is undeniable on this thesis and on many of the other authors which are referenced.

This thesis can be divided into two parts. In the first two chapters, I argue how the traditional view of moral agency misses out on certain aspects of moral agency. One such way is that the traditional view does not seem to provide a central and productive role to emotions. It does not necessarily deny that emotions play a part in the moral domain. It is

¹ I take these examples from Varela (1992: 5)

however our rational reasoning that makes an action moral. The limitations of the traditional view become apparent when examining the cases where we act against our judgements in a specific moral action on the basis of certain (moral) emotions. There is also a large body of empirical evidence that suggest that emotions are necessary for moral agency.

I argue that the standard view of moral agency seems to exclude certain aspects of moral agency, because the standard view holds three assumptions that become problematic for moral agency in light of cases where we act against our judgements. First, a controlled process is conscious and only our controlled processes can be considered rational. This stand in contrast to automatic processes, which are unconscious. Second, our emotions are the result of automatic processes and our moral deliberations are conscious processes. Third, our moral evaluations are the result of either conscious processes or automatic processes. In examining evidence from the research in the cognitive sciences, I find good reasons for discarding these assumptions. In this I do not discard the traditional view of moral agency or deny that we are rational agents. In discarding these assumptions, I believe it becomes possible to rethink moral agency in such a way that it is more inclusive of those aspects that escape the traditional view. I end the first part of this thesis with the conclusion that moral thinking unjustifiably is seen as distinct from emotion and affect. However, in experience thinking and feeling are strongly connected. This means, in my view, we can relate in a moral way to the world, without necessarily involving discursive thoughts.

The second part of this thesis forms the positive part of this project. In the second part, I bring together literature on embodiment, cognition and ethics in order to rethink moral agency. In effect I am building on Francisco J. Varela's *Ethical Know-How* (1992), where he offers a first attempts to formulate a notion of moral agency from the perspective of the enactive approach. *The enactivist approach* is a programme best described as trying to unify the cognitive sciences with phenomenological insights from the philosophy of Maurice Merleau-Ponty. In this new view, cognition is not abstract symbolic manipulation, but affective and contextual, arising from the interaction between an acting embodied organism and its environment. Through the enactivist approach, I address the assumptions that were made explicit in the first part. In the enactivist view cognition is seen as the activity of sense-making, which entails not only the emergence of a (bodily) self, but also an environment as a place of significance and value. This notion of cognition no longer assumes a fundamental distinction between cognition and affect. If the interaction with an environment is constitutive for sense-making, then this should also hold for ethical sense-making. Colombetti and Torrence (2009) argue that ethical appraisal should be seen as a form of *participatory sense-making*, i.e. sense-making in the social domain through coordination with other people. I argue that ethics should be seen as a dynamical process of interaction with other ethical agents. This means that ethical sense-making should first and foremost be understood as a skill through which we engage with others. To be an ethical expert is to be a skilful participant of an ethical community. Drawing on Dreyfus and Varela, I argue that ethics also involves unreflective ethical know-how. I will conclude this thesis with the argument that within the social domain we enact an ethical world that structures our skilful social interactions. However, in doing this we are also shaped by the same (ethical) interactions.

This thesis does not contain a complete account of moral agency. Many of the topics addressed, such as autonomy and the interaction between moral reflection and immediate coping, deserve further discussion. This thesis should be read as a proposal for an approach to the question of what is moral agency, in which the ideas of the many authors here discussed are formulated in a single programme from which we can engage the longstanding debate of moral agency.

Thesis Outline

In chapter 1, I argue that the emotions play a central and productive role when it comes to moral agency. This role is undervalued in the traditional view of moral agency. I start with discussing the case of Huckleberry Finn, where Huck acts against his judgement and helps his friend Jim. Since his action is not based on moral reasons on a narrow construction of 'reasons', it is problematic to consider his actions as morally motivated from the perspective of the traditional view of moral agency. To further support the point that emotions play a central and productive role, I discuss two bodies of empirical research. Psychopaths who consistently act in an antisocial way not only show deficiencies in practical rationality, but also in emotional responses. Additionally, cases of *moral dumbfounding*, where a subject makes a moral judgement without adequate reasons, suggest that an emotional response is sufficient for moral evaluation. Moreover, I argue that emotions are important for moral experience. It is likely that without an appropriate emotional response, the behaviour of the other is unintelligible to us. Emotions sometimes also help us to improve our moral evaluations.

In chapter 2, I trace the limitations of the traditional view of moral agency to problematic assumptions about emotions and cognition, and controlled processes and automatic processes. I will discuss the moral psychological theories of Joshua Greene and Jonathan Haidt. I discuss these two authors in particular because they are influential and stand at the beginning of the *moral psychological* turn leaving a considerable influence on the debate. Although both wrote extensively on the subject of moral judgement, their work also provides a criticism of the standard view of moral agency, either arguing that it should be adjusted or that we are not the deliberative moral agents the standard view supposes. In this criticism they both rely extensively on (their) empirical research and put the standard view of moral agency into doubt. Examining these two authors serves two points. First, both authors show certain limitations of the standard view of moral agency. Second, despite the criticism of both authors to the standard view, they still share certain assumptions with the standard view - concerning the relationship between rationality and emotions, and the relationship between controlled and automatic processes – which become apparent when examining the empirical evidence. By closely examining the empirical evidence both Haidt and Greene appeal to, I aim to make these assumptions explicit and show there are good grounds for rejecting them when it comes to moral agency.

In chapter 3, I address the notion of appraisal. I believe that in order to rethink moral agency, we must first develop a notion of (ethical) appraisal that can account for cases like Huckleberry Finn. Huck did not deliberate in the sense of discursive thought and did not form the moral judgement that he should save his friend on the basis of deliberations. However, I do believe it's justified to say that Huck's feelings for his friend Jim can be considered a moral appraisal in the sense that Huck relates himself morally to the world. In order to arrive to this inclusive notion of appraisal, I first examine the enactivists conception of cognition as the *activity of sense-making*, which integrates cognition and affect. When an organism has cognitive capacities, it is *autonomous* in the sense that the organism not only self-organises itself but also creates an environment of meaning and value. Giovanni Colombetti enriches this account by discussing the particular role of emotions. According to her, emotions should be seen as *self-organizing* configurations of an organism, where the organism relates itself to the world. She argues that this means that there is no longer a fundamental distinction between arousal, appraisal and action. An appraisal is therefore, in Colombetti's view, an

evaluation of our relation with the world through the body. Since an emotion is already a cognitive process and all cognitive processes have an affective dimension, an appraisal always has an emotional aspect and this implies that ethical thinking or appraising involves emotions. I conclude with a discussion of the human cognition is not only *socially embedded*, but also *socially extended*. If our mind is socially extended, then so are our sense-making activities, and therefore our ethical thinking. De Jaegher and Di Paolo call activities of collective sense-making *participatory sense-making*.

In chapter 4, I will formulate an alternative notion of moral agency. I follow Colombetti and Torrence’s line of argument that an ethical appraisal should be seen as a form of participatory sense-making. Therefore, I argue that ethics and ethical sense-making should be seen as a *dynamic interaction* between two or more agents. This I contrast to what I call *the linear model of moral agency*, which conceives ethical sense-making as an individual input-output process. I argue that ethical agency must first and foremost be understood as a skill for interacting with others. I therefore apply the notion of participatory sense-making to the theories of Dreyfus and Varela, who both argued that ethics is skill and mainly involves ethical know-how. I conclude with a description of how ethical agents enact an ethical world, and as individuals are shaped by the ethical world. A consequence of this is that the definition of autonomy that has been used up till now should be adjusted, because that definition can only conceptualise autonomy at the level of the body and cannot account for the field of social interactions that are not fully part of the body, nor fully part of the (social) environment.

Abbreviations

DLPFC	Dorsolateral prefrontal cortex
DP-theory	Dual-process theory of morality
DST	Dynamical Systems theory
fMRI	functional Magnetic Resonance Imaging
PFC	Prefrontal cortex
SIM	Social Intuitionist model

List of Figures

The linear model of moral agency	39
The dynamic model of ethical agency	40

1. Moral Emotions

In the (philosophical) literature a moral agent is often viewed or conceptualised in the ideal case as autonomous, in conscious control, reflective in nature and acting on reasons. This I call the *standard view of moral agency*. According to the standard view, our capacity for moral agency lies in our rational capacities. This view does not seem to capture all aspects of moral agency. In this chapter, I address several empirical studies and philosophical views on the nature of emotions in ethics, which taken together suggest that our moral agency is not just dependent on our rational capacity but at least as much on our emotional capacities. By bringing together multiple discussions on or relating to moral agency, I want to argue for a positive role of emotions in the moral domain that seems to be lacking in the standard view and that is in agreement with the empirical evidence. In order to understand humans as moral agents, you also have to understand humans as emotional agents. This chapter will show that any account of moral agency needs to integrate the emotions and reasons; I explore such an account in the latter chapters.

I first discuss the standard view of moral agency and its limited view on ethical life (§1.1). Next, I discuss recent research into psychopathy (§1.2). Psychopaths consistently display a pattern of irresponsible and antisocial behaviour and thus immoral behaviour. Psychopaths show not only deficiencies in practical reasoning, but also lack important social and moral emotions. This suggests that both (practical) reasoning and emotions are constitutive of moral agency. Furthermore, I discuss the phenomenon of *moral dumbfounding* (§1.3). Moral dumbfounding is the phenomenon that people hold on to a moral attitude without any supporting reasons. Moral dumbfounding experiments allegedly show that in actual practice an emotional response is sufficient to cause a moral attitude. Finally, I conclude this chapter with two examinations of moral experience (§1.4). The first is the thought experiment of Moral Mary, who has no innate moral attitudes. Although Mary would be in principle capable of making a moral appraisal, she would have a hard time functioning in our every-day moral world. The second examination is a return to the case of Huckleberry Finn. This case is an example where our emotions can improve our moral appraisals and can even be considered rational.

1.1 The Standard View of Moral Agency

It is safe to say that we often think of ourselves as deliberative agents. This idea assumes that our conscious intentions (or will), and our reasons have an impact upon how we act. At the basis of the standard view of moral agency is the notion of agent-control; i.e., the view that “actions are subject to agent-control if and only if they are the products of acts of will.” (Pettit, 2007: 77) This means that an agent can initiate an act on demand or will an act as a conscious choice. We can think about what is the right thing to do, i.e., we are aware of our deliberations, or how to judge a situation, and we can act upon those deliberation. It is this conscious control that makes us moral agents; consciousness is required in order to be autonomous and morally responsible. (Clark, 2013: 239) It generally makes a difference in how we judge persons if, for instance, a person knocks down a precious vase on purpose or by accident. Those adhering to the standard view are the first to admit that we do not always act on reasons or that not every action is preceded by a process of conscious deliberation. Still, many believe that our deliberative awareness and our capacity for conscious control are

the abilities that make us moral agents. (Sie, 2009: 519) Even if we do not always engage in moral reasoning, it is our capacity for moral reasoning that makes us moral agents. This does not mean that any action has to be preceded by a process of deliberation, but that we should always be able to know or reconstruct our reasons for a particular action, and that those reasons take the form as if they are directly the result of a process of deliberation. Thus according to the standard view, we are genuine moral agents when we act upon moral reasons and an action is moral when the action is performed on the basis of a moral reason. For a moral action, we must not only have conscious control over what we are doing, but we must also be aware of the reasons we have for that particular action. Therefore, a moral agent must first be understood (or idealised) as a reflective agent, who engages in moral reasoning.

However, there are certain aspects of our ethical life that are not directly captured by the standard view of moral agency. There are many cases that the standard view would characterise as irrational or not even containing morally motivated actions, whereas they clearly count as moral. To illustrate this, consider the case of Mark Twain's Huckleberry Finn. At one point in this story, Huck tries to help his friend Jim run away from his slave owners. However, he has a bout of conscience. In Huck's world helping a slave escape is akin to stealing property from the slaveholders. There is no question for Huck that assisting slaves in attaining freedom is considered to be morally wrong. Since Huck believes this, he feels guilty. He believes that turning his friend Jim in to the authorities is the right thing to do. So when the opportunity arrives Huck pretends to reconnoitre, while his real aim is to turn Jim in. When he encounters slave hunters he finds himself unable to give up his friend and instead lies to slave hunters in order to save his friend. He sees his compassion for Jim as a weak, ignorant and wicked felony.

Jonathan Bennett (1974) - who views this case as a conflict between emotion and judgement – describes Huck's action as: "...he simply *fails* to do what he believes to be right". (4) From the standpoint of the standard view, Huck's actions cannot be understood as morally motivated. In the standard view we can only ascribe moral motivations to people if those motivations are derived from conscious reasons. However, it seems correct to claim that Huck acts with a *moral* motive, but he does not know that it is a moral motive. He has come to see his friend Jim not as property of someone else, but as a human being with needs and wants. He does fail to adjust his beliefs accordingly.

When we construe moral agency as the ability to form beliefs by reflection and practical judgement by practical deliberation, it leads to missing out important aspects of the ethical life and what it means to be a moral agent. One of those aspects is an adequate description of how our emotions relate to the moral domain. In the remainder of this chapter I examine the role emotions as it is presented by other philosophers and empirical research.

Not all versions of the standard view deny that emotions play a role, but the emotions do not play a constitutive role. Huck's feelings for his friend Jim cannot be called a judgement in the ordinary sense. Therefore I try to avoid speaking of moral judgement and instead mainly speak of moral *appraisal*. An appraisal is somewhat akin to a judgement but does not share the same rational commitments and is therefore more inclusive of examples like the case of Huckleberry Finn.²

² Appraisal must first and foremost be understood in an inclusive sense. Appraisal refers to all forms of moral evaluation and also to moral judgements. Therefore, my use of appraisal is neither meant to deny the existence of moral judgements, nor is it meant as a definition of judgement. It can be said that a judgement is a form of appraisal but not all appraisals are judgements. Although it is common in the fields of metaethics and moral

At the turn of the millennium, a new interdisciplinary approach to ethics gained ground in the form of *moral psychology*. Moral psychology can be characterised twofold: first, as empirically scrutinizing claims from philosophy. Second, as the engagement of psychology with important philosophical considerations.³ Hence there has been quite a lot of empirical research into the role of emotion in morality. In the remainder of this chapter I discuss, besides philosophical arguments, some of the more important empirical findings that suggest that there is an intimate connection between morality and emotions. I start with a discussion on research into psychopaths and persons with lesions in their ventromedial frontal cortices, which has interesting implications for the role of emotions in not only our functioning as moral agents, but also our moral development.

1.2 Psychopaths

In this section, I examine some of the research into those who closely resemble amoralists. How we conceptualize humans as moral agents is generally considered to be dependent upon which capacities we believe to be necessary for moral behaviour. In other words, how human agency is characterised depends on whether morality is seen as having a rational foundation or a sentimentalist foundation. The reason for this is that, if our moral appraisals are expressions of emotions, then any emotional impairment could cause a moral impairment. If our moral appraisals are judgements based on reasons, then any rational impairment could cause a moral impairment. Thus someone who is an amoralist - i.e. a person who is unable to be a capable moral actor, because such a person either lacks the capacity for morality or the motivation to behave morally – would tell us a lot about our underlying moral capacities. I hold that emotions are indispensable for our moral practices. If emotions are indispensable, a pathological amoralist should display (some) emotional impairments.

Psychopaths come closest to amoralism. Psychopathy is defined as a pattern of irresponsible and antisocial behaviour. This behaviour is linked to a lack of emotional responses, as psychopaths show an early onset of aggressive behaviour that is not tempered by feelings of guilt towards or empathy with the victim. Furthermore, psychopaths are distinctively emotionally impaired in the sense that they do not display typical emotional responses when hearing words pertaining to violence. (Grey et al., 2003) Psychopaths display normal response to threat cues (i.e. a picture of a shark), but have reduced responses to distressed cues (i.e. a picture of a crying child). (Green, 2014: 200) Psychopaths also perform poorly in what is referred to as the 'moral/conventional' distinction. This distinction between moral and conventional is something most of us intuitively feel and make. (Blair, 1995: 20-21; Nucci, 1981) Moral rules differ from other rules. We judge someone who does not shake your hand differently than a person who murders a child. Transgressions of a nonconventional norm are treated more seriously and the validity of moral norms seems not to be based upon authority. (Sauer, 2011: 98) The ability to make a distinction between moral and conventional norms is seen by many authors as crucial for the ability to make moral evaluations. Any full account of moral agency must take into account the evidence from research into psychopathy.

psychology to speak of judgements, I think there is much to be gained if we instead see our moral evaluations primarily as appraisals. I will return to the subject of appraisal in chapter 3 and moral appraisal in chapter 4.

³ Specific theories from moral psychology or heavily influenced by moral psychology will be addressed in Section 2.

James Blair (1995) showed how exactly psychopaths fail to make the moral/conventional distinction. This impairment in making a moral/conventional distinction becomes apparent in the justifications psychopaths give of why certain transgressions were morally or conventionally wrong. First, their justification of why a certain norm transgression is wrong is less likely to make references to the pain and discomfort of the victims of that transgression, in comparison with justifications for the wrongness of the same norm transgression given by people from a control group. (idem: 13) Thus there is a tendency among psychopaths to ignore the victim's welfare when asked to justify their answers about moral wrongs. A second, and more surprising finding, was that psychopaths tend to judge all violations of conventional norms also as moral. (idem: 23) Research subjects suffering from psychopathy were presented with stories such as a child hitting another child (moral transgression) and a child talking in class even though the teacher instructed the child not to (conventional transgression). As expected, the psychopaths did not make any distinction between the two kinds of story. Surprisingly, the psychopaths tended to judge all the violations as moral violations and authority independent. At first, this outcome was problematic, because Blair predicted that the psychopaths were not responsive to moral cues and therefore unable to respond morally. However, Blair has argued convincingly that this does not mean that psychopaths have an increased moral sensitivity. First, the psychopaths were able to ignore the moral/conventional distinction, where the control group, which consisted of incarcerated nonpsychopaths, consistently responded to the distinction. (ibidem) Secondly, the psychopaths were incarcerated and therefore highly motivated to demonstrate that the treatment they received was effective in order to secure a release. Blair suggests that this desire in coupled with the inability to make the moral/conventional distinction led the psychopaths to judge all transgressions as moral. (ibidem)

Despite the role of emotional impairments in psychopaths, we cannot conclude from this that research into psychopathy demonstrates that rationality does not play a role in morality. Other criteria that are used to diagnose psychopathy are a lack of realistic long-term goals and impulsive decisions that get in the way of more important goals. They are unable to weigh the pros and cons of an action, not taking the possible consequences into account. Self-control is generally associated with the cognitive capacity to keep one's long-term goals in view and avoid factors that might undermine ones resolve. (Kenneth, 2006: 77) Psychopaths are also chronic liars. They cheat and manipulate when there is no chance of success. These deficiencies in practical rationality lead to irrational behaviour where psychopaths systematically act against their best interests. (Sauer, 2013: 241-242) Jeanette Kenneth (2006) argues that these impairments in practical rationality are also relevant for the moral domain. (77-78) One important aspect of self-control is inhibitory control or effortful control (i.e. the ability to alter ones behavioural response in order to better achieve a goal) which begin to emerge in the end of the first year of life and before children are able to make the moral/conventional distinction. Poor self-regulation is associated with conduct disorder and anti-social traits. Effortful control has also been strongly associated with the development of a conscience, which psychopaths notably lack. Research by Walter Mischell (1981) shows that there is a meaningful development of children's knowledge of self-control tactics.) Take, e.g., the following experiment that is now famously known as the marshmallow test; by the age of five or six children know that looking at a marshmallow or thinking about the two marshmallows they were promised if they leave the one marshmallow alone, makes waiting difficult. Once children develop the moral/conventional distinction, they are also beginning to develop the capacity to reflect on and evaluate their preferences and determine whether or not they provide meaningful action. Kenneth (2006) argues that research into the

development of self-control capacities, like the above, demonstrate that rational capacities are “as critical to human moral agency and moral judgements as the more basic capacity to recognise and respond affectively to distress in others which we share with many other animals.” (78)

The research into psychopathy does not resolve the debate over the role of emotions and rationality for moral agency, as a psychopath shows clear deficiencies in affectivity, such as a lack of empathy, but clear impairments in practical rationality as well. A sentimentalist, who believes emotions form the basis of morality, can argue that the deficient practical rationality in psychopaths is the result of them not being able to pick up affective cues from other humans. A rationalist, who believes our rational capacities form the basis of morality, could maintain that what is impaired in psychopaths is the capacity needed to connect unpleasant feelings with some of their own misdeeds. (Vujošević, 2014: 7) What is clear however is that psychopaths lack the capacities for moral emotions or have poor moral emotions, fail to consistently make the moral/conventional distinction⁴ and fail to be morally competent subjects. Therefore I believe that trying to understand psychopathy only from a rationalistic perspective is insufficient. This seems to suggest that the emotions play a significant and perhaps necessary (but not sufficient) part when it comes to moral agency.

The importance of emotions for moral agency also seems to be supported by research into individuals whose *ventromedial frontal cortices* were damaged and as a result showed sociopathic behaviour. The ventromedial prefrontal cortex is primarily associated with regulating emotions. Damasio, Tranel and Damasio (1990) studied patients with previously normal personalities, who after procuring brain damage displayed abnormal social conduct. The hypothesis is that these patients were unable to generate *somatic markers* due to damage to their emotional circuitry in their brain; in particular damage to their ventromedial frontal cortices. A somatic marker is a physiological state that is associated with an emotion and helps us in decision making. A somatic marker is generated when a person experiences punishment or rewards in a specific social situation. When a situation arises that is associated with a certain somatic marker, the marker allows us to quickly ascertain the situation as positive or negative (Bechara & Damasio, 2005: 346) Thus in the case of social interactions, these somatic markers allow a person to quickly and subconsciously determine whether a certain behaviour is appropriate or not; hence an inability to form a somatic marker would lead to a deteriorated moral character. One patient named EVR was a successful and happily married man who suffered a brain tumour. After surgery his basic intelligence and standard memory was not affected. However, his social conduct was heavily affected by the brain damage. EVR was unable to hold a job, divorced twice and was diagnosed with a sociopathic disorder. (idem: 82) There are also cases of early onset of damaged ventromedial frontal cortices. Anderson et al. (1999) examined subjects, who procured prefrontal cortex lesions before they were 16 months old. Those with early onset damage displayed the same impaired social behaviour like the cases where such damage occurs in adulthood despite having normal basic cognitive abilities. However, subjects with early onset damage also grew up with defective social and moral reasoning. This seems to suggest that

⁴ The research into psychopathy shows that both emotions and reasoning seemingly play a fundamental role in moral behaviour, even if one does not accept the moral/ conventional distinction. Although the distinction is widely accepted in the relevant literature, it is not without controversy. Most of the research discussed here consisted of research done on incarcerated psychopaths. However, there are cases where a psychopath is able to function successfully in a society and other studies suggest that even the tested inmates are capable of making the moral/conventional distinction. See for an overview Langdon & Delmas (2012).

the damage prevented the acquisition or impaired the acquisition of complex social conventions and moral rules. (ibidem)

To summarise, the research into psychopathy as well as the research into cases of people with damage to their ventromedial frontal cortices seem to suggest that emotions are not only necessary for our moral development, but also for our general moral functioning. The empirical evidence discussed here suggests that emotions play an important and constitutive role, that is, our reflective capabilities do not suffice for moral agency. Earlier I argued that the standard view is not inclusive enough to capture the whole spectrum of ethical life. Another way emotions are taken to be indispensable for moral agency is via their role in moral appraisal.

1.3 Moral dumbfounding

According to Jonathan Haidt emotionally charged intuitions can cause moral appraisals. He drew this conclusion after experimentations into what Murphy et al. call the phenomenon of *moral dumbfounding*.⁵ Moral dumbfounding is the phenomenon where people hold on to a moral attitude, without supporting reasons. (2000) Haidt, Murphy & Björkland presented research subjects with the following case:

“Julie and Mark are brother and sister. They are travelling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love?” (Haidt, 2001: 814)

American college students were asked to consider if this case was morally wrong. The majority of the research subjects regarded the actions of Julie and Mark as immoral and only 20% of the respondents thought their actions were ok. (Murphy et al., 2000) When asked for reasons, the people who judged it wrong pointed to the dangers of incest or that Julie and Mark would both be hurt, perhaps even emotionally. The researcher explained to each of the research subjects why their arguments failed. Eventually, when they were reminded that the story excluded these objections, most still held one to the initial judgment (now 32% of the respondents thought the actions of the actions were ok), claiming: ‘I don’t know, I can’t explain it. I just know it’s wrong.’ (Haidt, 2001: 814; Murphy et al., 2000)

In the case of moral dumbfounding people hold a moral attitude in the absence of any rational justification. Arguments are given, but most people who judged the actions of Julie and Mark as wrong, never wavered in their conviction, despite the fact that they were not able to provide adequate reasons in support of this judgment. They stick to their initial gut reaction. Haidt argues that since the reasons given by the participants no longer support the appraisal and the participants are still unwilling to change their attitude, the attitude cannot

⁵ Haidt’s claims that experiments into moral dumbfounding not only show that emotionally charged intuitions can cause moral attitudes, but all our moral attitudes are caused by the emotionally charged intuitions. This forms the basis of Haidt’s Social Intuitionist model of moral judgement (SIM). A more extensive discussion of the SIM is to be found in the next chapter.

have been caused by the reasons that were given. Therefore, the moral reasoning involved in cases of moral dumbfounding has to be *post hoc*. (Haidt, 2001: 817) The reasons that were given cannot constitute the moral attitude and are only given to support the initial judgment. Other cases where used that contained an action that was contentious but not harmful to anyone, for instance eating a deceased pet. The subjects were not able to provide (adequate) reasons to support their attitude and the subjects did not change their moral attitudes when those reasons were debunked.

Another possible interpretation of dumbfounding experiments is that the participants have very good reasons for their view on incest; however these reasons are not available to subjects to report on, they 'work' unconsciously. A lot of problem-solving is done unconsciously and we do not always have insight in how we arrive at our appraisals in other domains. According to Prinz (2007), this interpretation is very unlikely since there is no evidence that much reasoning is going on, let alone unconscious moral reasoning. (30) It would be difficult to maintain that the participants could not access those reasons after intense moral reflection. Secondly, what could those reasons be? All the other reasons the participant gave for objecting to consensual incest, were refuted by the experimenters. The most likely explanation for moral dumbfounding is that the moral attitude was not directly the result of a process of moral deliberation.⁶

1.4 Moral Mary

In the last two sections, I made a case based on empirical evidence of why it seems that emotions should be given a central and productive role when it comes to moral appraisal. I argued that without emotions we cannot be fully capable moral agents and there are cases where we make a moral appraisal without the (direct) involvement of moral deliberation or conscious reasons. I also argue that there is also a phenomenological case to be made for the essential role that emotions play in ethical life. What would moral experience be like without the emotions?

Jesse Prinz (2007) provides us with a thought experiment about a woman named Mary. (38) In this case she is not colour blind but has never received any moral education; she has no innate moral attitudes.⁷ Thus Mary lacks the concept of right and wrong and does not possess the appropriate bodily feelings. However, she is an intelligent adult and is motivated to learn everything about morality. So she reads Kant, Mill, and other normative ethicists. In time, Mary becomes very adept at normative ethics and for every action she can determine whether that action would maximize utility or whether that action would lead to practical contradictions if pursued by all agents. The question is whether Mary is able to make a genuine moral appraisal on the basis of all she read, despite having an impoverished moral life? In Prinz' view Mary could wonder if an action is morally right or wrong, but she cannot know. She can know that an action maximizes utility, but can remain unsure whether an action is morally right. The same problem would hold were Mary to know knows that an

⁶ This does not mean that reasoning is not involved at all in a constructive way — although that is the conclusion that Haidt draws — merely that the moral appraisal was not the result of a deliberative process but of an emotional one. A more extensive discussion can be found in section 2.2

⁷ Moral Mary is an adaptation on the famous thought experiment of 'Mary's room' first proposed by Frank Jackson. In the original experiment Mary was a scientist who knows everything about colour, but has never had an experience of colour because she has lived her whole life in a black and white room. The question then is if Mary leaves the room does she learn something new?

action Y leads to a practical contradiction if everybody would perform that action. However, she can still wonder if this action is morally right. If the Millian and Kantian recommendations for action come into conflict, she has no way of choosing which one. Prinz' moral Mary is an interesting reformulation of Moore's famous naturalistic fallacy. Prinz argues that because moral Mary is unable to learn through Kantianism and utilitarianism what is good (she can always question it), it is therefore the moral emotions that inform us whether something is good.

I do not agree with Prinz that the moral Mary thought experiment makes a strong case for emotionism⁸. Prinz argues that reading Kant and Mill does not give Mary a sense of right and wrong. Since she lacks a sense of right and wrong she is unable to choose between Kantianism and utilitarianism. The thought experiment presupposes that Mary lacks the concept of what is right and wrong and therefore the possibility of moral deliberation is denied to her because she simply lacks the concept and not necessarily because she lacks the adequate moral emotions. Also the issue is whether Mary would be able to choose between utilitarianism and Kantianism is not relevant to debate on moral agency. The issue is whether Mary (even if she applied any of these moral theories in practice) would be able to make genuine moral appraisal. (Joyce, 2009: 3) Mary's inability to choose between either one of these theories does not necessarily mean that she is incapable of making a genuine moral appraisal.

Although the moral Mary thought experiment fails to make a strong case for Prinz' emotionism, I believe that Prinz does bring an interesting aspect of the role of emotions in moral experience to light. Intuitively it is clear that Mary is lacking something if she makes a moral appraisal about something without an appropriate emotional response.⁹ When we, for instance, become aware of child abuse, a normal reaction would be to become angry, perhaps even disgusted. We expect other people to become angry when they witness child abuse. A person who believes, like most people, that child abuse is a severe moral violation and does not become angry when witnessing child abuse is unintelligible to us. Whether or not Mary is a genuine moral agent, it is doubtful Mary would be seen as such by others. We expect someone to show emotional responses with respect to the things they care about. Not showing an adequate emotional response towards a moral issue can be a valid reason to question the sincerity of the 'non-emotional' person. One can argue that Mary makes moral appraisals of a different kind, one without the involvement of emotions. However, given the evidence from psychopaths and EVR-like patients, which suggests that emotions are indispensable to our moral functioning, and that Mary would most likely not be able to fully function in a typical ethical society; it is more plausible to deny Mary moral competence. Emotions make up an important part of our moral experiences, not only in our experience of ourselves as moral agents, but in my view just as important or even more so in the experience of others as moral agents.

In our everyday life, emotions at the very least seem not only to enrich the moral experience, but also seem to be indispensable for our moral practices. Emotions can help us improve our moral appraisals, like in the case of Huckleberry Finn. In the end Huck is swayed by his sympathies for his friend Jim. It is his sympathies for his friend that led Huck towards a better moral appraisal of the situation and better moral decision making. As Sauer

⁸ Emotionism refers to Prinz' sentimentalist view that emotions are central to our moral functioning.

⁹ I do not wish to imply and neither does Prinz wish to imply that in every case there is a strong emotional reaction. However often there is and most of the time we expect from other people and ourselves a strong emotional reaction. Most philosophers and psychologist would agree that emotions are involved in our everyday moral practises, although this role should not always be understood in a productive of constitutive way.

explains, it is not uncommon to be swayed by or to introspect our emotions like Huck does when we are encountering a moral problem. Take the following example:

“Imagine a friend telling you that she attended a mutual acquaintance’s wedding gown fitting. Your friend thought the dress was dreadful: unflattering, tasteless, and tacky. But she lied to the bride-to-be, telling her that she thought it was wonderful: flattering, tasteful, and classy. How do you figure out whether you thought her behaviour was ok?” (Sauer, 2013: 164)

We do not have a list of principles at hand which we can then easily consult. On the contrary, we take into consideration what we know and then act upon that which feels right. (idem: 164-165)

In §1.1 I argued that the case of Huckleberry Finn cannot be accounted for by the standard view since his feelings for Jim are not the result of a deliberation and even overwrite his ‘moral’ deliberation. The standard view can only portray Huck as acting on mere inclination and not on the basis of genuine moral motivation. But what about the following interpretation of this case? Huck does not act on mere inclination. He has come to see Jim as a fully-fledged human being and it is this awareness that makes Huck unable to turn Jim in. His action appears to be motivated by a practical inference of which he is not aware. Huck is doing the right thing against his better judgement.

However, according to the standard view Huck’s action would be considered irrational or arational, because we should act upon our judgements. A rational being deliberates before he acts and acts on *good* reasons. Nomy Arpaly (2003), who also draws on the example of Huckleberry Finn, notes that cases like Huck show that we do not have to equate general mental inferences with deliberation. (20-23) The case of Huckleberry Finn suggest that we can be rational without deliberation, calling into question the assumption that only deliberative processes and the outcome of deliberative processes are rational. Huck’s actions are not motivated by his deliberations; neither is the practical inference a consequence of the deliberation. Another observation Arpaly makes is that deliberation can be a source of irrationality:

“‘Rational deliberation’ is a common phrase in philosophical works. ‘Irrational deliberation’ is not. Yet we all suspect that deliberation is at times a source of irrationality.” (Arpaly, 2003: 27)

From our modern-day perspective, the judgement in the case of Huckleberry Finn can be considered false or even irrational. Huck considers Jim to be property, because that is the dominant view of society around him and this dominant view is based on racist and exploitative attitudes falsely considered to be moral. Jim is not property but a human being! Should Huck turn Jim in because he construes that action as moral, we can judge him as acting on bad reasons and thus acting irrational. In this case deliberation does not necessarily lead to a rational outcome. It seems to be not only moral but more rational to help Jim escape, despite the fact that this action is motivated by an emotion contrary to his judgements. By conceptualizing moral agents as rational agents who deliberate, we miss out on important aspects of not only moral inferences, but ethical life as a whole.

To summarise, cases like Huckleberry Finn cannot be adequately described from the perspective of the standard view of moral agency. By focusing on conscious rational deliberation, the standard view misses out on important emotional aspects of our moral considerations. In this chapter I explain in what sense we can understand emotions to play

an important part in ethical life and how to situate the empirical literature in this respect. Firstly, research into psychopathy suggests that we competent moral agency requires our emotions. Secondly, not all our moral appraisals can be justified by moral reasons as they are understood by the standard view of moral agency. Moral dumbfounding shows that the way in which the standard view conceptualises actual moral decision making is inadequate, because in those cases the moral appraisal is not caused by a process of deliberation and the subjects involved are unable to provide (adequate) reasons. Thirdly, from our moral experience and the moral Mary thought experiment we can learn that our emotions play an important part in making our moral behaviour and interactions intelligible. Emotions are important for communicating and understanding the moral responses of a person. Without an appropriate emotional response a person would most likely not be considered intelligible. On the contrary, it is a good reason to deny someone moral competence all together. Fourth, there appear to be cases where our emotions can help us improve our moral appraisals. Sometimes it can even be more rational to act on an emotion than a judgement.

The standard view of moral agency captures an important part of moral life. Sometimes we do deliberate and act on moral reasons. However, our (moral) emotions play an important role as well. If we want to understand humans as moral agents, we also have to understand humans as emotional agents.

2. Moral Psychology

In the previous chapter, I argued on the basis of empirical research that emotions play an indispensable role to our moral agency. Any adequate conception of moral agency needs to integrate emotions with rationality. Phenomena like moral dumbfounding, which suggest that certain moral judgements are the result of intuitive emotional processes and not rational deliberative processes, are somewhat difficult to explain from the standard view of moral agency. I am not the first to criticise the standard view, as it has received criticism from neuroscience and moral psychology. According to the standard view of moral agency, we are moral agents because we are capable of being aware of our deliberations and able to consciously act on those deliberations. In a sense, that would mean on the interpretation of the standard view we are no moral agents at all.

In this chapter, I discuss two authors from the field of moral psychology whose work offer important criticisms of the standard view, namely Joshua Greene and Jonathan Haidt. Greene argues that neuroscience shows that we have two separate systems for moral processing: a fast and automatic system, and a slow and conscious system (§2.1). Haidt argues that we are not the moral reasoners we think we are, because our moral judgements are caused by unconscious and intuitive processes. The moral reasons we exchange in our justification of our moral responses are therefore confabulations because we make up reasons after the intuitive moral response has taken place (§2.2). I have three reasons for discussing these two authors. First, they both have written extensively on the subject of moral judgement and have been influential in shaping the current debate by both relying on a *dual-process model* of moral judgement, which states that there are two different mental processes involved in the formation of a moral judgement. Secondly, both Haidt and Greene accept, just as I have, empirical evidence, which shows that we are not the strong deliberate reasoners the standard view supposes. Thirdly, both offer an integrated account of rationality and emotion. However, I argue that their accounts are not adequate.

I show that these authors still share certain assumptions with the standard view of moral agency. These assumptions are: seeing emotions and cognition as two separated processes that either compete or cooperate with each other, a strong contrasting of controlled and automatic processes, and thereby equating the former with consciousness and the latter with subconscious processes. I argue that the neurophysiological and psychological evidence Haidt and Greene appeal to do not support these assumptions. I conclude this chapter by arguing that in order to formulate a notion of moral agency that can accommodate cases like Huckleberry Finn, we need a notion of moral agency that does not adhere to these assumptions.

Although this chapter contains a comprehensive discussion of the neurophysiological basis of cognition and emotion, I am hesitant to provide exact definitions for both, because of the different definitions used in the scientific and philosophical literature discussed here, and addressing the differences between those definitions is not necessary for my thesis. When I use the term cognition in this chapter I mean those processes that are normally associated with thinking, e.g., decision making, reflection, learning, etc. When I use the term emotion in this chapter, I refer to short, bounded episodes of intense feeling, combined with physiological changes. What is important for now is that in the (meta-)ethical debate moral evaluations are traditionally attributed to either rational processes (and therefore also cognitive) or emotional processes. As a result cognition and emotion have been contrasted in a semantic context, for instance 'moral judgements are the expression of an emotion' and/or

a causal context, for instance 'moral judgements caused by a process of rational reflection.' In this context, emotions are seen as arational, i.e., not apt for rational evaluation.

2.1 Greene's Dual-Process Theory

Greene's aim is to come to an understanding of the mind in physical terms. He is especially interested in the neural correlates of both reason and emotion in moral judgement. In order to come to a better understanding of the mind, he employs in his research neuroimaging techniques and behavioural studies. His most famous experiment involves studying the responses of persons to moral dilemmas, while they undergo a fMRI¹⁰ scan. (Greene et al., 2001) Most notable amongst the moral dilemmas was the *trolley- and footbridge-dilemma*¹¹. The trolley dilemma goes as follows:

'You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman. If you do nothing the trolley will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman. Is it appropriate for you to hit the switch in order to avoid the deaths of the five workmen?'

The footbridge dilemma is a variation on the trolley dilemma:

'A runaway trolley is heading down the tracks toward five workmen who will be killed if the trolley proceeds on its present course. You are on a footbridge over the tracks, in between the approaching trolley and the five workmen. Next to you on this footbridge is a stranger who happens to be very large. The only way to save the lives of the five workmen is to push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if you do this, but the five workmen will be saved. Is it appropriate for you to push the stranger on to the tracks in order to save the five workmen?'¹²

When confronted with the trolley dilemma, most people tend to agree to pull the lever, so that the five workmen are saved (a utilitarian consideration). However, when confronted with the footbridge dilemma, most people say 'no' against pushing the large man of the footbridge, therefore not saving the five workmen. The trolley dilemma and the footbridge dilemma are construed by Greene as a dilemma between a deontological choice versus a utilitarian choice. The choices of 'not pulling the lever' and 'do not push the man' are construed as deontological, because these choices are concerned with whether an act is intrinsically good. The choices of 'pulling the lever' and 'push the man' are construed as utilitarian choices, because only the consequences are taken into account.

¹⁰ Functional Magnetic Resonance Imaging is a technique to measure brain activity by detecting changes associated with blood flow.

¹¹ The trolley-dilemma was first introduced by Phillipa Foot (1967) and later extensively analysed by Judith Thompson (1985)

¹² Both the trolley-dilemma and the footbridge-dilemma are formulated as used in the experiment by Greene et al. (2001). Retrieved from: <http://science.sciencemag.org/content/suppl/2001/09/13/293.5537.2105.DC1>

According to Greene, the discrepancy between our intuitions about these two dilemmas can be explained by a distinction between *personal* and *impersonal* moral cases. Personal moral actions are those that could lead to serious bodily harm to a particular member or members of particular group of people. (idem: 2107) Personal cases like the footbridge dilemma tend to engage people's emotions in contrast to impersonal cases. This is because, so argues Greene, personal cases are more emotionally salient.¹³ (idem: 2106) The thought of pushing someone to his death is more emotionally salient than switching a lever that will produce the same result. Deontological evaluations, according to Greene, are the result of fast emotional processes and therefore automatic. Utilitarian evaluations, on the contrary, are the result of slow controlled cognitive processes. Thus deontological evaluations do not involve rational thought and moral evaluation involves two separate processes, which compete against each other. If our deontological judgements are the expression of emotions, then they cannot be rational, because our emotions cannot be rational. In this section, I argue that Greene is committed to a strong separation of emotions and cognition in his dual-process approach, and that this is a theoretical commitment we need not necessarily have to share. I start out by explaining dual-process theory after which I discuss the underlying neurophysiological evidence for dual-process theory.

2.1.1 Dual-process and dual-systems theory

Dual-process approaches have been highly influential within cognitive psychology. There is no single unifying dual-process theory for the human mind, but dual-process theories have been commonly employed in areas of psychology like learning, reasoning, social cognition and decision making. (Frankish, 2010: 916) Typically a dual-process-theorist holds that there are two distinctive processing modes, usually designated as system 1 and system 2 processes.¹⁴ System 1 is typically defined as unconscious, rapid, automatic and high capacity. System 2 is usually seen as slow, deliberative and conscious. (Evans, 2008: 256) Cognitive biases are usually (but not always exclusively) attributed to system 1 processes. System 2 processes are seen in contrast as logical, abstract, analytic and requiring limited working memory. System 1 processes are involved in quick judgements and are seen as heuristic and system 2 processes are seen as deliberative and reflective. Thus system 1 processes are seen as low effort, high capacity processes and system 2 as high effort, low capacity processes. The former are considered as associative and the latter as rule-based. System 1 processes are usually considered evolutionary older (shared with animals or closest to animals) than system 2 processes. Attitudes generated by system 1 are implicit and hard to change, while attitudes generated by system 2 are explicit and easier to change.

Greene argues that this model is also applicable to moral evaluation. Utilitarian judgements are the result of system 2 and deontological judgements are the result of system 1. Although traditional applications of dual-process theory usually were not concerned with emotions – dual-process theories are usually seen as describing two different forms of cognitive processing – it is not difficult to place the emotions in system 1 processing. Greene not only posits two different types of processing systems in moral evaluations but also thinks the underlying cognitive structure is different for each process. Therefore, Greene's Dual-

¹³ In response to criticism, Greene no longer speaks of the personal/impersonal distinction, but of personal force. (Greene et al., 2009) However I will not address the difference between personal force and personal/impersonal here since I am primarily interested in the underlying distinction between two different mental processes.

¹⁴ Although many different names have been proposed in order to contrast two kinds of thinking.

Process theory of morality (DP-theory) can not only be described as a dual-process, but also as a dual-systems theory.¹⁵ A dual-systems theory posits not only two processes, but also two distinct underlying cognitive systems – in Greene’s case an emotional system and a cognitive system. For Greene these processes should have their own distinct neurophysiological basis. In Greene’s DP-theory this means there is a distinction between emotional and cognitive regions in the brain. Given that the underlying cognitive structures are different and the competitive nature of the model, one can speak of two minds in one brain. According to Greene, when we make a deontological judgement, brain regions associated with emotions should display more activity on fMRI scans. And when we make a utilitarian evaluation regions associated with cognition (or more specific working memory) should display more activity on fMRI scans. (Greene, 2001: 2106)

Greene’s believes that in theory, it would be possible to make a map of the brain detailing which areas of the brain perform emotional tasks and which areas perform cognitive tasks. If in the performance of another task a brain region associated with emotion becomes more active, then that task can be said to be emotional. Greene argues as follows: there are certain brain regions that show an increase in activity in a fMRI scan when experiencing emotions or engaging in emotionally charged tasks. For instance, when we experience fear there is an increase in activity in the part of the brain called the amygdala. The amygdala can be associated with the emotion of fear, meaning that this specific part of the brain is (partly) responsible for the emotion of fear. According to Greene’s (2015) research, deontological evaluations of moral dilemmas show increased activity in those brain regions associated with emotional activity. (1016-1017) Since both emotions and deontological evaluations are associated with similar brain regions, Greene concludes, deontological evaluations must be emotional evaluations. When subjects in Greene’s experiments underwent a fMRI-scan while being confronted with (moral) dilemmas, Greene (2015) found in cases of personal dilemmas increased activity in the amygdala, medial prefrontal cortex, medial parietal cortex and the temporal parietal junction, the parts usually associated with emotions. Increased activity in the dorsolateral prefrontal cortex was found for utilitarian evaluations, the part that is usually associated with cognition.

2.1.2. The neuropsychological basis of emotions and cognition

In order for Greene’s inference to be viable, the brain regions in question have to function selectively, i.e. only showing increased activity for certain similar tasks. Underlying Greene’s model is the tacit assumption that brain regions involving cognitive and emotional tasks can be clearly separated from each other. A region in the brain, for example the amygdala, should only be involved with emotions and not cognition or vice versa. You cannot infer that a certain task is either emotional or cognitive in nature, if the brain does not show a clear distinction between emotion and cognition on a neurophysiological level. A recent meta-analysis study by Luiz Pessoa (2008) has shown that defining the emotional brain is difficult, because the designated areas are complex and involved in numerous functions. The amygdala has traditionally been seen as a core affective region strongly linked to fear processing. Greene found in fMRI experiments a consistent activation of the amygdala when subjects made a deontological evaluation. However, it appears that the amygdala is not only involved in emotional tasks, but also in tasks traditionally seen as cognitive. The amygdala also supports functions that are closely linked to attention and associative learning. (Pessoa,

¹⁵ For more on dual-process and dual-systems theory see Frankish (2010)

2008: 149) The amygdala's role lies more likely in detecting motivationally salient stimuli. Lindquist, Wager, Kober, Bliss-Moreau and Barrett (2012) found that the amygdala is also strongly associated with perception. (130-132) Amygdala responses are strongly dependent on attention, and linked to perception such that the responses are not simply predicted by the physical characteristics of the stimulus. This means that controlling the attention to and cognitively changing the meaning of emotionally evocative stimuli greatly affects amygdala responses. The amygdala enables monitoring, updating and integrating sensory signals. (Pessoa, 2008: 149-150)

The dorsolateral prefrontal cortex (DLPFC), which Greene associates with utilitarian considerations, is part of a network that is involved in top-down, goal-directed control of attention. The DLPFC is known to be involved in working memory. However the DLPFC also shows increased activity during emotion experiences and perception. (Lindquist et al., 2012: 137). The DLPFC is active during mental states in which subjects attended to emotional feeling or perceptions. In fact, some areas in the DLPFC are more likely to show increased activity during instances of anger. (idem: 138)

It appears that activations in brain areas do not neatly cohere to the distinction between emotion, cognition and perception. Normally perception (or the brain regions associated with perception) is seen as separate from affect. Although the perception of an object, for instance a hungry bear running towards you, can lead to an affective response, for instance cowering in fear, they are distinctive steps. Only after perceiving something a cognitive or affective process follows that establishes the salience, relevance or value of what is perceived. This would mean that in the perception of the bear only the regions for vision are active and the fear response would lead to an increase in activity in brain regions associated with affect. Barret & Bar (2015) proposed on the basis of their data that affective states seem to influence the processing of sensory stimulation. (1331) It is not the case that an object is first identified and then in a subsequent step the salience, relevance or value is established. Affective responses support vision from the very moment visual stimulation occurs. If perception and affect are fundamentally separated, brain regions associated with affect should not be involved in the perception of an object.

Furthermore, it also appears that emotion and cognition cannot be separated from each other on a neurophysiological level. The prefrontal cortex (PFC) is strongly associated with cognition in that the PFC is critical for cognitive control and the manipulation and maintenance of information. However, many PFC structures are also involved in emotions. (Pessoa, 2008: 150) Emotional and cognitive contributions to executive control cannot be separated. (idem: 153) This means that in situations where we have to plan or to make decisions, we cannot identify in the brain which part of the decision or plan is cognitive or emotional. Emotions and affectivity are integrated in the brain. Many of the regions that are part of the 'emotional' brain are also involved in task that are normally considered to be only cognitive like autobiographical memory, decision making, language, self-referential processing etc. (Lindquist et al., 2012: 139) This also implies that these cognitive functions play a routine role in constructing experiences and perceptions of emotions. When we experience a certain emotion, there is also an increase in activity in areas associated with cognition. Barrett (2013) argues in light of these findings:

"There is no 'affective' brain, 'social' brain, or 'cognitive' brain. Each human has one brain whose functional properties can be understood differently for different time scales and levels of organization." (387)

As a result of these considerations, we should abandon the picture that emotion influences cognition during decision making. Cognition and emotion (or their underlying processes) cooperate to realize a behavioural outcome. A picture where emotion and cognition battle in the brain when a moral decision has to be made is not validated by studies of the brain. (Lindquist et al., 2012: 142) Pessoa (2008) concludes that behaviours might be characterised in a distinction between cognition and emotion, while true integration of emotion and cognition takes place, blurring the line between the two. (154) Psychopaths and patients with VMPFC-lesions not only showed deficiencies in emotions but also in (practical) reasoning, suggesting a complicated and close relation between emotions and rationality.

Another consequence of the above given picture of the brain is that the distinctions between controlled and automatic processes becomes more problematic. Emotions are typically seen as the outcome of automatic processes and cognition is seen as a controlled process. If a behaviour has an emotional basis, then the underlying processes are not under our control. This indicates a clear distinction between uncontrolled emotions and cognitive processes which we can control. However, if there is no clear distinction between cognition and emotions, and behaviour is orchestrated by many different brain areas, then there is no clear difference anymore between controlled and automatic processes. Barrett et al. (2004) suggest that even conscious experience might not be an indication of controlled processing and people can engage in controlled processing when they do not experience themselves doing so. (563)

Because Greene assumes a strict distinction between emotional and cognitive brain regions and he associates deontological judgements with the former and utilitarian judgements with the latter, the DP-model conflates three distinctions. (Sauer, 2013: 59) (i) Automatic and controlled processes are conflated with respectively (ii) emotional and cognitive processes and, as Hanno Sauer (ibidem) argues, respectively (iii) unjustified and justified processes of judgement formation.¹⁶ Moral competency lies in the use of system 2 processes over system 1 processes. (idem: 35-36) Meaning that better moral judges rely more on belief-based system 2 reasoning than the heuristic emotions of system 1. In Greene's view, a competent moral agent engages more in utilitarian reasoning, than relying on deontological emotional judgements. However, this point does not cohere with research into belief-logic conflict problems¹⁷ where dual-process theories of rationality are applied. People with higher ability reasoning skills do not engage more in system 2 reasoning than people with lower reasoning skills, but are simply more effective when they employ system 2 reasoning. (Evans, 2008: 264-265) Therefore, it is not the case that competent moral agents necessarily engage in more moral reasoning than incompetent moral agents. More importantly, the conflation of emotional with automatic and cognitive with controlled is not something that we should adhere to.

The neuroscientific research discussed in this section shows that Green should not conflate automatic processes with emotions, as this is not supported by the neuroscientific research discussed in this section. Moreover, there are many examples of automatic processes that are generally considered to be non-emotional. For instance, intuition in logic,

¹⁶ Besides arguing for deontological evaluations as emotional and utilitarian evaluations as rational, Green takes the argument further. He claims that one's deontological arguments should be discounted in favour of our more rational utilitarian evaluations, because our deontological judgements are not truth-tracking and respond to morally non-relevant factors. (Green, 2015)

¹⁷ A belief-logic conflict is the phenomenon where people are inclined to accept the conclusions of an argument based upon how well that conclusion fits with their beliefs, instead of judging the validity of a conclusion on the basis of its premises.

language and physics. We also often employ judgemental heuristics, where there is no easy distinction to be made between controlled and uncontrolled processes.¹⁸

2.2 Social Intuitionist Model

Greene scepticism is aimed towards the standard notion of moral agency in that not all our moral responses are the outcome of cognitive processes. Deontological judgements come from the emotions, but that does not mean we cannot engage in moral reasoning. Only that deontological evaluations are not caused by a cognitive and rational process. Thus Greene leaves open the possibility for us to be rational moral agents in the traditional sense. Jonathan Haidt takes the scepticism towards the standard notion of moral agency a step further. Where Greene draws into doubt the rationality of deontological moral evaluations, Haidt draws into doubt the rationality of all our moral evaluations.

Haidt's scepticism is the result of research into the phenomenon called moral dumbfounding.¹⁹ (Murphy et al., 2000) In these experiments people were unable to provide reasons for their moral attitudes and most of them did not change their moral attitudes when confronted with this fact. For Haidt cases of moral dumbfounding are not the exception, but rather the norm. From this line of reasoning Haidt developed his Social Intuitionist model of moral judgement (SIM). In this section, I will show that the SIM draws a strong contrast between subconscious processes and conscious processes; the former are equated with emotions and seen as automatic, the latter with slow rational thinking, which are seen as a controlled process. I start out by explaining how the SIM aims to integrate emotions with rationality. I argue, contrary to what Haidt supposes, that we have a remarkable amount of control over our automatic processes. Hence, moral reasoning and emotions are more intricately involved with each other than Haidt supposes.

2.2.1 SIM as a dual-process theory

We can ask people for reasons to support their moral appraisals with reasons, which in most cases they are able to provide. However, according to Haidt these reasons are often superficial and post hoc. When you press those people on their moral values, these values turn out not based on deep moral reasons which are the outcome of a moral deliberation. Rather, these values come from strongly embedded sentiments. Our moral attitudes are not the result of moral deliberation, but our moral attitudes are caused by a quick flash of emotionally fuelled moral intuitions. (Haidt, 2001)

The quick flash of intuition that leads to a moral judgement is the first link in the SIM and is called 'the intuitive judgement link'. In total there are six links in the model: (1) the *intuitive judgement link*, (2) the *post hoc reasoning link*, (3) the *reason persuasion link*, (4) the *social persuasion link*, (5) the *reasoned judgement link* and the *private reflection link*. The *intuitive judgement link* describes the sudden appearance of a moral judgment, through a subconscious process. Evaluations happen rapidly before any consciousness processing has taken place; only the results appear in consciousness. Intuitions are triggered implicitly and automatically. This means a person does not know what exactly causes his or her moral evaluation; we have no insight into what drives our behaviour, because our typical moral

¹⁸ These examples are from Sauer (2013)

¹⁹ See section 1.3 for a more extensive discussion of research into the phenomenon of moral dumbfounding

evaluations are the result of subconscious processes. Intuitions are useful, because they require little effort and cognitive power.

If Haidt is correct, reasoning plays no role in the formation of our typical moral evaluations. Reasoning only comes into play after an initial evaluation has been made. The second link in Haidt's model is called the *post hoc reasoning link*. The purpose for moral reasoning is not to generate moral judgments, but to provide *justification*. Thus reasoning is an effortful process in which a person searches for arguments that will support an already-made judgment. Hence, reasoning is post hoc.

Haidt argues that there is an important social aspect to moral evaluation. The first two links describe how moral judgments are formed through our intuitions and the justifying role our reasons play. However, these two links do not describe how an agent changes his moral judgments. Revision of our moral judgment primarily occurs through social interaction and is typically not an individual affair. The third link, *the reason persuasion link*, describes how our moral reasoning when verbally uttered can affect other people's moral intuitions. Moral discussions and arguments do not convince through logically compelling arguments, but trigger new affectively valenced intuitions in the listener. (Haidt, 2001: 819) The fourth link is *the social persuasion link*. This link describes how people conform their intuitions to group norms. The moral evaluations we have a directly shaped by the judgments of others; especially judgments made by acquaintances, friends and family. This works even if no reasoned persuasion is used. These four links make up the core of the SIM.

Haidt does not deny that moral reasoning exists. Reasoning mainly serves to influence the intuition of others and even then moral reasoning is not necessary. Moral reasoning has an essential social structure. Furthermore, in the model reasoning is depicted as causally ineffective in producing moral evaluations. Even in the reason persuasion link our moral reasons do not directly cause a moral judgement, but activate other intuitions. Thus there appears to be no genuine moral reasoning in the SIM, because in order to speak of genuine moral reasoning, it is generally believed that the moral evaluation has to be caused by the moral reasoning. (Sauer, 2013: 77) Merely being able to justify a moral evaluation is not enough.

The final two links of Haidt's model together with the post hoc reasoning link and the reason persuasion link involve moral reasoning. The fifth link is the *reasoned judgement link* and offers room for *private* moral reasoning. Only in cases where the initial evaluation is weak and reasoning capacity is high, a judgment may be overwritten by the sheer force of logic. (Haidt, 2001: 819) Haidt emphasises that these cases are rare. And even in these cases the moral reasoning is derivative of the initial moral evaluation; the reasoning merely involves taking new information into consideration. The reasoned persuasion link can only be activated after an initial evaluation has already been formed. The sixth and final link is *the private reflection link*. When thinking over moral cases, competing intuitions can arise within an agent. Reason may trigger an intuition that contradicts the initial intuition. What follows can be characterised as an internal dialogue. Either the strongest intuition wins out in forming the judgment or reason chooses among the alternatives on the basis of the application of a conscious rule or principle.

Just like Greene's DP-theory, SIM is an application of dual-process theory. However, in Haidt's SIM there is no simple contrast between emotions and cognition. Haidt's view is more in line with 'traditional' applications of dual-process theories, in that he describes two different cognitive systems. The first is a quick, intuitive and emotionally fuelled cognitive system, which is responsible for our moral evaluations. The second is a slower, conscious deliberative system that is responsible for the justification of the moral evaluations. Contrary

to Greene's DP-theory there are no two competing systems, but system 1 and system 2 cooperate in realising moral behaviour. It is important to note that intuition and reasoning are two strictly separate cognitive systems; only the intuitions involve affective states and only rationality is involved with system 2. Thus the SIM is an example of a cooperative dual-process theory.

2.2.2 Control and unconscious processes

When we reason, according to Haidt (2001), we are like a lawyer defending a client, instead of a judge or scientist seeking the truth. (820) If the nature of our reasoning is post hoc and not effective in the formation of our moral evaluations, then moral reasons are confabulatory. What looks like causally effective reasoning is nothing more than a social charade to influence people on a subconscious level. (Sauer, 2013: 81) It is important to note that the question thus far in the discussion with Haidt and Greene on the rationality of our moral evaluations, is just as much a question of automaticity versus control. Because our moral intuitions are automatic they are not rational and neither are the moral evaluations which result from those intuitions. Reasoning and therefore rationality, is understood by them as occurring consciously. (Haidt, 2001: 818) If we believe that moral agency means our moral evaluations have to be caused by moral reasoning, then this implies that we are not moral agents.

I do not argue against the empirical findings on which SIM is based. I believe Haidt is right in stating that our moral evaluations generally occur in a quick intuition-like process. Most of our moral evaluations occur automatically. However, moral reasoning seems to play a larger role than Haidt supposes. Kenneth & Fine (2008) argue that reasoning or controlled processes have a larger influence on automatic processes than what can be characterised as confabulatory. Even if Haidt is right and individual judgements are directly caused by intuitive processes, moral reasoning can still influence those underlying intuitions. Evidence from social cognitive psychological research suggests that we have some measure of control over our automatic processes through strategic overriding. This can take the form '*up-front*' mental control or '*after-the-fact*' correction. (Kenneth & Fine, 2008: 88; Govorun & Payne, 2006: 130) The last-mentioned sometimes occur in cases of what is called 'mental contamination'. Cases where irrelevant factors might influence an evaluation, people are able and willing to discount the irrelevant factors that may distort the evaluation. (Schwartz & Clore, 1983) When a person's mood leads to a certain bias, and the attention of that individual is drawn to that mood as a possible source of that bias, the bias can be corrected. There is other research suggesting that when people are made aware that they have a tendency to make certain types of evaluation in a certain biased way, they will effortfully override these intuitions, provided they are motivated to be unprejudiced. (Kenneth & Fine, 2008: 89) If we apply these insights to Haidt's SIM, it is possible to block the tight connection between intuitions and (moral) evaluations. This blocking is an effortful, conscious process and is also not comparable with the reasoned judgement link, because there are many cases where an initial intuition can be very strong, for instance in cases of racial prejudice. Thus people can engage in '*after-the-fact*' adjustments of initial social and moral evaluations.

On top of that, there is also evidence that people have some '*up-front*' control over the activation of certain intuitions. A person can make a conscious decision to what kind of situation that person exposes himself. We can consciously control the input by choosing our surroundings. For instance, if you are on a diet you can choose to skip the ice-cream aisle of the supermarket out of fear you might succumb to the temptation. Just as you can avoid

smokers if you are trying to quit smoking. (Pizarro & Bloom, 2003) The same holds for the moral domain, where we can avoid situations that would morally compromise us or seek out situations that would evoke our sympathies. Thus reasoning can influence our moral evaluations by controlling the situations that would elicit certain intuitions.

'Up-front' control is not only directed at external conditions but also at internal conditions. Automatic processes can be pre-consciously controlled in accord with consciously held goals. Keith Payne (2005) has done research into the self-regulatory processes that enable people to limit the influence of automatically activated information. One of the tests Payne employed was the *weapon identification task*. In the test, participants were shown a pair of pictures flashed briefly on a monitor. The participants were instructed not to respond to the first picture, which is the picture of a face that is either white or black. The second picture was that of either a gun or a tool. The participants had to classify each target object as either a gun or a tool by pressing an arrow key. (idem: 492) For most participants seeing the face of a black man led to a false positive of identifying a picture of a gun when the picture of the face of a black man was followed by a picture of a tool. However Payne found that people highly motivated to avoid acting with prejudice and with higher general executive ability, were found to make fewer mistakes in the stereotyping task. (idem: 491) Since they were motivated to avoid discrimination, those people were able to constrain their processing to the relevant information, rather than being driven by irrelevant but activated information. Kenneth & Fine argue that evidence from experiments like the weapon identification task does not allow us to draw the conclusions that Haidt draws. According to them, these results should be interpreted as showing that "although the deployment of 'up-front' control is neither consciously willed nor accessible, it depends on the availability of controlled processing resources." (Kenneth & Fine, 2009: 92) Preconscious control can be the causal consequence of a prior or current reflective endorsement of a certain goal. Closely related to the above mentioned examples of 'up-front' control is the application of called 'if-then' plans. (Gollwitzer et al., 2009: 616) Implementation of these 'if-then' plans or 'implementation intentions' can have a desired effect on people's automatic reactions. An 'if-then' plan is a plan that links a critical situation with a goal-directed response. (Gollwitzer, 1999: 493) For instance, if I encounter something fearsome, then I will remain calm and relaxed. One study showed that a group of participants were able to reduce reactions of disgust and fear towards fearsome stimuli by forming the 'implementation intention' of staying calm when exposed to the fear inducing stimuli. (Sauer, 2013: 112)

From the above presented findings, we can learn that rationalisations are slowly transposed on our intuitions. This indicates that we do have some control over our automatic reactions. We are able to modify our moral intuitions in accordance with reasoned goals we might possess. With this insight we can somewhat rehabilitate those who participated in the moral dumbfounding experiments. The case of incest between brother and sister is quite an extreme case. Although the case of incest between brother and sister used as a vignette by Murphy et al. (2000) stipulates that no harm was done, in everyday life incest is very much associated with harmfulness. The fact that the participants are unable to immediately produce adequate reasons for their moral evaluation does not mean that their response is completely irrational. The automatic evaluation of incest as morally wrong can be based on reasons that participants have learned. Those reasons become part of our automatic responses in the sense that those reasons form those responses in question. The forming of automatic responses or habits can be a very thoughtful and therefore considered to be part of a rational conscious process. It is therefore understandable that it is difficult for the participants to abandon those reasons in cases like the incest between brother and sister,

which is by design not a scenario that is likely under real-world conditions. People are unable to give adequate reasons for them judging Julie and Mark as immoral, because people cannot fathom incest not being harmful. Our intuitions are developed under real world conditions, conditions that do not apply to the thought-experiment. The fact that most of our moral evaluations are the result of intuitive-like processes where reason is post-hoc does not mean that there are no reasons involved. This of course does not mean that all instances of post hoc reasoning influence our intuitions. Many instances of post hoc reasoning are probably confabulatory. However, our reasons are not always, or not necessarily confabulatory.

2.3 Awareness and Conscious Control

Seeing moral appraisals or moral judgements as the exclusive result of a cognitive process or as the exclusive result of an emotional process is problematic given the bodies of research from the neurosciences and moral psychology discussed in this chapter. Equally problematic is to ascribe control exclusively to our cognitive processes. These theoretical assumptions that Greene, Haidt and the standard view of morality adhere to is not something we need to share.

Let me summarize. First, I argued that there is sufficient neuroscientific evidence to question this strong separation between emotions and cognition. The neuroscientific research discussed in this chapter showed that you cannot separate areas in the brain associated with emotions from those associated with cognition. This means that the distinction between emotions and cognition cannot be justified on the basis of brain research. This implies that the separation of emotions and cognition is not something we have to commit to.

Secondly, I also argued that consciousness is not necessary for control. For example, Haidt posits a strong distinction between controlled and uncontrolled processes. According to Haidt (and I believe he is right on this point), our typical moral evaluations are quick and automatic. In his view, most of our moral judgements are caused by our moral intuitions and not our moral reasoning. The processes behind our moral evaluations are subconscious and happen without our control. However, I agree with Kenneth & Fine, who argued that even though our moral evaluations are more like gut-feelings, this does not mean that we do not have some sort of rational control over them. There is a continuous complex interaction between moral reasoning and quick intuitions. This means that moral agency cannot be understood in terms of a continuous stream of *isolated* moral evaluations. An agent's moral attitudes are modified and re-evaluated over time, so that an agent's moral attitude cannot be attributed to either one of the systems. Haidt conceptualizes ethics as a flash of intuitions. However, ethics is not something we do sometimes, but is implicit in everything we do. Barrett et al. (2004) argued:

“[C]onscious experience is not diagnostic for distinguishing the presence of automatic and controlled processes. [...] We argue that [...] people can engage in controlled processing even when they do not experience themselves doing so.” (563)

An evaluation of current neuroscientific evidence shows that it is not easy to draw a distinction between emotions and cognition on the basis of brain processes. Also, it is not easy to draw distinction between processes which we control and processes which are

automatic. Thus the assumption that we are only in control if we consciously deliberate is not something we necessarily have to accept. We do not have to be aware of what we are doing in order to be in control. These assumptions are of course interrelated. First, in order to speak of control a process has to be conscious and deliberative. This stands in contrast to automatic processes, which we have no control over because they are subconscious. Second, emotions are automatic processes and stand in contrast to cognition. Third, moral evaluations are caused by either controlled or automatic processes. If moral appraisals are no longer only the result of a cognitive or an emotional process, we need a new definition of appraisal. This will be the subject of the next chapter.

3. Embodied Cognition, Emotions and the Social Mind

In the previous chapter, I argued that moral appraisal (or judgements) are not adequately conceptualised as only emotional in nature or rational in nature given the bodies of research from the cognitive sciences. Therefore, I argued that we need a new outlook on moral agency that incorporates the emotions as well as our rational capacities. I believe the *enactivist approach* offers interesting and fruitful conceptual tools to explore a new conception of moral agency. The enactivist approach combines certain findings of the cognitive sciences with the phenomenology as developed in the works of Merleau-Ponty.

In this chapter I, will address the concept of appraisal through an enactivist framework and more specifically Giovanni Colombetti's work. As I explain, according to the enactivists cognition is understood as *embodied* (§3.1). That is, cognition must be understood as a dynamical interaction between an organism and its environment. According to Colombetti this means that cognition is fundamentally affective. Moreover, emotions itself are also seen as a cognitive activity (§3.2). Emotions cannot be separated from appraisals.

If cognition is understood as a dynamical interaction between an organism and its environment, that means that our cognitive processes are extended into (or involve) the environment of the organism. Cognition is therefore neither skull-bound nor body-bound. I discuss how this has consequences for how cognitive capacities are shaped in relation to the environment (§3.3). For humans this environment is a social environment and therefore our cognitive processes are shaped through social interaction. Our (moral) cognition is fundamentally social.

I conclude this chapter with some remarks on the possible intersubjective and embodied nature of moral appraisal, which will serve as a jumping off point for the next chapter, where I explore an embodied, situated account of ethical agency.

3.1 Enactivism

I start this section with a general discussion of *enactivism*, before moving on to the more specific *autopoietic enactivism*. Enactivism was first put forward by Varela, Thompson & Rosch (1991) in the book *The Embodied Mind* as an alternative to *classical cognitivism*. Cognitivism is a position within the cognitive sciences that holds that intelligence is computation and that cognition can actually be conceptualised as computations or manipulations of symbolic representations. (idem: 40) A mental symbol 'stands' for something (in the world) and is therefore a representation. In the cognitivist view, the characteristics of the body are regarded as unimportant as long as the system performs the functions for cognition. What is missing in this picture is a constitutive role for human experience or consciousness. Just like a computer does not need consciousness to compute — only the right programme — so it seems that the human mind (or brain) also does not need consciousness. Enactivism is not only a philosophical programme, but is an interdisciplinary approach taking ideas from notably the cognitive sciences, biology and the philosophical tradition of phenomenology, in particular the works of Maurice Merleau-Ponty. Merleau-Ponty (1945) made a distinction between the *objective body*, i.e. the body experienced as a physiological structure (Körper), and the *phenomenal body* (Leib), i.e. the body as a subject. (121) The latter he also calls the *lived body* and is the body experienced

from a first person point of view. The lived body is also the body experienced *pre-reflectively*, that is in the background of my other experiences. For instance, “when I pick up a cup of hot tea, I feel the hot smooth surface of the porcelain and the heat penetrating my fingers, and the sensations linger for a time after I have put the cup back down on the table.” (Thompson, 2007: 250) In this example I experience not only the cup of tea, but also my fingers and hand through which the experience of the hot cup of tea is made possible. I am aware of my body as that through which various experiences are possible. According to the enactivists and Merleau-Ponty we see our bodies both as physical structures and as lived experiential structures. (Varela et al., 1991: xv) These two sides of *embodiment* are not mutually exclusive. Instead we continuously circulate back and forth between these two perspectives.

In the enactivist approach, the mind is *enacted* or brought forth by the whole living organism; the mind and cognition are embodied, i.e., dependent upon and constituted by features of the physical body. Any study of the mind cannot ignore *lived experience* and has to develop adequate methods for its investigation. This does not mean that the enactive approach is specifically interested in lived experiences, but that a phenomenological analysis should always be a part of (scientific) research into the mind. For the enactivists, cognition is not abstract symbol manipulation but the activity of *sense-making*, where the world is transformed by an organism into an *environment (Umwelt)* and thereby the organism enacts their own domain of meaning and value. The organism’s environment is the sense it makes of the world and the significance and valence are the result of the global actions of the organism. (Thompson, 2007: 158)

According to Steve Torrance (2006), two major collections of ideas in enactivism can be distinguished: a broad view of what it is to be an agent with a mind and a more focussed account of the nature of perception and perceptual experience. The first strand is often called *autopoietic enactivism* and sees autopoiesis, which is a form of self-organisation, as central to explaining the fundamental properties of our mental life. The second strand is often called *sensorimotor enactivism* and is not so much focussed on autopoiesis, but explains perception in terms of sensorimotor contingencies, that is the way, in which stimulation changes contingently upon actions of the organism. O’Regan & Noë, (2001) who are proponents of this idea, argue that perception cannot be reduced to the representation of sensory information in certain brain areas. Instead, they claim that vision is “a mode of exploration that is mediated by knowledge of what [they] call sensorimotor contingencies.” (940) Perceiving, in this view, is characterised as the exercise of practical knowledge, i.e. knowledge of sensorimotor contingencies. Perception is therefore just as much a motor process as a sensory one. Although there are many more strands in the enactivist approach, I restrict the discussion to autopoietic enactivism. I also focus on those ideas that relate most closely to the discussion of the topic my thesis, which is an exploration of an embodied situated account of moral agency.

According to the enactivist approach cognition emerges from the interactions between an organism and its environment. These interactions are called *sense-making activities*. Autopoietic enactivism aims to connect cognition to (simple) biological functions. For this the autopoietic enactivist use the term autopoiesis. Autopoiesis is a specific form of self-organisation., that is, the process through which an organism makes itself distinct from its immediate surroundings not only for itself but also for an observer. To clarify, we can make all kinds of distinctions on the basis of convention. For instance, in meteorology we can differentiate between Cirrus clouds and Cumulus clouds. Thus we individuate (clouds) from the outside by convention. We can also refer to the body in different ways; we can refer to the given anatomical structure, or physiological body, or a lived body in a given world.

Autopoietic enactivism states that all organisms are *self-organising*; autopoiesis should therefore be understood as explaining the nature of all living systems. When an organism self-organises it not only actively maintains itself, but also maintains a distinction between itself and the environment. (Di Paolo & Thompson, 2014: 69) The ability to self-organise is called *autonomy*. Thus autonomy in the enactivist sense is not based in some capacity for (self-)reflection. Evan Thompson (2007) defines an autonomous system as one whose components “(i) recursively depend on each other for their generation and their realization as a network, (ii) constitute the system as a unity in whatever domain they exist, and (iii) determine a domain of possible interactions with the environment.” (44) An autopoietic system is an autonomous system that generates both itself and its boundary and regulates the system’s interaction with the environment by way of that boundary. (idem: 126) The simplest example of an autopoietic system is that of a single living cell. A living cell has a semipermeable membrane, which acts as a boundary through which a cell regulates its interaction with the outside world.

In order for a system to be autonomous it must meet two conditions: a system must be operationally closed and it must be adaptive. Operational closure means that every constituent process depends on some other process in the system. The example of a living cell can illustrate this. A cell stands out from its molecular environment by creating a boundary. These boundaries are constructed by metabolic processes within the cell that sustain the cell. At the same time, the metabolic processes are made possible by the very boundaries they create. The metabolic processes enable the boundaries and in turn the boundaries enable the metabolic processes. Should one of these processes stop (the cell no longer self-organises), the cellular components will no longer form a unity and are no longer distinguishable from the molecular environment. (Thompson, 2007: 44)

Adaptivity refers to the capability to regulate the operationally closed processes. (Di Paolo & Thompson, 2014: 73) An adaptive autonomous system strives to maintain its own identity and therefore its existence in precarious conditions, which it registers as better or worse for itself. Certain conditions promote autonomy while other conditions inhibit it. An autonomous system establishes a certain perspective from which interactions with the world acquire a normative status. (ibidem) An illustration can be found in the example of a motile bacteria swimming uphill in a food gradient of sugar as given by Evan Thompson. (2007: 157-158) A motile bacterium always orients itself towards the zone of the greatest sugar concentration. The sugar is not simply just a part of the world the bacteria move in, but the bacteria have a special relationship with sugar. Sugar is important for these bacteria because of the way their metabolism realizes their autonomous organisation. (Thompson & Stapleton, 2009: 25) The significance of sugar for these bacteria is not an intrinsic property of the sugar molecule, but is a relational feature tied to the bacteria’s metabolism. Although sugar is presents in the environment, it gets the status of food in the environment that the organism itself brings into existence. Even simple bacteria regulate their interactions with the world in such a way that they transform the world into a place of significance, meaning and value. (ibidem)

Significance, meaning, and value do not pre-exist (what is food for one organism can be nothing or poisonous to another), but is constituted by living organism on the basis of its own norms that follow from being an autonomous system. The enactivist see this regulation of the interactions between the organism and its environment (i.e., sense-making) as a form of knowledge. Hence, the enactivist speak of cognition. An organism can only exist if it succeeds in finding in the world an adequate environment. We live in a world that is meaningful. A glass of water is not a mere object in the world, but something you can drink.

Other objects not directly related to our survival are also meaningful to us: for example, a chair is something you can sit on. Through the process of self-producing a body enacts a meaningful environment. Thus autonomy and the activity of sense-making are two sides of the same coin. Cognition is therefore the activity of sense-making and basic cognition is a matter of “establishing relevance through the need to maintain an identity.” (Di Paolo & Thompson, 2014:73) According to the autopoietic enactivism there is a deep continuity of life and mind; living is sense-making. Furthermore, cognition and experiential consciousness are part of the same process: that of the lived, embodied action of the organism within its world. Just by being alive and staying alive, an *environment or Umwelt* emerges from the world. (Thompson, 2007: 153).

3.2 Embodied Emotions

Although the enactivists emphasise the embodiment of cognition, so far nothing has been said of emotions. Emotion theorists usually separate arousal, action and appraisal when it comes to emotions. The term ‘arousal’ typically refers to the activity of the autonomic nervous system. The term ‘action’ refers to motor activity that is changed in the musculoskeletal system and facial expressions. ‘Appraisal’ refers to the evaluative part of an emotion and is typically seen as occurring ‘in someone’s head’. (Colombetti & Thompson, 2007: 531) Appraisal is therefore generally understood as separate from the body. In this section, I first discuss Giovanni Colombetti’s general view on emotion. After which I examine her proposal for embodied appraisal in which she tries to integrate the arousal, action and appraisal aspects of emotion.

3.2.1 Emotions as self-organisation

Colombetti (2014) observes that if cognition as the activity of sense-making is an embodied action, then cognition is fundamentally affective. This means that sense-making comprises emotion as much as cognition. Sense-making not only involves conduct in relation to what has salience and value; salience and value imply valence. Sense-making involves *action tendencies*; that what has meaning for us and therefore valence attracts and repels and therefore encourages approach or avoidance. If cognition is described as embodied action and comprises action tendencies, then cognition is essentially affective. Motivated action when it involves affect is a mode of self-regulation. (Thompson & Stapleton, 2009: 26) In Colombetti’s view the mind is fundamentally affective and our emotions and moods are grounded in our ability to be affected.

“[Affectivity] is a broader phenomenon that permeates the mind, necessarily and not merely contingently. The mind, as embodied, is intrinsically or constitutively affective; you cannot take affectivity away from it and still have a mind. Affectivity [...] refers broadly to a *lack of indifference*, and rather a *sensibility* or *interest* for one’s existence.” (Colombetti, 2014: 1)

To be affected is to be ‘touched’ or something has to strike you as meaningful, relevant or salient. Even the simplest forms of life are affective since affectivity depends on the organisational properties of life; Colombetti uses the term *primordial affectivity* to refer to this aspect of our existence. (idem: 2)

According to Colombetti (2014), emotional episodes are best understood as “self-organizing patterns of the organism, best described with the conceptual tools of dynamical systems theory.” (53) Dynamical systems theory (DST) is a branch of mathematics that describes dynamical systems. A dynamic system is a system of elements that changes over time. Self-organisation can be conceptualised in terms of *first-order constraints* and *second-order constraints*.

“First-order constraints refer to the influence that a system’s components exert on one another. [...] Second-order constraints refer to the global influence that macrolevel patterns or *forms* exert on the system’s components.” (idem: 56)

Underlying Colombetti’s theory of emotion is the notion that even the simplest organism stands in a regulatory interaction with its environment. Colombetti notion of affectivity is in line with Spinoza’s notion of *conatus*. Spinoza maintained that all things in existence, even a non-living things has a conatus, i.e. a thing that endeavours to persevere in its being. (Ethics III.6) All existence is intrinsically striving and this striving has in humans both a mental and bodily expression. In Colombetti’s view, affectivity is a property of the organisation of living systems. From the framework of DST she can now conceptualise emotions as self-organizing configurations of the organism, or what she calls *emotion forms*. (Colombetti, 2014: 82) These patterns are flexible in the sense that they can vary from individual to individual and between cultures. The processes constituting them organize themselves differently, depending on context. However, some of them may also reliably appear across cultures. The range of possible emotions is dependent on the state of the organism and therefore evolutionarily and developmentally constrained. Emotion forms are second-order constraints that recruit or entrain various processes (neural, muscular, autonomic etc.) into highly integrated configurations or patterns. (idem: 69)

3.2.2 Embodied appraisal

Colombetti’s view on emotion and the broader enactivist account of the mind that integrates experiential and physical aspects has important implications for our understanding of appraisal. If emotions are understood as self-organizing configurations of the organism, there is no longer a distinction between the arousal part and the appraisal part of the emotion. In other words, an appraisal is constituted by the organism as a whole, meaning when it comes to appraisal you cannot separate the cognitive and emotional parts. This is in line with a meta-analysis of neuroscientific studies by Pessoa, which showed that at the level of the brain where cognition and emotion are integrated at a neural level. Barret & Bar²⁰ hypothesised that vision is supported by affective responses from the get go, thereby arguing against the notion that perception precedes interpretation and appraisal. This implies that appraisal is structurally entangled with emotions (and therefore the body)

Emotion and appraisal are also difficult to separate from a phenomenological perspective. Colombetti observes the following:

“It seems to me that it is not possible to clearly distinguish appraisal and emotion in experience, and in particular it is misleading to suggest that a separate appraisal can

²⁰ For a discussion of both Pessoa and Barret & Bar see chapter 2.1.2.

produce or elicit an emotion in a “linear” way, as it were- that is, to suggest that one *first* (consciously) evaluates something as being a loss, for example, and *then* feels sadness.” (2014: 107)

The point Colombetti draws our attention to is that the phenomenality and intentionality of emotions cannot be seen as separate. When the content of an emotion changes, so the felt quality changes. Examples of this are when fear turns into relief or excitement turns into disappointment. Emotions can be intentional in the sense of an evaluation of our relation with the world through the body. However, the body itself can also be its own intentional object; although emotions are usually felt in the form of the latter. (Colombetti, 2014: 113) The body is not always felt prominently, which does not mean that we are no longer aware of it. Take the sensations of exploratory touch with the hand. When I move my hand over a certain surface, say for instance wood, I can feel its texture. The texture of the wood is the intentional object of my experience, but my hand is not completely absent from the experience itself; my hand is experienced as that through which the texture of the wood is felt.

The experience of bodily feeling can also be diffuse in the sense that feelings are felt throughout the body. (Colombetti, 2014: 118) Take for example the experience of joy, which is not felt in a particular place in the body. The experience of joy might include urges to jump and run; in that case the body is felt as wanting to move. Philosophers typically see these feelings – or better known as *action tendencies* or experiences of *action readiness*- as an important motivational element of emotions. Take for example when you are angry, the urge to hit somebody goes together with the tensing of the muscles and you clench your fists. In this sense (many) emotions have a strong motivational character in that they prepare us for an action. This does not mean that there is a causal relationship between emotions and actions in the sense that a certain action always follows after a certain emotion. The action tendencies are however part of the appraisal.

3.3 The Socially Extended Mind

In the enactivist approach, cognition is not a process that happens in the skull, but is a relational process of sense-making that takes place between the organism and its environment. (Thompson & Stapleton, 2008: 26) Cognition belongs to the relational domain in which an organism as a unity relates to the wider context of its environment. Therefore, cognition is not only embodied, but also *extended*, i.e., the environment is constitutive for cognition. Cognition is not something that only happens in the brain, but also involves elements of the environment; our mind is extended. Since the environment is a constitutive part of many thought processes, these thought processes cannot be reduced to brain mechanics. In this section, I discuss how the mind is *socially extended* according to Shaun Gallagher. If our mind is socially extended, than this implies that the mind in its development is shaped by social interactions. To illustrate this, I also take a look at the developmental roots of the socially extended mind as discussed by Joel Krueger. I conclude this section with the conceptualisation of social encounters as processes of joint sense-making.

3.3.1 Mental institutions

The concept of the extended mind was first introduced by Clark & Chalmers. (1998) Consider the case of Inga and Otto. Inga wants to go to an exhibit at the Museum of Modern Art. She remembers that the museum is on 53rd street, so she walks to 53rd street. Otto also hears about the exhibit in the Museum of Natural Arts and decides to go. However, Otto suffers from Alzheimer's disease and has to rely on information from the environment to help structure his mind. Otto always carries a notebook with him where writes down new information and looks up old information when he needs to. So for the exhibition Otto consults his notebook, which says that the museum is located on 53rd street. According to Clark & Chalmers, the cases of Inga and the case of Otto are analogous. (idem: 13) Inga believes that the museum is on 53rd street, and she believed this before she consulted her memory. The belief was memorised, waiting to be accessed. The same can be said of Otto. However, it is his notebook that serves the function of his biological memory. Just as Inga believed that the museum was at 53rd street before she consulted her memory, it can be said that Otto believed that the museum was at 53rd street before he consulted his notebook. Clark & Chalmers argue that Otto is best regarded as an extended system, a coupling between a biological organism and external resources. (idem: 18) This shows that cognition does not stop at the skin and the skull, but extends in Otto's case to his notebook. His notebook is part of his cognitive system.

Shaun Gallagher takes a more, in his own words, liberal interpretation of Clark & Chalmers' concept of the extended mind. According to him, the mind is not only extended in the sense that it supervenes on processes in the environment, but the mind is also socially extended to include processes that occur within certain social and cultural institutions.²¹(Gallagher, 2013: 4) In these so called *mental institutions* certain cognitive processes are accomplished and without these institutions such cognitive processes would no longer take place. Mental institutions include "cognitive practices that are produced in specific time and places. [A mental institution is] activated in ways that extend our cognitive processes when we engage with them (that is when we interact with, or are enactively coupled to them in the right way." (idem: 6) Mental institutions encode the collective wisdom of our cultural environment and enable some of these processes occurring within these institutions. Gallagher suggests the legal system as a good example of such a mental institution.

The hypothesis of the socially extended mind suggests that if we want to study the mind, we should explore how social and cultural practices shape the mind. Without such practices our cognitive processes would be different. In other words, in studying the mind we should look for those institutions that are formative in the early years of a child. Gallagher suggest that the family can be considered as ontogenetically the first mental institution. Thus studying the family can uncover the developmental roots of the extended mind by studying how basic embodied and situated processes of primary and secondary intersubjectivity can shape a child's cognitive habits. This research trajectory was taken up by Joel Krueger (2013), who examined the embodied practices that comprise early infant-caregiver interaction. According to him, there are physical interventions that caregivers use to regulate attention and emotion. These physical interventions allow an infant to perform cognitive tasks

²¹ Gallagher not only differs with Clark and Chalmers on whether or not cognition is extended to social institutions, but Gallagher also rejects the belief-desire psychology that Chalmers adheres to. For Chalmers a cognitive process is extended, when the process outside of the head is similar to a similar process that would happen in the head. However, Gallagher believes that the extended mind hypothesis challenges the standard belief-desire psychology. I will delve any further into this discussion as it falls outside the scope of my thesis. For more on the difference between Chalmers and Gallagher see Gallagher (2013).

she could otherwise not perform by herself. Secondly, Krueger argues, these physical interventions encode “the norms, values, and patterned practices distinctive of their specific sociocultural milieu.” (idem: 40) He explains this with the following example: the cognition of students is shaped within the mental institution of education. The attention of students is (partly) managed by the spatial configurations of their environment. Students must learn what to pay attention to and what to ignore given a certain task. Classrooms place the instructor in front of the room, which is organised in this specific way to guide visual attention to the teacher, thereby simplifying perceptual choices and heighten attentional focus. Teaching practises within the institution of education codify the cumulative wisdom of previous generations of teachers; thus the institution and its environmental arrangement instantiates a *sociocultural norm* – such as the teacher should be the centre of attention and authority while speaking. (ibidem) (Physical) interventions of caregivers function as embodied social practices that scaffold the infant’s capacity for attention and emotional regulation and this scaffolding is, according to Krueger, an external mechanism for both cognitive enhancement as well as cultural entrainment. (idem: 42)

Like Gallagher, Krueger maintains that mental institutions operate from a child’s birth²², and the family can be seen as a cognitive enhancing mental institution. The cultural entrainment also begins from day one. The cognitive scaffolding provided by caregivers’ embodied practices is crucial for the infant’s psychosocial development during the first months when infants have little control over their attention or emotions. (idem: 43) This becomes apparent in the activity of breastfeeding, which is one of the earliest complex forms of social interaction. Early infant attention is primarily controlled by objects and events outside of the child. Contrast this to adults who have top-down, voluntary control over their attention. I can be distracted during a certain task, but I can draw my attention away from the distraction. Young infants do not possess these skills. An important element of the activity of breastfeeding is the rhythmic cycles and back-and-forth interplay between mother and child. In this interplay, the mother and child form a coupled social system. Humans are the only mammals that breastfeeds in short bursts. Human infants suckle in burst of 4-10 sucks at about one per second, separated by pauses of about 4-15 seconds. Mothers instinctively try to intervene in these burst-patterns by jiggling the baby when breastfeeding or jiggling the bottle when feeding with a bottle. (Kaye, 1982: 37-38) The jiggling and stopping behaviour of the mother elicits a burst of sucking from the baby. Although the rhythmic interplay between mother and child is mechanical, this ‘basic’ interplay is already a complicated and meaningful interaction. The baby and the mother react to and influence each other’s behaviour.

“Within the dynamics of this exchange mothers sculpt the infant’s attention: their behaviour is organised by the mother’s touch and physical prompting. The infant is guided to notice salient environmental affordances (e.g., the nipple affording feeding) that, in light of her underdeveloped endogenous attention and lack of behavioral [sic] organization, she might not otherwise pick up on.” (Krueger, 2013: 43)

It is the touch and gentle caresses of the mother to which the infants respond, and mothers adept to the bout-pause behaviour of the infant’s sucking; the infant is not merely a passive participant but reliably postpones her sucking until the mother ends her tactile behaviour, thus making the activity of (breast)feeding a mutually-governed interaction, or a form of

²² Probably even before they are born, I would speculate, since an unborn baby is already exposed to sounds external to the womb and more importantly in constant interaction with the mother.

participatory sense-making (see next section). (ibidem) Thus Krueger argues that this feeding practice is an instance of an embodied practice, embedded within the mental institution of the family that generates an extended cognitive process that develop a child's voluntary attention. In the dynamic between mother and child new cognitive abilities emerge that would not have arisen outside of this coupled system. (idem, 44) The child gains a new level of attentional focussing that exceeds the current phase of the child's cognitive development. These low-level perceptual and affective processes (between mother and child) are crucial for building up a child's interpersonal competence and form the foundation of various (embodied) skills that inform our ability to smoothly engage with others in our adult life.

3.3.2 Participatory sense-making

This coupling between mother and child is an early example of participatory sense-making. Participatory sense-making is a term first coined by De Jaegher & Di Paolo (2007), and describes the activity of sense-making in the social domain. The above given account of mother and a child suggest that interaction is central to our cognitive development. De Jaegher & Di Paolo aimed to put the process of interaction between humans at the centre of their investigation and in turn describe those processes with concepts from dynamical systems theory. According to them, "interactions are processes extended in time with a rich structure that is only apparent at the relational level." (idem: 490) One important notion used in DST is that of *coordination*; coordination is, in terms of DST, a form of *coupling* between systems. Two or more dynamical systems are coupled when they reciprocally influence and constrain their behaviour over time, such that they can be modelled over time. (Colombetti, 2014: 55) Coordination is then understood as the non-accidental *correlation* between the behaviours of two or more systems that are in sustained coupling, or have been coupled or have both been coupled to a common system. A correlation is a coherence in the behaviour of two or more systems over and above what is expected of those systems. (De Jaegher & Di Paolo, 2007: 490) The phenomenon of coordination is often illustrated with the example of pendulum clocks mounted on a wall. These pendulum clocks, when allowed to oscillate freely, synchronise when in each other's vicinity. This happens through minute vibrations the clocks provoke on the walls. That happens because the frequency of the vibrations caused by one pendulum influences the movement of the other pendulum. What is important to understand is that coordination is something that is generally easily achieved and does not require advanced cognitive skills, but can even be achieved via simple mechanical means. (ibidem) Synchronisation is not the only form of coordination; coordination also includes such things as mirroring, anticipation, imitation etc. Coordination does not have to be permanent, coordination may break down and there can be different levels of coordination. De Jaegher & Di Paolo call the acquiring of coherence between the sense-making activities of interactors through their interaction participatory sense-making. Thus they define participatory sense-making as "the coordination of intentional activity in interaction, whereby individual sense-making processes are affected and new domains of social sense-making may be generated that were not available to the each individual on her own." (idem: 497) Note that social interaction is not merely a coupling between agents; the coupling has to be regulated by the interactors. Thus the transfer of body heat between those waiting for a bus at a crowded stop are not interacting socially in this respect, since the transfer of body heat is not regulated. De Jaegher & Di Paolo provide the following example:

“Consider the situation in a narrow corridor when two people walking in opposite directions have to get past each other. They have to decide whether to continue walking as they are, or shift their movement to the right or the left. Occasionally such encounters unfold like this. Instead of choosing complementary movements that would allow them to carry on walking, the individuals move into mirroring positions at the same time. This unintended coordinated change in individual positions creates a symmetrical mirroring relation. This symmetry, in combination with spatial constraints of the corridor, increases the likelihood that the next move will also be a mirroring one (there are not many other moves available). Thus, the coordination maintains a property of the relational dynamics that forces the individuals to keep facing each other and consequently remain in interaction (in spite of, or rather because of, their efforts to break from this situation).” (idem: 493)

It is important to note that both agents in this encounter are attempting to stop interacting, but the interaction continues to self-sustain in spite of this. Only when the symmetry in movements is broken, for instance by one person inviting the other to move first, does the interaction break down. De Jaeger & Di Paolo claim that the coordination patterns (the coordinated lateral shifts in position) sustain the interaction, which in turn promotes through its relational symmetry certain patterns of coordination. (ibidem) Thus, both persons mutually influence each other. In this example, the interaction is not reducible to the actions of the individuals, but establishes a *relational domain* with its own properties that constrain and modulate individual behaviour. However, this does not mean that there are no individual elements in the process of interaction. Within a social engagement the interactors have to remain autonomous throughout the engagement in order to properly speak of social interaction. (idem: 494) To illustrate this, think of two dancers. Dancing involves moving the other and being moved by the other. Even though there is a leader and a follower, following is part of the agreement and is not reducible to being placed in a certain position by the other. If the follower had no autonomy, there would be no dance, but the interaction would be more akin to one person dragging a doll over the dancefloor. What is important here is that the interaction is not only structured by the agent interactors, but in turn structures the behaviour of the agent interactors themselves. Interaction is a *structured and structuring* process.

Colombetti & Torrance are quick to point out that participatory sense-making as an embodied activity has an inescapable *affective dimension*.

“Yet if sense-making [...] is inherently affective already at the level of the individual organism, then participatory sense-making is not just ‘often’ but *always* affect-laden; autonomous organisms bring to their encounter their own forms of cognitive as well as *affective* understanding, and as a consequence affectivity is perturbed and transformed as the encounter unfolds, and as it generates its own meaning.” (Colombetti & Torrance, 2009, 507)

That participatory sense-making is affect-laden becomes clear in the example given above of the mother feeding her infant. It is the interplay between mother and child that endow the nipple or the bottle with the meaning of affording feeding. Note that this interaction between mother and infant happens without any discursive thought. I do not mean that the mother lacks all discursive thought while interacting. Only that the specific coordination between mother and child happens affectively without any specific discursive thought preceding the coordination, i.e. the jiggling done by the mother happens instinctively. The experience is

fundamentally affective and pre-reflective, in the sense of not reflected upon but still characterising the experience. This experience is an experience of *connectedness*. Participatory sense-making involves a concrete encounter between the self and the other that is fundamentally a process of empathy.

Empathy is a term that has many different conceptions. Commonly a psychological distinction is made between *basic empathy* and *cognitive empathy*.²³ The former is characterised as the ability of emotional contagion. Emotional contagion is the ‘primitive’ ability to *feel* what others are feeling. This generally happens involuntary and automatic. Cognitive empathy is the capacity to understand the mental state of another, ‘based on’ taking on their perspective. I will however not draw upon a psychological conception of empathy, but a *phenomenological* one. A phenomenological conception of empathy is not only interested in the relevant cognitive and affective processes involving self and others, but takes the lived experience of an other’s experience as central. I take the notion of emotional contagion as too narrow to characterise the basic experience of empathy. The notion of cognitive empathy implies that we can only gain access to other people’s mental states via a process of mind reading, i.e. inferring that the other is experiencing something that is alike to my experiences. Instead, I take basic empathy to be the experience of the other as a subject. From a phenomenological approach, empathy is a form of intentionality that discloses an experience that is not ours. Thus a phenomenological approach assumes that that empathy is not reducible to some additive combination of perception and inference. (Thompson, 2007: 386) To support this notion of empathy, I will rely on Edith Stein’s phenomenology of empathy.

Stein (1917) emphasised that empathy is not a kind of judgement, but empathy is a kind of act of perceiving *sui generis*. (11) Empathy is feeling the other’s subjectivity non-primordially (in the sense that the feelings of the other are not felt as my own feelings). Empathy can be seen as any intentional act that discloses or presents the other’s experience. The experience of the other cannot be disclosed in its original first-person subjectivity, but is experienced as a second-person perspective – hence the nonprimordially of empathy. In empathy another (human) being is experienced “directly as a person – that is, an intentional being whose bodily gestures and actions are expressive of his or her experiences or states of mind.” (Thompson, 2007: 386) This notion of empathy corresponds to what Evan Thompson calls *affective and sensorimotor coupling*. Examples like the interplay between mother and child while feeding are not merely mechanical interactions, but also examples of empathically experiencing another as a person. Thompson also distinguishes two other forms of empathy, namely *imaginary transposition* into the other’s place (which comes close to the traditional conception of cognitive empathy) and *understanding* the other as an alter for whom I am an other. (idem: 382-411)²⁴ Imaginary transposition involves taking up the other’s perspective. The workings of this kind of empathy become apparent in the emotion of envy.

“In envy, we sense the other’s pleasure in possessing the good we desire, via a phenomenally inextricable mixture of imagining and taking up the other’s experience

²³ This distinction between basic empathy and cognitive empathy is not meant to exhaust all views on what empathy means. Nor do I wish to imply that it is agreed upon by all that both of these are forms of empathy.

²⁴ Thompson also distinguishes *moral perception* as a form of empathy. I will (indirectly) address this issue in the next chapter.

from *her* point of view, and imagining *ourselves* in her position.” (Colombetti & Torrance: 2009: 513)

The presence of the other – i.e. the understanding of you as an *other* me, and me as an other you – is not always transparent. Not all interactions occur smoothly. There are occasions where the alterity of the other is experienced, that is the experience of the other is more ‘opaque’. This means that the experience of the other can be conflictual in nature. Colombetti & Torrance give the example of antipathy as a form of ‘dissonant connectedness’ in which basic empathy and sense of alterity are subtly entwined. (2009: 512) It is the experience of the other that we resist and want to disengage with. The presence of empathy in participatory sense-making does not mean that the interaction is moral or pro-social. In fact, empathy is necessary to enjoy someone’s suffering. Neither does participatory sense-making imply that the interaction consist always of morally good behaviour. On the contrary, participatory sense-making can end up in the destruction of the autonomy of one of the interactors, such in the case of torture.²⁵ The point is, as Colombetti & Torrance remark, “that ‘feelings of connectedness’ involve a *complex interplay* of various levels of empathy or other grasping.” Processes of empathy underlie all (face-to-face) social interactions and therefore all ethical interactions. Thus we experience other people as autonomous subjects with their own experiences; in the encounter we see the other as an agent. This experience of others in participatory sense-making is pervasively affective, but also shows an ethical dimension. Participatory sense-making always happens against a background of broad social ‘norms’ and practices, which often have a strong ethical character.

To summarise, according to the enactivists cognition is the activity of sense-making and is an embodied, situated process. Furthermore, affect and cognition are no longer distinct. If sense-making can occur socially, then this implies that appraisals (as a form of sense-making) can be the outcome of a social interaction. Furthermore, an emotion and an appraisal can no longer be separated from each other and this also holds for a moral appraisal. The enactivist framework also implies that our ability to make appraise morally is dependent on the autonomous nature of living systems, since all our sense-making activities are dependent upon our ability to self-organise. From this framework an embodied, situated view of moral agency arises that is constituted by both rational and emotional processes. I will develop this view in the next chapter.

²⁵Although ones the autonomy of one of the interactors is destroyed the interaction ceases to be a social interaction.

4. Ethical Know-How

"It is not an ethical 'ought' that conduct should be social. It is social, whether bad or good."

John Dewey (1922: 17)

My aim in this thesis is to explore an alternative to the standard view of moral agency, which avoids the problematic assumptions underlying the standard view I discussed in chapter 2, such as the emotion-cognition distinction, the strong contrasting of controlled versus automatic processes, equating the former with consciousness and the latter with subconscious processes. In chapter 3, I showed that in the enactivist approach, thinking and emotions are integrated, and thinking must also be understood as socially embedded and extended. Building upon what has been argued hitherto, I present, in this chapter, a sketch of an embodied and situated account of ethical agency. I argue that ethics can be conceptualised as a skilful activity or a skilful coping that is dependent on constitutive processes of the body and is constituted by (our history of) social interactions. Ethical agency is generally not restricted to the practice of making judgments on whether or not an action is right, but of immediate coping with the social world.

Since I approach the notion of moral agency primarily with the aid of an enactivist framework, I will use Francisco Varela's conception of ethics from an enactivist background as a starting point. Varela conceptualises the skills employed in ethical activity as *ethical know-how*.²⁶ (1992) He criticises the standard conception of morality as only employing abstract prescriptive principles; ethics is more a matter of a person unthinkingly rushing to help at the scene of an accident. In this spontaneous coping there is no *I* performing a deliberate action. From the enactivist framework cognition is seen as grounded in the concrete activity of the whole organism. Cognition is therefore seen as embodied action. From this perspective the enactivists, in line with Merleau-Ponty, consider the world not as pregiven, but enacted through our history of structural coupling. (Varela, 1992: 17) All living systems are characterized as making sense of their world by being autonomous and therefore all living systems have a certain perspective on the world. Therefore, as a result of engaging in sense-making our relationship to the world is always meaningful. Our primary way of relating to things is skilful (and therefore requires know-how) and bodily. According to Merleau-Ponty and the enactivists, we perceive things not merely by its objective and determinate features, but we perceive things pragmatically in the light of a contextual motor goal effected by one's body. (Merleau-Ponty, 1945: 127; Thompson, 2007: 247) We see the world in possibilities for action, or what Gibson (1979) calls an *affordance*; a cup is something to grasp, a glass of water is something to drink from. An affordance leads to a readiness-for-action.

Because of the transparency of our immediate coping, ethical know-how is not in the foreground of our experience. *We know-how by means of our habits*. According to Varela: (1992) "we always operate in some kind of immediacy of a given situation." (9) Most of the time, we do not have to think about what we are doing. Take our example of seeing

²⁶ Know-how is contrasted with know-that. The difference between know-how and know-that was first clearly formulated by John Dewey (1922), and is most known from Gilbert Ryle's essay *Knowing How and Knowing That*. Know-how is knowledge you have when you can truly do something, like knowing how to ride a bicycle. Know-that is knowledge of some fact, like 'the earth is the third planet from the sun'. (Ryle, 1945-1946) Know-that is usually construed as propositional in nature.

somebody fall over. In Varela's view we always act within specific instead of general situations. Our complex of know-how results in having a readiness-for-action (or action tendency) proper to every specific lived situation and we move from one readiness-for-action to another. Varela calls a readiness-for-action a *microidentity* and its corresponding lived situation a *microworld*. (idem: 10) These microworlds have a recurrence (a familiarity) and a stream of moving from one microworld to another embodies appropriate action. Within this picture appropriate refers to the ability we have to read the demands and dynamics of each situation we go through every moment of the day and be able to respond accordingly. The point Varela wants to make is that 'who we are' consist of already constituted microworlds and "who we are', at any moment cannot be divorced from what other things and who other people are to us." (ibidem) Microworlds are shaped by our social interactions. Only when we go to a foreign country, there is no recurrence of a microworld, but in our everyday experience we are surrounded by what is familiar: family, school and restaurants all form microworlds. This means that microworlds and microidentities are historically constituted (this does not mean that microworlds are fixed and not subject to change). (ibidem)

Varela's account is in turn heavily influenced by the works of Hubert Dreyfus. Dreyfus also claims that skills form an important part of ethics and applied his understanding of engaged coping to ethics, calling it the *phenomenology of ethical expertise*. (Dreyfus & Dreyfus, 1991) Dreyfus was interested in activities that did not issue from a process of reflection and deliberation. In Dreyfus views activities such as driving or playing chess are first learned through applying rules of the activity explicitly. With practice proficiency increases and once one reaches the level of an expert, the person is no longer conscious of the rules he is following and he then is completely immersed in the activity.

In light of this thesis Varela's account offers two interesting lines of investigation: the social aspect of ethical agency and ethical agency as a skilful activity. I start out with a discussion of the social aspects of ethical agency (§4.1). Varela's notions of microworlds suggest that ethics has a fundamental social aspect. I argue that these microworlds are the result of what De Jaegher & Di Paolo called participatory sense-making. I believe the use of participatory-sense-making as a conceptual tool is useful tool for analysing our social and ethical interactions. In this I follow Colombetti and Torrance (2009) in that ethical appraisal is a kind of participatory sense-making. Therefore an ethical appraisal should be understood as the result of a social interaction. Because ethical appraisal is to be understood as the outcome of a social interaction, I argue that ethics is fundamentally intersubjective. In order to clarify this point I contrast two models of agency. The first model I call the *dynamic model of ethical agency*, where ethical sense-making is understood as a dynamic interaction between two or more agents. The second model I call the *linear model of moral agency*, where ethical sense-making is seen as an individual input-output process. The main difference is that in the dynamical model instead of the individual, the interaction is taken as central. In my discussion of moral agency, I conceive ethics as broader than encompassing what action is morally right or wrong to perform.²⁷

²⁷ I maintain a distinction between ethics and the moral, but this distinction should not be seen as absolute or even as a clear distinction. I do think that a distinction is useful, mainly when it comes to the motivational aspect of actions. Everyone will agree, e.g., that drinking a cup of tea is (typically) not a moral matter. The act of drinking tea can take place in a social context with moral importance, implying certain moral obligations. (For instance, various Asian tea ceremonies) However, not drinking tea is in itself not a moral transgression I refer to ethics because there are cases which we do not want to dub moral and this allows me to distinguish between cases which are explicitly ethical and moral, and those which are implicitly ethical.

Furthermore, I discuss how to conceptualize ethical agency as a *skilful activity* (§4.2). I discuss Dreyfuss' notion of ethical expertise and address certain shortcomings to his theory. Drawing on Hubert Dreyfus phenomenology of ethical expertise and Varela's account of immediate coping, I argue that ethics involves *unreflective ethical know-how* and should therefore be considered more as a skill. To be an ethical expert is to be a skilful participant of an ethical community. Thus an ethical expert relies less on explicit reasoning processes than a novice.

After exploring these two lines of investigation, I conclude with discussing the relation between an ethical agent and what I call the *ethical world* (§4.3). I argue that ethical expertise gives us access to the ethical world, which is the whole of collective ethical sense-making practices. The ethical world is a joint achievement that has to be enacted by the ethical agents of a community. However in turn, these agents, as individuals, are shaped by the ethical world. Therefore, not only the ethical world is co-constituted but also the agents themselves.

As I already stated, in this chapter I formulate a possible alternative account of ethical agency. I do not deny the important role reason and reflection play; they are essential for our ethical practices. As humans we are thinking agents, who as individuals can have an impact on what is right or wrong. However, I will focus in this account on unreflective skilful action. It is important to understand the human mind not only as socially embedded, but also how our mind is formed by the very social encounters we as humans live in. As stated, I take ethics as essentially interpersonal; therefore our ethical appraisal must be understood as an interaction between two or more ethical agents against an already established background of norms and practices. To interact with other humans is a skill everyone needs to learn and we learn those skills by living in an ethical world. Ethics is a skill that allows us to access the ethical world. I will argue that accessing the ethical world is the same as enacting ethical meaning.

4.1 Two Models of Moral Agency

If human cognition is socially extended, then not only is our mind influenced by social processes, but social processes shape our cognition and allow us to perform or participate in cognitive tasks we are incapable of performing individually. This means that our cognition is shaped by, what Gallagher calls, mental institutions from day one. Thompson (2007) remarks on the nature of cognition as socially scaffolded:

“One of the most important reasons that human mentality cannot be reduced simply to what goes on inside the brain of an individual is that human mental activity is fundamentally social and cultural. Culture is no mere external addition or support to cognition; it is woven into the very fabric of each human mind from the beginning.”
(403)

Krueger demonstrated how processes of participatory sense-making shape and extend the cognitive abilities of infants. If our cognitive abilities are shaped from infancy by social processes, so are our ethical capacities shaped by these same social practices. Thus our ethics and ethical appraisal have an intersubjective basis. In this section, I follow Colombetti

& Torrance argument that ethical appraisal is a form of participatory sense-making and will base the *dynamical model* of ethical agency on this point. However, in order to clarify my position I will first discuss the model with which this contrasts and what I call the *linear model* of moral agency. The standard view of moral agency is such an example of the linear model.

4.1.1 The linear model of moral agency

In chapter one, I argued the standard view on moral agency as encompassing all aspects of the ethical life. In the standard view making moral judgements on the basis of moral deliberation is what makes us moral agents. I also argued that the standard view maintains the assumption that there is a strict separation between cognitive and emotional processes. This assumption is not only to be found in the traditional view, but also in many moral psychological accounts such as Haidt's SIM and Greene's DP-theory. Not only do these theories share the same basic assumption of a separation between cognition and emotion, I argue that these models rely on the same theoretical framework, which I call the 'linear model of moral agency' (see figure 1).

Not only do the standard view and the above discussed dual-process theories share many assumptions, but they also seem to function according to a basic input-output schema. Hence, I call this framework the 'linear model of moral agency'. First, an event or object is observed by the moral agent. This perception is still neutral and serves as the input. Then morally relevant properties are identified (or attributed) and the some form of moral processing takes place. Moral processing should be understood in the broadest sense possible; moral processing can take the form of voluntary conscious deliberation or automatic intuitive responses. The output of this processing is a moral judgement. This moral judgement can lead to an action, but an action does not necessarily follow from the judgement. Thus perception, cognition (or in the case of noncognitive theories emotive/intuitive responses) and action are distinctly separated from each other. It should be noted that this account does not fully do justice to the many sophisticated forms moral processing can take or the nuanced interactions between all elements many models posits. Even Haidt's SIM, which is generally considered to be a simple sentimentalist model, has different ways in which our moral processing can be shaped and even our judgements can influence our moral processing. Nor do I mean to equate the moral psychologist with the standard view. Both deal with different question, the former are interested in providing a causal story of moral judgement, the latter a semantic account. However the point is that the linear model of morel agency forms the core of the models of moral agency as understood by the standard view, the SIM and Greene's DP-model.²⁸

Another important element of the linear model is its individuocentric nature. This means that the linear model and all models based on it take the individual as its centre. Taking again Haidt's Social Intuitionist model as an example, we can still see a bias to explain moral behaviour from the standpoint of just the individual. Despite the fact that Haidt introduces an important social aspect in his model, namely that moral reasoning mainly occurs socially, the moral judgements are only caused by an internal process. The social aspect of Haidt's model lies in the ability to influence the moral processing of other individuals. Although the engagement of the individual with its social environment is important when it comes to moral reasoning or the adjustment of intuitions through other

²⁸ I would even argue that the linear model holds for almost every contemporary theory of moral agency. However, in this thesis I will keep the discussion contained to these three.

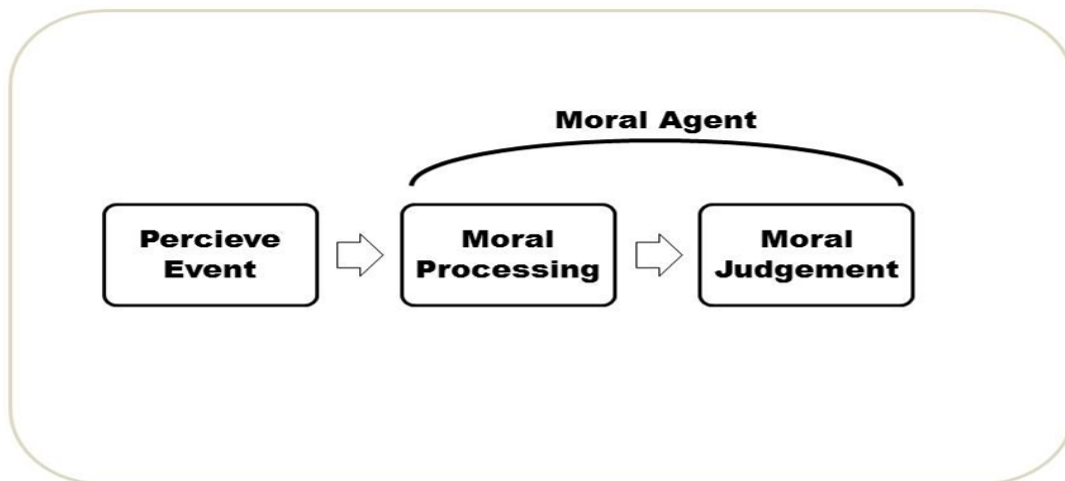


Figure 1 The linear model of moral agency

means, the social interaction does not play a constitutive role in judgement formation. It is the formation of a moral judgement through a correct form of moral processing that determines whether someone is a moral agent or not. This means that a social interaction is in principle not a necessary part of the formation of a moral judgement. Therefore a moral agent in Haidt's SIM is socially embedded, but moral agency is not *socially extended*.

Moral agency in SIM ultimately lies in the individuals own capacities and not the engagement of the individual with its social environment. In the linear view, ethics is concerned with how an individual should act or what kind of person an individual should be. An ethical society from this point of view is built up from the individual ethical acts, which hopefully form a coherent whole (or at the least harmonious). Thus any ethical rule or analysis is aimed at the action of the individual. There is nothing wrong with taking this individuocentric stance per se, however I argue, this stance cannot account for the dynamics of human social sense-making that make up the bulk of human ethical behaviour. The ethical implications of participatory sense-making cannot be thematised from the linear model since all ethical sense-making in this model happens at a fundamental individual level. Although models based on the linear model can take into account the interaction of people (to a certain extent), it cannot be the focus of the models. Thus seeing that the mind is socially extended, there is no reason to take the individuocentric stance as default when it comes to the formation of a moral appraisal.

4.1.2 The dynamic model of ethical agency

Take again the example of De Jaegher and Di Paolo (2007) of two people moving towards each other in a small hallway. (493) The mirroring actions between the interactors gain an autonomy of their own. The coordination is not done purposefully, but leads to a negotiation of the space both interactors move in. While both interactors mirror each other in their swaying from left to right, eventually this specific interaction has to be terminated otherwise they would not be able to pass each other. Colombetti & Torrance (2009) argue that this very simple case of interaction-autonomy contains elements that can be applied to a very broad range of situations, not all of them just belonging to situations of simple physical or bodily coordination. (519) Not only are the dynamics of movement negotiated, who gives way to the other, but this simple interaction also involves ethical dynamics. In this interaction, and in every interaction, broad social cultural practices play a background role. Both agent

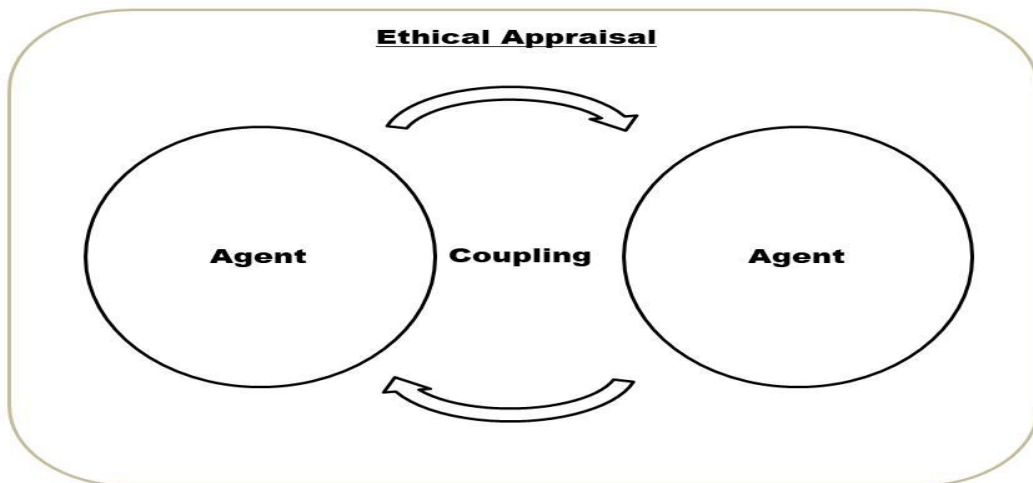


Figure 2 The dynamic model of ethical agency

interactors implicitly abide to the norm of 'not pushing each other out of the way'. Thus the interaction does not gain an aggressive character; neither of the interactors sees the other as invading their space. The interactors do not ascribe malign intentions to the other, because each of the interactors adheres to an implicitly present ethical norm about how to negotiate each's personal space. The social standing between the two interactors also plays a significant role in the creating of meaning in this interaction. The interaction will unfold differently if the interaction is between two students who stand on equal social footing or between a professor and a first-year student. Thus Colombetti & Torrance comment on this example:

"These various implicit social-ethical norms are reaffirmed and reinterpreted by the agents' interaction. So, even in this very simple kind of case, the interactors find themselves engaged in an implicitly *moral* transaction, rather than simply a straightforwardly 'physical' manoeuvre, or a social encounter with no ethical overtones." (Ibidem)

When it comes to ethical appraisal, they propose two shifts. (idem: 523) The first shift is to see the ethical valuation of a given situation as emerging as much from the interaction of the participants as from the autonomous decision-making or original authorships of the participants themselves. This means that the focus of appraisal is less on the individuals and more on the ethical qualities of the interaction itself. The second shift contains the notion that social interaction between agents is never ethically neutral.

Based upon Colombetti & Torrance's reappraisal of ethical appraisal, I propose in a dynamic model of ethical agency (figure 2). The dynamic model takes the interaction as central and not the individual moral agent. This account centres on participatory sense-making and the notion that the ethical dimension of a certain interaction or situation arises (partly) from the meanings which emerge from the interactors. This does not mean that the interactors are not autonomous and therefore not agents, the autonomy of the interactors is necessary in order to properly speak of a socio-ethical interaction. The interaction depicted in figure 1 is of a dyadic nature, but this does not mean that all ethical interactions are between two persons.

Thus the main difference between the dynamical model and the linear model is that the dynamical model is fundamentally intersubjective. Contrary to the linear model, where an individual forms a judgement, in the dynamic model the appraisal is the result of the

coordination between the moral agents involved. Thus an ethical appraisal is not only co-constituted by the interactors, but the meaning generated by the encounter gains its own autonomy.

Another difference with the linear model is that the event or object, which was perceived, is not separated from the moral processing. No event or object is perceived in an ethically neutral way, since a (moral) agent constitutes a meaningful environment of its own. In the linear model, an event is first perceived, before any meaning is ascribed to it. Secondly, the event that requires moral appraisal usually revolves around other agents, who bring their own meanings to the interaction. Thirdly, the meaning of the event or object changes during the interaction. Hence the event cannot be separated from the moral appraisal. To illustrate this point I turn to an example given by Maureen Sie (2013: 276-277):

“Some time ago my colleague, Arno, left his bicycle pump in the corner of the café where he had lunch. He wanted to have his hands free on his after-lunch walk, and the barman of the café promised to keep an eye on the pump. When Arno came back, the pump had disappeared. Upon asking the barman where his pump went, the barman shrugged his shoulders; he could not remember (and was busy anyway). In response to this attitude, Arno blamed the barman for being inattentive; did he not ask him explicitly to keep an eye on it? The barman, feeling guilty now, tried to remember what happened to the pump and realized he might have allowed a customer who often visits the café, to use the pump. He vaguely remembers her entering the pub looking for someone to help her with a punctured tire. He now realizes she must have forgotten to bring it back after using it. The barman promises Arno to call her to account at her next visit. A week later, a friendly young woman rang at Arno’s door, handed over his pump and apologized for having taken it. She had assumed that the owner of the pump had forgotten it and would not miss it. Also, she had assumed that the barman, who she knows from her regular visits to the pub, meant her to take it home when he said that she “could take it.” Hence, she had taken the pump on the road to pump up her punctured tire a second time on her way home. It was only after the barman asked her whether she took the pump and why, that she realized that the owner did miss the pump and that the barman had not intended to allow her to take it as her own.”²⁹

For Arno, the protagonist of this story, the meaning of the event ‘the women taking the bicycle pump at the pub’ changes meaning through the interactions with the other participants. By blaming the bartender, Arno succeeds in drawing the attention of the bartender to the significance that the losing the pump for Arno has; the bartender no longer shrugs his shoulder. Now the sharing the significance of the lost pump with the bartender. Although the bartender has come to see the moral significance of losing the pump, the event has also changed in meaning for Arno. Where at first Arno saw the bartender as breaking a promise, now both Arno and the bartender share the view that the woman who stole the pump is to blame. Thus the moral transgression is no longer the broken promise of the bartender, but the theft of the woman. The interaction again changes meaning when Arno directly interacts with the women. Now the event is seen as a simple misunderstanding.

The interaction between Arno, the bartender and the women cannot be reduced to the individual actions of the interactors. The meaning that is generated through this

²⁹ Sie uses this example to illustrate how through blame moral responsibility is ‘ascribed’ to one of the participants.

interaction is not fully under the control of Arno or any of the other interactors. Nor can the autonomy of the interactors be denied, each bring their own ethical norms and expectations into the interaction. No ethical interactions take place without a broad range of ethical practices from which the agents operate and which allow them to interact with each other. The production of meaning is not only an individual enterprise and “participatory sense-making is always shaped by super-individual norms and institutional practices.” (Gallagher, 2013: 8) The dynamics of participatory sense-making partly gain their character from the wider context of social norms the interactors inherited. Participatory sense-making is in part re-enacting these broader social and ethical practices. Without these shared practices the above sketched interaction would not be possible. What is also important to note is the contribution of the role of emotion to the sense-making process. The guilty feelings of the barman allowed him and Arno to come to a shared appraisal of the situation. This implies that cognition³⁰ and emotion are not strictly separated. Thus an ethical appraisal is not only a pattern of self-organisation in which the agent (re-)evaluates his relationship with the world, but most of all an evaluation and organisation of the social world and its inhabitants.

I already argued for the deep connection between emotions and ethics, and I share Colombetti & Torrance assessment that “the pervasively affective nature of participatory sense-making determines the ethical character of encounters.” (2009: 524) No social interaction is ethically neutral. This also means that ethical sense-making is a communal project; our appraisals and norms are never fully our own. Thus the social world is also an ethical world. These socio-ethical practices are mind-independent, that is that they do not belong to an individual; they have a kind of autonomy and are fundamentally intersubjective. Ethical sense-making is constituted by an ethical negotiating against a broad background of established norms and practices. This negotiating or coordination between moral agents need not be discursive, but always contains an affective richness.

4.2 Ethics as a Skill

The examples of two persons who encounter each other in a narrow corridor and the example of Huckleberry Finn who helps his friend Jim (as discussed in chapter 1) suggest that ethical sense-making does not always require judgements as consciously held beliefs. The ethical appraisals in these cases do not require cognition as understood in the typical sense. Not all ethical appraisals issue from a process of reflection and deliberation. On the contrary, navigating a corridor seems to involve skilful action instead of reflection. We can do many things without thinking about them, we know how to use utensils at the table, how to use a door, how to engage in conversation with others. We know *how* to do these things without reflecting on them or even bringing them to the foreground of our consciences. The knowledge employed in these kinds of activities is best described as *skilful* knowledge or know-how. When at the table we employ skills to sit, to use the utensils; we also employ skills to observe the correct etiquette when sitting at the table and how to perform a conversation.

As already discussed, one influential account of ethical skill and know-how can be found in the works of Dreyfuss. In this section I will address and expand upon Dreyfuss' notion of skill and know-how. Dreyfus was interested in activities that did not issue from a process of reflection and deliberation. In Dreyfus view activities such as driving or playing

³⁰ Or what is typically considered to be a cognitive process.

chess are first learned through applying rules of the activity explicitly. With practice proficiency increases and once one reaches the level of an expert the rules are no longer on the person's mind and he is completely immersed in the activity.

A novice chess player is a thinker, who determines actions on general rules that are given by a teacher. The chess player learns the value of each piece. A more advanced learner recognises meaningful situations, such as a chess player recognizing a weakened position on the king's side. A proficient player stops reflecting on problematic situations as a detached observer and no longer looks for principles to guide his behaviour. A proficient chess player may know instinctively when to attack, but still has to deliberate about how best to do so. The final stage of learning is that of expertise. An expert knows how to perform an action without calculating and comparing alternatives.

"It seems that beginners make judgements using strict rules and features, but that with talent and a great deal of involved experience the beginner develops into an expert who sees intuitively what to do without applying rules and making judgements at all. [...] an expert does not *solve problems*. He does not *reason*. He does not even act deliberately. Rather he spontaneously does what normally has worked and, naturally, it normally works." (Dreyfus & Dreyfus, 1991: 235)

The expert is involved in the activity in such a way that he has left the conscious mind behind. A chess grand master does not make moves for a reason, in Dreyfus view, they simply act. Any focussed attention on the task at hand merely disrupts the flow and the performance. This stands in contrast to the novice, where focussing attention on the task improves the performance. This means that an ethical novice needs to learn some of the ethics of a community by following strict rules, but once the ethical novice becomes an expert, he would leave rules and principles behind and his ethical responses would be more spontaneous. To illustrate this, imagine yourself coming to a strange land. You find yourself unfamiliar with the local customs and morals. You do not know how to greet other persons, how to haggle at a market, what the societal faux pas are, or the norms governing other social interactions. However, a friend can show you the rules of the market; you can learn rules from an etiquette book, etc. Even when employing these rules you never seem to fully participate in the strange society and all its nuances escape you. Encounters occur clumsily and your social behaviour requires your explicit attention. However, as time goes by you get accustomed to the local customs. The strange land loses its strangeness more and more and becomes familiar. No longer do you need to decipher the actions of the locals, they simply appear as meaningful to you. No longer are you a researcher surrounded by unfamiliar phenomena, but you become a participant; many of phenomena that were strange don't just become familiar but fade to the background of your consciousness. Thus, echoing Aristotle, Dreyfus claims that ethical expert performance is doing what those who already are accepted as ethical experts do and approve. (Dreyfus & Dreyfus, 1991: 237) An ethical expert, just like the chess grand master, would be unable to give you the reasons why he acted.

Although Dreyfus is right in stating that in many cases an expert would be unable to reconstruct the rules or reasons for his action, this does not mean that the action was not rule-governed. Alva Noë argues that the question of whether or not someone acted upon a rule does not depend on whether she (consciously) entertained the rule. Drawing on Wittgenstein's discussion on rule-following³¹, he claims the following:

³¹ See Philosophical Investigations §185-§243

“It is to over-intellectualize the workings of the intellect to suppose that every exercise of understanding requires a deliberate act of contemplation of an explicitly formulated rule. Such an over-intellectualized conception of the intellect leaves out the possibility that intellectual skills themselves may admit of expertise and effortless exercise.”³² (Noë, 2012: 119)

The inability to make a rule explicit does not mean, contrary to what Dreyfus argues, that reason and conceptual understanding are not present. Thus the mistake of cognitivist's traditions, against which Dreyfus argues, is not that skills involve reasons and understanding, but that “they take for granted that the intellect's operation is always deliberate and detached.” (idem: 120) However, if unreflective skills are thoughtful and involve a form of understanding, as Noë argues, then our skills seem to be conceptual or at the least proto-conceptual. If an unreflective sense of the right action is considered to be a reason for us to respond to certain aspects of a situation, then the responsiveness seems to be conceptual, i.e. involving concepts. The problem is that concepts are usually only attributed to judgements and explicit reasoning. In this view, we use concepts to gain knowledge from our (sensory) experience and can, thanks to these concepts, reflect on experience and articulate them. Perceptual consciousness therefore does not rely on concepts since it precedes thinking, but we need those concepts to understand our experience and talk about them. Knowledge involves concepts. Thus gaining knowledge means bringing things under concepts through a deliberative act i.e. a judgement. This seems contrary to the above formulation of ethical know-how as practical unreflective knowledge. An action is normally considered skilful if it consists in the exercise of concepts in the judgemental sense as to mere habits or mere movements. This is apparent in the traditional view of moral agency, where a certain behaviour can only truly be considered an (ethical) action if it is preceded by a thoughtful process of deliberation. However, ethical know-how does not seem to involve concepts or contain propositional knowledge. The problem here is twofold: First, if knowledge or understanding involves the use of linguistic concepts, can know-how be considered knowledge in full sense of its meaning? Second, we can reflect on our ethical practices and can talk about (many) of our ethical practices. So, how does or reflective linguistic ethical understanding relate to our unreflective know-how?

Noë (2015) calls the approach that sees concepts as judgements the *intellectualist* approach. (2-4) This approach stands in contrast to what he calls *crypto-intellectualist* approach, which he attributes to Dreyfus. (idem: 9) The crypto-intellectualist also shares the assumption that concepts are judgements, but only attributes them to the novice. Noë rejects both approaches and instead proposes *concept pluralism* as an alternative. He sees both approaches as over-intellectualising the intellect (see the above given quote). Noë does not share their assumption that the world is already given and everything is present to us. Perception already involves skills and understanding; understanding and skills enable us to bring the world into focus for perceptual consciousness. (Noë, 2012: 115) Thus conceptual understanding means having the skills to gain access to the world and involves a continuous

³² Noë uses the term understanding here to refer to practical knowledge. He does not make a distinction between understanding and knowledge. (Noë, 2012:24) For Noë, know-how, as a form of knowledge, is an achievement of understanding. When it comes to know-how, for instance cycling, having the skill of riding a bike means you understand riding a bike on a practical level. When it comes to know-how I follow Noë for practical purposes not make a distinction between know-how and understanding. If one finds the term understanding for practical knowledge a bridge too far, one might use the term proto-understanding.

engagement with the world. Think for instance of reading; we do not experience the words on a paper as scribbles, but the words appear as meaningful to us when we are skilled in the language in question. We immediately experience the words and sentences as meaningful. Thus we have access to a text in a non-judgemental³³ way. To an illiterate, the text could just as well be invisible. For Noë, then there is not a large difference between thought and perception, for they are different styles of access to the world. The approach Noë proposes is pluralistic in the sense that supposes that there are different modes of access to the world and therefore different modes of understanding, of which judgement is one of them. (Noë, 2015: 11) Human understanding is a continuous working process, but its presence is fragile. The world is not something that is present and readily understood. Through concepts we have to bring the world into presence (or into view). Thus concepts are skills or techniques through which we gain access to the world. However we can fail to bring the world into presence, hence presence and understanding is fragile; our understanding of the world can always be disrupted. But understanding and presence is also robust, a continuous conversation where the world opens up and closes. We can always gain access to the world through skilful engagement. (Noë, 2012: 2)

Noë's notions of understanding, presence and fragility offer a promising way to integrate skills and understanding.³⁴ However, Noë's approach is limited in his account of subjectivity in that he hardly addresses who or what is understanding. According to Thompson (2007), Noë's account is incomplete in that "[k]nowledge implies a knower or agent or self that embodies this knowledge." (206) For Noë understanding is done by the individual (he does not elaborate on who or what this individual is), but our ethical knowledge and understanding is, as I argued, fundamentally intersubjective.³⁵ Ethical meaning is generated through a dynamical and interpersonal interaction between two agents. This means that our ethical understanding is not socially embedded but socially co-constituted. Our ethical understanding is shaped through shared socio-ethical norms, meaning that our ethical understanding is never really the accomplishment of a single individual. This also means that the process of ethical sense-making is a fragile one. Ethical appraisal is dependent on the coordination with others and is not only dependent on what others do, but what we do with others.

Not only does Dreyfus underestimate the role of understanding in skilful activity, his analogy of ethical expertise with general skills like playing chess has limitations from a developmental perspective. We do not learn most of our ethical practices through a process of strict rule-following, like the novice chess player. While the acquisition of chess skills is dependent on reflective processes, our acquisition of ethical skills is for the large part not dependent on a process of explicit rule learning. When we are born we are not strangers in an unfamiliar land, but we are born into a world that is already meaningful. Thus gaining ethical expertise is not so much learning when to apply the correct rule, but to develop skills to engage with the ethical world.

Nigel DeSouza (2013) argues that it is important to make a distinction between *pre-reflective ethical know-how* and *unreflective (ethical) expertise*. (280) Unreflective (ethical)

³³ Not in a narrow understanding of judgements

³⁴ There is of course much more to say about the topic of conceptuality and my discussion grazes not only the mere surface of the debate, but also that of Noë's position and hardly does justice to the debate. However, an extensive discussing of what exactly conceptuality is and the conceptual and nonconceptual nature of experience goes beyond the aim of this thesis.

³⁵ I would argue that all our understanding is interpersonal; however I will limit my account here to ethical understanding.

expertise can be formed through learning rules, like the novice chess player advancing to the level of expertise. Pre-reflective ethical know-how cannot be mediated by processes of reflection since it is logically prior to reflection, meaning that we do not acquire pre-reflective ethical know-how through explicit rule-following. The acquiring of unreflective ethical expertise is dependent on our pre-reflective ethical know-how. Pre-reflective know-how does not mean that the know-how is *reflexive*, in the sense of automated like a reflex and something we are cannot be aware of. This is the mistake Dreyfus seems to make in thinking that awareness can only be reflective. However, in skilful coping we experience our activity in a non-reflective way; experience also comprises a pre-reflective self-awareness such as a pre-reflective bodily self-consciousness.³⁶ (Thompson, 2007: 314-315) Experiences of basic empathy can also be best described as pre-reflective.

I follow DeSouza and take pre-reflective know-how as an ontogenetic condition to be an ethical agent and the pre-reflective know-how is acquired from infancy, meaning that our pre-reflective ethical know-how precedes the reflective normative practices. To repeat the point I made earlier, we are born in a world of (ethical) meaning that is a world permeated with norms, conventions and values. Remember the example of a mother (breast)feeding an infant. Not only does the interaction occur without reflection, but the interaction as an embodied process is also imbued with wider socio-ethical norms and practices, such as what constitutes good parenting. This does not mean that the child has immediate access to all ethical and normative dimensions of the interaction. What it means is that from early childhood on, a person is immediately engaged in practices of (ethical) participatory sense-making. It is important to note that not only our cognition is shaped through participatory sense-making, but also our ethical agency. Just like a mother shapes the attention of her child by making the bottle meaningful, so does Arno draw the attention of the barman on the ethical significance of the lost bicycle pump. In both cases a similar process of participatory sense-making occurs. We experience ethical actions (or the ethical dimension of actions) first in a pre-reflective way, before we develop capacities to gain access to those actions in a reflective and explicit rule-oriented way. In other words, we gain a sense for what is ethical, before we ethically theorise and reflect; we have a pre-reflective sense of what is the correct or right thing to do in a certain situation. DeSouza summarises the point as follow:

“[Pre-reflective ethical know-how forms] the background that is constitutive of our ability later on to make sense of ethical rules and concepts and, relatedly, to think abstractly and critically about ethical and moral questions. The pre-reflective background never just gives way or cedes its place to mature moral reasoning. It is the very foundation on which such reasoning is at all possible.” (DeSouza, 2013: 288)

Critical ethical and moral reflection can only make sense on the basis of this pre-reflective background. Pre-reflective ethical know-how gives us an ‘inner’ sense on which all our ethical skills are build.

The discussion so far seems to suggest that conscious deliberation plays no (significant) role when it comes to ethics. However, as I argued, the process of ethical sense-making is fragile, meaning that the interactions often break down. It is especially in these instances of break-down in sense-making that deliberation is necessary. In these moments of break-down we are no longer experts; we need to analyse and deliberate like Dreyfus’

³⁶ To illustrate this notion, think about touch. When we touch something we not only feel that we touch a thing, but we also feel touched by the thing in question.

novice. A breakdown can occur when the coordination between two agents fail and an ethical interaction is not established. Ethical sense-making is an interpersonal process. However, persons have their own motivations and goals, which can conflict with other persons. A person can also be inattentive to the coordinating movements of other people, by having contradictory norms or simply not being receptive. A breakdown can also occur when the situation is unfamiliar to the agent (or agents) in question. Take again the foreigner in a strange land. The unfamiliarity with the local customs forces the foreigner to consciously think on his actions. Expertise requires a familiarity with the encountered practices.

In other words, we engage in deliberation and reflection when we encounter problems or situations we are unfamiliar with. Note that deliberation and reflection are also processes of participatory sense-making — remember Haidt (2001) arguing that moral reasoning is rarely an individual enterprise. (819) Engaging in reflection and deliberation is a part of our ethical skills, although our ethical sense-making is usually of the immediate coping variety, “which is transparent, stable, and grounded in our personal histories.” (Varela, 1992: 19) A breakdown then, to formulate it in Varela’s terms, is a moment where the (micro)world seizes to make sense and forces detachment; deliberation and explicit reasoning is necessary. A breakdown is always but a moment, since it immediately opens up the opportunity for new meaning hence meaning is never completely lost.

Our (ethical) practices consist of microworlds and microidentities, i.e. they consist of specific meanings and possibilities for action. For instance, the practise of promising comprises a meaningful world of trust and keepings one’s word, while at the same time making an action possible, i.e. keeping a promise (and at the same time limiting the actions of an individual). Practices allow us to appropriately navigate our social environments. Our practices involve a behavioural stance that is a specific repertoire of behavioural patterns, and a web of meaning which makes the specific behavioural stance appropriate. In our normal everyday coping these practices are not in the in the foreground of experience, but also not simply subconscious since they characterise the experience. To reflect means bringing the practice and it’s including meaningful situation to the foreground or centre of consciousness. This motion is in itself an abstracting move in the sense that we take a step back to look at the situation as a whole, we become detached from the immediate concrete situation. However, deliberation and reflection are dependent upon the concrete practises in that they build on a composition of microworlds and microidentities. Deliberation and reflection on a certain situation is important and necessary, until the situation becomes familiar to us, after which the outcome of the deliberation is enriched and becomes a variety of know-how.

4.3 The Ethical World and the Moral Agent

In my discussion of the ethical mind as socially extended, it may seem that we have lost the ethical agent as an individual to the ethical world. The ethical practices we can master are fundamentally interpersonal, leaving seemingly little room for the individual. In order to understand the ethical agent, I must also elaborate on what the *ethical world* is.

In order to come that an ethical appraisal, that is, to enact ethical sense-making, we engage in participatory sense-making. De Jaegher & Di Paolo rightly emphasise that participatory sense-making, like every social interaction, needs to be actively regulated by two or more individuals. This means that, although the interaction is autonomous, there would be no social interaction without two (or more) autonomous interactors, or ethical

agents. However, these interactions are never reducible to the interactors. Thus the ethical (and the social) world is not simply the existence of interaction between individuals. According to Steiner & Stewart (2009) that the social world or the “social domain is always defined by a set of structural norms; and moreover, that these norms are not only a set of constraints, but actually *constitute*, the possibility of enacting worlds that would just not exist without them.” (528) To this I would add that the structural norms are interpersonal practices that are enacted by members of the society or social group. It is the environments of these structures of norm-governed practices that in turn give meaning to all the social interactions. For an ethical agent to produce meaningful behaviour is to produce behaviour which is in line with these interpersonal practices, i.e. patterns of behaviour that are recognized as meaningful and constitute what is meaningful.

To illustrate this, I return to the example of a chess player, because chess is a simple model of a norm-governed practice. The game of chess is constituted by its rules, that is what counts as a move, when do you win etc. Dreyfus showed that being a good chess player involves more than knowing the rules, it requires recognizing strategies and seeing opportunities; becoming good in chess involves developing expertise. Chess playing is an embodied skill that requires know-how not only of the norms but also competencies like visualization, imagination, pattern recognition etc. The more skilful you become at chess, the more you are able to predict and understand what other players are up to in such a way that is impossible for novice players to do. According to Victoria McGeer (2015), there are three lessons to be taken from this example. (262) First, you shape your thoughts and actions in chess-compliant ways. Thus becoming skilful involves self-regulation. Second, competence requires some minimal ability to operate on your own and have a disposition to learn from corrective interventions of others. And third, even as you become more capable to operate on your own, you will still depend on the regulative interventions of others.³⁷ Each chess-player perceives the intentions of other players directly in their playing behaviour. They understand each other because they conform to the shared norms and practices of chess.

If we apply these lessons to ethics, the ethical world can be construed as the whole interpersonal norm-governed ethical practices. Ethics is the domain where humans collectively enact and re-enact meaning through their actions and interactions and in turn coordinate their individual and collective behaviour in the social world; ethics serves as the fabric of the social world. Thus the ethical world is the whole of autonomous interactions between ethical agents. The ethical world is a web of meaning and significance that evolved in a particular community (one does not have to belong to a single community) and underlies all social interaction, by both limiting action and making certain actions meaningful.

The socio-ethical world does not only play a contextual and developmental role, but also a constitutive one. Our mind is (partially) shaped through processes of participatory sense-making. This means that the self, which is the individual as a whole, consist of organised interpersonal norms and practices. This human self-organisation is achieved through the continuous interaction with others; we are socialised. Our cognitive development is dependent on the external practices; hence human cognition is social cognition. These ethical practices have a history of being enacted by other individuals and therefore precede the individuals in question. We are shaped as subjects through our interactions with other people. Therefore, it can be said that it is the social-ethical world that enables humans to be autonomous.

³⁷ Note that McGeer speaks in terms of capacities, as in capacities needed to be considered a moral agent.

Until now I have used the enactivist definition of autonomy, i.e. a system that self-organises its identity based upon operational closure.³⁸ However, this definition fails to adequately describe human autonomy in the ethical world, as this autonomy also comes from our interactions with other people. This means that autonomy cannot only be constituted by processes at the level of the body, which autopoietic enactivism proposes. This is most likely a point of contention between me and the autopoietic enactivism as described in the previous chapters. According to autopoietic enactivism, the metabolic processes precede all other forms of organisation. Our autonomy is derived from the operational closure of our metabolic processes and our normative relation with the world primarily depends on our bodily processes.³⁹ For the autopoietic enactivists as discussed in the previous chapters, autonomy is constituted by bodily processes, i.e., the metabolic processes. However, I do not believe you can speak of persons and/or autonomy in any meaningful sense without the social processes a person is embedded in. What I mean is that you cannot give primacy of metabolic processes over social interactions or vice versa. I argue that many of the social interactions humans are embedded are also constitutive for autonomy. To imagine a human derived of social processes is to leave a human only with basic instincts and death. The autonomous mind (or better said the autonomous body) is co-constituted between the individual and the social practices.

The identity of the self is always a relation between the individual and the socio-ethical world mediated through the body. The self relies on interactions with the others. Thus the different (ethical) meanings and practices generated in the socio-ethical world form your individual mind, but at the same time they are your practices and form your understanding as you enact them as a person. We self-organise through our skilful engagement with others. Both the self and the ethical world are co-constituted. This co-constitution means that both the individual and the world have a say in ethical sense-making. We cannot gain access to the ethical world without the interaction with others. What we care for is shaped by our history of interactions with other agents.

This access is fragile and therefore the ethical world is fragile, since the ethical world has to be enacted by its skilful participants. The ethical world is above all a joint enterprise and there will always be limitations and failures to achieve ethical sense-making. Coordination with other people might fail or we may know ourselves less than we think. However, the ethical world is also robust in that we always engage in participatory sense-making; we are always interacting with others. We are shaped by others and in turn shape others through our ethical interactions.

³⁸ See chapter 3.1.3 for the discussion on autonomy from the enactive approach

³⁹ See section 3.1.3.

5. Conclusions

In this thesis, I have shown that it is worthwhile to think about ethics and moral agency with the aid of empirical research. Useful conceptual tools for thinking about moral agency can be found in the enactivist approach. I have shown that certain assumptions of the standard view of moral agency, concerning to the relationship between emotion and cognition, and controlled and automatic processes, need not be shared. I have argued that an alternative view on moral agency is possible, which maintains a different relation between emotion and rationality in the moral domain. I have attempted to show this by sketching an embodied, situated account of ethical agency, with insights from the enactivist approach. I also argued that ethical life cannot be conceptualised as making isolated quick singular appraisals, but ethics is implicit in everything we do. Our moral appraisals are strongly interconnected with other people and fundamentally intersubjective; our moral cognition is socially extended. There is an ongoing moral exchange with other people and a continuous adjustment of moral attitudes. I agree with Colombetti and Torrance that ethical sense-making should be conceptualised as a form of participatory sense-making. In other words, humans, as living organisms, continuously assign meaning to their environments. This means that humans continuously enact ethical meaning. I proposed to call the whole of ethical sense-making practises 'the ethical world'. There is no clear sign when a person starts or stops engaging with the ethical world. Humans constantly move in and respond to what is around them. Each action not only affirms or disaffirms certain ethical practices and norms, but these practices and norms are also enacted in the actions performed. This process of enacting ethical meaning and affirming or changing ethical practices fundamentally intersubjective and must therefore be analysed from a relational perspective. Furthermore, I proposed that the ethical world in turn structures our social interactions. I have explained how we can understand ethics as a matter of skilful interacting with others and the knowledge employed in these activities is best described as ethical know-how. Ethics is an effective way of regulating each other's behaviour, through a process of continuous negotiation. The individual shapes the ethical world just as much as the individual is shaped by the ethical world.

This thesis touched upon many subjects concerning moral agency. Moral agency understood as the application of ethical know-how has many implications that warrant further discussion. First, when it comes to studying ethics and ethical behaviour we have to put the interaction central. This also has consequences for answering normative questions. For instance, merely assigning blame to individuals is not enough, we also have to look to the practices involved which promote specific morally bad or good behaviour. Moral responsibility for moral interaction is distributed throughout society (although this distribution is not equal). If a moral appraisal is co-constituted, that implies that all of the involved participants have a co-ownership of the ethical appraisal and the moral actions. That is what I mean with that ethics and morality is a joint enterprise.

Second, the question remains how we should think about the relationship between our ethical know-how and our ethical know-that. In other words, what is the nature of moral deliberation? Moral improvement seems to be first of all a matter of self-knowledge. We know ourselves through our skilful habits, i.e. our ethical know-how. However, there can be a mismatch between our habits and our moral beliefs. The way we act may not be in accordance to how we see ourselves as a moral person. Furthermore, ethics is a matter of continuous practice. Skills not only have to be acquired but also maintained. Our ethical skills are not simply capacities we acquire in childhood. Throughout our lives we have to make an

effortful attempt to practice and improve those skills. This task is impossible to achieve on one's own, but requires direct interaction with other moral agents.

Another consequence worth mentioning is that the notion of ethical know-how as interpersonal has implications for the notion of autonomy. Autonomy according to the (mainstream) enactivists has to be understood as the ability to self-organise through which a bodily self and a meaningful environment emerge. Autonomy requires a distinction between bodily processes that achieve operational closure and a distinct environment. However, I argued that although autonomy involves the self-organisation of a bodily self, in the case of humans it also necessarily involves the shared socio-ethical practice. Thus autonomy is better understood as the self-organisation within a social world. The emergence of an ethical self can only come about through the continuous engagement with others, meaning that the social interactions themselves are part of the organisation of the ethical agent. Thus the (ethical) self is not only the result of its bodily organisation, but is also co-constituted by the social practices and interactions it is embedded in.

Finally, I want to point out that in this thesis I mainly focussed on micro-interactions, i.e. direct interactions between one or more individuals, and the processes involved in those interactions. However, ethics concerns all human interactions and therefore any account of moral agency must also include the impact of macro-interactions, i.e. societal institutions, on a moral agent. Thus ethics is not only a matter of how we directly interact (or should interact) with others, but also how societal institutions shape us as moral agents and how we collectively form these institutions.

The notion of ethical know-how, as proposed by Varela and Dreyfuss, provides a promising way to rethink moral agency. However, some questions remain and other things have to be reconceptualised. Thus, this thesis must be understood as outlining an examination into the notion of moral agency. To this project, I hope to have contributed.

References

- Anderson, S.W., Bechara, A., Damasio, H., Tranel, D., Damasio, A.R. (1999) Impairment of social and moral behaviour related to early damage in human prefrontal cortex. *Nature Neuroscience*, vol. 2, no 11, 1032-1037
- Barrett, L.F. (2013) Psychological Construction: The Darwinian Approach to the Science of Emotion. *Emotion Review*, Vol. 5, No. 4, 379-389
- Barrett, L.F. & Bar, M. (2009) See it with feeling: affective predictions during object perception. *Phil. Trans. R. Soc. B*, 364, 1325-1334
- Barrett, L.F. & Tugade, M.M., Engle, R.W. (2004) Individual Differences in Working Memory Capacity and Dual-Process Theories of the Mind. *Psychological Bulletin*, Vol. 130, No. 4, 553-573
- Becharra, A. & Damasio, A.R. (2005) The somatic marker hypothesis: a neural theory of economic decision. *Games and Economic Behavior* 52, 336-372
- Bennett, J. (1974) The Conscience of Huckleberry Finn. *Philosophy* 49, 123–134
- Blair, R.J.R. (1995) A cognitive developmental approach to morality: investigating the psychopath. *Cognition* 57, 1-29
- Clark, A. & Chalmers, D. (1998) The Extended Mind. *Analysis*, Vol. 58, No. 1, 7-19
- Clark, T.W. (2013) Experience and Autonomy: Why Consciousness Does and Doesn't Matter. In: *Exploring the Illusion of Free Will and Moral Responsibility*. Ed: Caruso G.D., Plymouth: Lexington Books.
- Colombetti, G. (2014) *The feeling body: affective science meets the enactive mind*. Cambridge, MIT Press
- Colombetti, G. & Thompson, E. (2007) The Feeling Body: Towards an Enactive Approach to Emotion. In: *Body in Mind, Mind in Body: Developmental Perspectives on Embodiment and Consciousness*. Eds: Overton, W. & Mueller, U., Lawrence Erlbaum
- Colombetti, G. & Torrance, S. (2009) Emotion and Ethics: An inter-(en)active approach. *Phenomenology and the Cognitive Sciences*, 8, 505-526
- Damasio, A.R. (2004) William James and the modern neurobiology of emotion. In: *Emotion, Evolution, Rationality*. Ed.: Evans, D. & Cruse, P., Oxford: Oxford University Press.
- Damasio, A.R., Tranel, D., Damasio H. (1990) Individuals with sociopathic behaviour caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, 41, 81-94
- De Jaegher, H. & Di Paolo, E. (2007) Participatory Sense-Making: An Enactive Approach to Social Cognition. *Phenomenology and the Cognitive Sciences*, 6, 485-507
- DeSouza, N. (2013) Pre-Reflective Ethical Know-How. *Ethical Theory and Moral Practice* 16, 279-294

- Dewey, J. (1922) *Human Nature and Conduct: An Introduction to Social Psychology*. London: G. Allen & Unwin [2007].
- Di Paolo, E. & Thompson, E. (2014) The Enactive Approach. In: *The Routledge Handbook of Embodied Cognition*. Ed.: Shapiro, L., New York: Routledge
- Dreyfus, H.L. & Dreyfus S.E. (1991) Towards a phenomenology of ethical expertise. *Human Studies*, Vol. 14, 229-250
- Evans, J.St.B.T. (2008) Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, 59, 255-278
- Foot, P. (1967) The Problem of Abortion and the Doctrine of the Double Effect. *Oxford review*, no. 5
- Frankish, K. (2010) Dual-Process and Dual-System Theories of Reasoning. *Philosophy Compass*, 5/10, 914-926
- Gallagher, S. (2013) The socially extended mind. *Cognitive Systems Research*, Vol. 25-26, 4-12
- Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*. Boston: Houghten Mifflin
- Greene, J.D., Sommerville, B.R., Nystrom, L.E., Darley, J.M., Cohen, J.D. (2001) An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, vol. 293, 2105-2108
- Greene J.D., Cushman, F.A., Stewart, L.E., Lowenberg, K., Nystrom, L.E., Cohen, J.D. (2009) Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition* 111, 364-371
- Greene, J.D. (2015) The Cognitive Neuroscience of Moral Judgment and Decision Making. In: *The Moral Brain: A Multidisciplinary Perspective*. Ed.: Decety, J. & Wheatley T., Cambridge: MIT Press
- Grey, N.S., Macculloch, M.J., Smith, J., Morris, M. & Snowden, R.J. (2003) Violence viewed by psychopathic murderers. *Nature*, 423, 497-498
- Gollwitzer, P.M. (1999) Implementation Intentions. Strong Effects of Simple Plans. *American Psychologist*, Vol. 54, No. 7, 493-503
- Gollwitzer, P.M., Sheeran, P., Michalski, V., Seifert, A.E. (2009) When Intentions Go Public. Does Social Reality Widen the Intention-Behavior Gap? *Psychological Science*, Vol. 20, No. 5, 612-618
- Govorun, O. & Payne, B.K. (2006) Ego-depletion and Prejudice: Separating Automatic and Controlled Components. *Social Cognition*, Vol. 24, No. 2, 111-136
- Haidt, J. (2001) The Emotional Dog and Its Rational Tail. *Psychological Review*, vol. 108, no. 4, 814-834
- Joyce, R. (2009) Review of Jesse Prinz's the Emotional Construction of Morals. *Mind*, 118, 508-518

- Kaye, K. (1982) *The Mental and Social Life of Babies: How Parents Create Persons*. Chicago: The University of Chicago Press
- Kenneth, J. (2006) Do psychopaths really threaten moral rationalism? *Philosophical Explorations*, Vol. 9, No. 1, 69-82
- Kenneth, J. & Fine, C. (2008) Will the Real Moral Judgment Please Stand Up? *Ethical Theory and Moral Practice* 12 (1):77–96
- Krueger, J. (2013) Ontogenesis of the socially extended mind. *Cognitive Systems Research*, Vol. 25-26, 40-46
- Langdon, R. & Delmas, K. (2012) Moral Reasoning and Psychopathic Tendencies in the General Community. In: *Emotions, Imagination, and Moral Reasoning*. Ed: Langdon, R. & Mackenzie C., New York: Psychology Press
- Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F. The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, 35, 121-202
- Merleau-Ponty, M. (1945) *Phenomenology of Perception (Phénoménologie de la Perception)*. Transl.: Smith, C. London: Routledge Classics [2005]
- McGeer, V. (2015) Mind-making practices: the social infrastructure of self-knowing agency and responsibility. *Philosophical Exploration* Vol. 18, No. 2, 259-281
- Mischel, W. (1981) Metacognition and the rules of delay. In: *Social cognitive development*. Ed.: J.H. Flavell & I. Ross, Cambridge: Cambridge University Press
- Murphy, S., Haidt, J. & Björklund, F. (2000) Moral dumbfounding: When intuition finds no reason. Unpublished Manuscript, Department of Philosophy, University of Virginia
- Nucci, L. (1981) Conceptions of Personal Issues: A Domain Distinct from Moral or Societal Concepts. *Child Development*, Vol. 52, No. 1, 114-121
- Noë, A. (2012) *Varieties of presence*. Cambridge: Harvard University Press
- Noë, A. (2015) Concept Pluralism, Direct Perception and the Fragility of Presence. In: *Open Mind*. Eds.: Metzinger, T. & Windt, J.M., Frankfurt a. M.: GER: Mind Group.
- O'Regan, J.K. & Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behavioural and Brain Sciences*, Vol. 24, 939-1031
- Pessoa, L. (2008) On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, vol. 9, 148-158
- Pettit, P. (2007) Neuroscience and Agent-Control. In: *Distributed Cognition and the Will: Individual Volition and Social Context*. Ed.: Ross, D., Spurrett, D., Kincaid, H., Stephens, G.L., Cambridge: MIT Press.
- Pizarro, D.A. & Bloom, P. (2003) The Intelligence of the Moral Intuitions: Comment on Haidt (2001). *Psychological Review*, Vol. 110, No. 1, 193-196

- Prinz, J.J. (2007) *The Emotional Construction of Morals*. Oxford: Oxford University Press [2013]
- Ryle, G. (1945-1946) Knowing How and Knowing That: The Presidential Address. *Proceedings of the Aristotelian Society, New Series, Vol. 46, 1-16*
- Sauer, H. (2011) Psychopaths and Filthy Desks. *Ethical Theory and Moral Practice 15, 95-115*
- Sauer, H. (2013) *Educated Intuitions: A Rationalist Theory of Moral Judgement*. Dissertation, Groningen.
- Schwartz, N. & Clore, G.L. (1983) Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States. *Journal of Personality and Social Psychology, Vol. 45, No. 3, 513-523*
- Sie, M. (2009) Moral Agency, Conscious Control, and Deliberative Awareness. *Inquiry, Vol. 52, No. 5, 516-531*
- Sie, M. (2013) Free will an illusion? an answer from a pragmatic sentimentalist point of view. In: *Exploring the illusions of free will and moral responsibility*. Ed: Caruso, G.D., Plymouth: Lexington Books
- Spinoza, B. de (1677) *The Ethics. (Ethica Ordine Geometrico Demonstrata)*. Transl. Elwes, R.H.M., Project Gutenberg [2003]
Retrieved from: <http://www.gutenberg.org/files/3800/3800-h/3800-h.htm#chap03>
- Stein, E. (1917) *On the Problem of Empathy (Zum Problem der Einfühlung)*. Transl. Stein, W., Washington: ICS Publications [1989]
- Steiner, P. & Stewart, J. (2009) From autonomy to heteronomy (and back): The enaction of social life. *Phenomenology and Cognitive Sciences 8, 527-550*
- Thompson, E. (2007) *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, Harvard University Press
- Thompson, E. & Stapleton, M. (2009) Making Sense of Sense-Making: Reflections on Enactive and Extended Mind Theories. *Topoi, Vol. 28, 23-30*
- Thompson, J. (1985) The Trolley Problem. *The Yale Law Journal, Vol. 94, No. 6, 1395-1415*
- Torrance, S. (2006) In search of the enactive: Introduction to special issue on enactive experience. *Phenomenology and the Cognitive Sciences 4, 357-368*
- Varela, F.J., Thompson, E., Rosh, E. (1991) *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MIT Press [1993]
- Varela, F.J. (1992) *Ethical Know-How: Action, Wisdom, and Cognition (Un Know-How per l'Ethica)*. Transl.: Edizone Laterza, Stanford, Stanford University Press, [1999]
- Vujošević, M. (2014) Conscience as the rational deficit of psychopaths. *Philosophical Psychology, DOI: 10.1080/09515089.2014.983221*

Wittgenstein, L. (1953) *Philosophical Investigations*. Transl.: Anscombe, G.E.M., Oxford: Basil Blackwell [1972]