# Master's Thesis

*About an interdisciplinary approach to morality, and what it adds to the toolbox of the applied ethicist*

Universiteit Utrecht

*Nina Marloes van den Berg*

December 2019

Master's thesis

*About an interdisciplinary approach to morality,*

*and what it adds to the toolbox of the applied ethicist*

January 2020

Applied Ethics

Faculty of Humanities

Graduate School of Humanities

University of Utrecht

Nina Marloes van den Berg

Student number: 6216668

Supervisor: dr. Frans Stafleu

Second reader: dr. Hanno Sauer

*“The very beginning of Genesis tells us that God created man in order to give him dominion over fish and fowl and all creatures. Of course, Genesis was written by a man, not a horse”*

- **Milan Kundera, The unbearable lightness of being (1984)**

# Table of contents

# Introduction

## *Preface*

In March 2019 the Dutch government funded institute for social research, het Sociaal en Cultureel planbureau (SCP) launched a quarterly rapport called *burgerperspectieven*, 'Citizens' perspectives', in which the result of a continuous research project reflecting the opinion of the Dutch people on multiple matters is presented. By means of surveys and focus groups, the institute follows the developments taking place in society and the attitudes that shape the discourse on public matters in the Netherlands. The report presented in March 2019 goes into detail about the different opinions on matters like refugee policies, climate policies and forms of institutionalized racism that are reflected in Dutch culture. The SCP questioned whether these differences in opinion create social oppositions, how those oppositions are experienced by civilians and whether they feel as if this amounts to increased conflicts and tensions within society.

The SCP found that the vast majority of the Dutch interviewees (75%) felt as though the general sphere in the public debate has hardened over the years, that polarisation is growing, and that it becomes increasingly more important to 'pick a side', pointing to the incommensurability of opinions that divide the debates. These changes are attributed to the influence of social media, a hardened, overall mentality and the growing multicultural character of Dutch society (Dekker, Den Ridder, & SCP, 2019. p.39).

From a meta perspective, the applied ethicist can approach such polarized public discussions for example by exercising reason and engaging in philosophical, unnatural modes of thought to evaluate the logical verifiability of the arguments on both sides of the discussion. They could reflect on the available facts weighing matters of harm, rights and justice, and construct arguments that ideally reach a state of universal applicability. Or the ethicist could try to reach a reflective equilibrium among various beliefs or argue for different positions based on how well those positions would fit in Kant's Kingdom of Ends (1998/1785, 4:439) under Rawls' veil of ignorance (1971, p. 136) or as an outcome of Bentham's utilitarian calculus (1999/1789, p.31-34). Engaging in philosophical exercises of this kind is a reflection of a traditional, top-down approach to morality and ethics that has been popular in the sciences and humanities since the Enlightenment.

This approach, that draws heavily on assumptions of human exceptionalism such as the unique human ability to reason, takes a central position in the MA applied ethics at the University of Utrecht, as most of the theories that are cultivated in the courses are rooted in this tradition.

In this thesis however I will argue for a different approach to morality and ethics, an approach that flows naturally from interdisciplinary research on morality and can, in contrast with the former perspective, be called 'bottom-up'.
By incorporating research results from the fields of primatology, evolutionary biology and moral psychology I aim to offer a layered understanding of the phenomenon of human morality in this thesis. I will illustrate how empirical research on the *arche* of morality, the very roots of it, can help us to better understand what function morality serves and how it serves that function.
I will argue that, by applying knowledge from different scientific fields that treat the topic of morality, the applied ethicist is able to get a more complete understanding of what has shaped and continues to shape human morality, which in turn will be helpful in a general sense to better understand, evaluate, and guide ethical debates. As this capacity should be of specific interest to the applied ethicist, I will conclude this thesis by arguing that a course on interdisciplinary approaches to morality and non-western ethical traditions should take a more prominent position in the central curriculum of the MA applied ethics at the university of Utrecht.

## *Method*

The central question to my research project is: *What does an interdisciplinary, bottom-up approach to morality attribute to the toolbox of the moral philosopher?*
I will make use of Frans de Waal's work on the evolutionary roots of morality and Jonathan Haidt's work on the automatic, cognitive processes involved in moral judgment to give an overview of what such an interdisciplinary approach comprehends.
I will illustrate what their work attributes to our current understanding of human morality, how it challenges some of the implicit assumptions underlying most prominent and influential theories in the Western philosophical tradition, and how they invite an alternative perspective to morality which can be conducive to the practical philosopher.
This alternative perspective, which will be called the 'bottom-up approach' is contrasted with a traditional 'top-down approach'. The motivation to differentiate between both perspectives, in which the former attains to a view of morality as process that comes from within and the latter as a capacity that is installed from the outside, is motivated by the disciplinary perspectives of evolutionary biology and primatology.
As evolution rarely throws anything out (de Waal, 2006, p.21) the traits or capacities that underlie our present behaviour have a core, and that core or rudimental form shows continuity with certain capacities found in other species. Morality is seen as a phenomenon that flows naturally from the social tendencies of primates, as opposed to morality being viewed upon as an independent,

uniquely human construct that was there all of a sudden (De Waal, 2018, p.249). Rather, it is the product of a continuously developing, cognitive layers-adding process in hominid evolution.

In all disciplines that are treated in this thesis there are similarities found amongst the theories that view upon morality as a phenomenon coming from within and theories that view upon morality as a phenomenon that is installed from the outside:

| **Biology** | **Psychology** | **Philosophy** |
|---|---|---|
| Social nature/Veneer theory | Intuitionism/Rationalism | Virtue ethics/Quandary ethics |

The main differences between both approaches in the sciences are that a bottom-up approach to morality is characterized by (1) a focus on *evolution*; it examines the evolutionary ancient mechanisms and advantages these gave to our ancestors to explore the notion of human morality, and (2) a focus on the *social nature* of our species and the continuity in capacities with the species that belong to the same taxonomic order as us. Because life in groups offers major selective advantages to primates but at the same time obstructs their individual goals, morality is seen as natural social glue serving our incentives to live together and cooperate with others.
On the contrary, a top-down approach to morality departs from (1) the unique human ability to *reason* to explain morality. The power of the mind to engage in conscious mental processes and derive logically valid judgements is said to motivate our moral judgements.
(2) the *individual*, autonomous and rational nature of humans is emphasised, both in what enables moral deliberation and in what morality ought to protect.

With the tendency to *"…put our most advanced traits on the pedestal, ignoring or even denying simpler antecedents*" (de Waal, 2006, p.23), the disciplines of moral psychology and philosophical ethics have favoured a top-down, reason-infused explanation of morality (Haidt, 2001; Horgan & Timmons, 2007) over evolution based explanations. In philosophical ethics this is expressed in the implicit assumption underlying most prominent theories that we, as a human species, have been endowed with the ability to be moral and are therefore raised infinitely above the animal kingdom, either through the hands of God (Genesis, 3:22) or because the ability to reason places us on top of the natural order of things (Aristotle, trans. 1995, p.118).

This made way for the longstanding tradition in philosophy to distinguish humans from animals and attributing additional value to the former, as is reflected in the following quote by Immanuel Kant (1996/1798):

'*The fact that the human being can have the representation "I" raises him infinitely above all the other beings on earth. By this he is a person… that is, a being altogether different in rank and dignity from things, such as irrational animals, with which one may deal and dispose at one's discretion'*

Characteristics that are said to constitute human uniqueness, such as personhood, empathic perspective taking, advanced language systems and the ability to reason, gave rise to the ambition of prominent Western philosophers during the Enlightenment to ground the foundations for ethics in that what sets us apart from the natural world: the human ability to rise above mere wantonness and use the faculty of reason to arrive at logically valid conclusions and morally bounding principles. The immense influence of the work of Immanuel Kant on Western philosophy led to the consensus of quandary ethics in moral psychology, a way of looking at morality as a means to solve dilemma's between free and equal agents, based on principles of justice and harm (Haidt, 2006, p.3). The perspective on human nature that underlies this is that morally mature humans are individual, rational agents capable of arriving at correct moral conclusions derived from abstract moral principles.

But in our everyday moral reasoning, do we really have the ability to bound behaviour based on these cognitive insights and rule-based outcomes? In other words, can reason alone motivate us to act or not to act? Is morality about problem solving only?

The influential work of Laurence Kohlberg (1971) in moral psychology implicitly confirmed this last question. Kohlberg, like many other psychologists during the cognitive revolution, believed strongly in the power of cognitive abilities and used it to explain progress in moral maturation during development. He believed that we are born with the empathic ability to engage in the process of role-taking, and that a child progressively gets better at solving moral dilemmas by applying conventional and eventually postconventional moral concepts to problems. He had a strong belief that the outcomes of rational processes, 'good' reasons, also motivated moral judgements. Kohlberg inherited these assumptions about morality, that it consists of applying rules, solving dilemmas and emphasising moral values like impartiality and abstraction, from moral philosophy (Haidt & Joseph, 2006, p.6).

This way of looking at morality was criticized in the 1980ties from within the discipline by Kohlberg's colleague and student Carol Gilligan, a Harvard psychologist whose publication *In a different voice*

(1982) is seen by many as the official starting point of the feminist turn in philosophical ethics (Shafer-Landau, 2012, p.274).

During her research on the difference between men and women's perspectives on morality, Gilligan noticed that many of her female research subjects fared poorly on Kohlberg's scale. Women almost never advanced beyond the third stage of development, the stage in which morality is strongly embedded in and motivated by the social domain. Gilligan found that women more often brought an attitude of care and sympathy to their moral reasoning and rarely used abstract moral principles or rules. Women emphasised values such as compassion, flexibility, cooperation, care, humility and mutual respect and continuously favoured these traits over the postconventional moral values of abstraction, impartiality and universalizability.

Gilligan argued that the female ethical perspective was not inferior because it could not reach the final stage of Kohlbergian moral maturation, but rather that Kohlberg's model was infused by well-entrenched prejudices of top-down, male dominated perspectives to morality, and that he had failed to notice and incorporate 'female'[1] values.

Within the discipline of philosophy, the publication of Gilligan's work gave rise to the school of *feminist ethics*, also known as the *ethics of care*, which is an approach to morality encompassing multiple viewpoints rather than it being one single, unified ethical theory. The main characteristic these viewpoints share is that they emphasise 'the female perspective[2]' and aim to understand and correct the gender binary that underlies current moral beliefs and the methodological approaches to ethical theory (Norlock, 2019, p. 1). They prioritize care and the importance of emotions in their theories and criticize the predominance of traditional values such as unification, impartiality, abstraction and competition. Rather than defining moral maturity as the ability to understand and apply these traditional values as Kohlberg did, feminist/care ethicists would rather argue that moral maturity is gained by "*facing life's difficult choices and not pretending that overly simple answers will solve our problems*" (Shafer-Landau, 2012, p. 281). The viewpoints of care ethical approaches fit well within a bottom-up perspective to morality as they emphasise the importance of sociality, care and empathy.

However, as the top-down approach to morality has prevailed for such a long time in the disciplines of philosophy and psychology, celebrating human rationality in multiple ways for its capacity to reach

[1] Brackets are added to denote that here I'm discussing values that are *traditionally* attributed to women. It does not mean that by nature, all women always use these values in their reasoning, nor does it mean that men never apply these values.

[2] Idem.

dazzling heights, the bottom-up perspectives proposed by Haidt and de Waal raise questions about the diminished value that is attributed to reason. Contra bottom-up theories, critics argue that deliberate reasoning remains the primary cause of moral judgements, for example because modern humans face moral problems that could never have been anticipated by evolution (Pizarro & Bloom, 2003, p.195)

After illustrating the main points of critique to a bottom-up approach to morality, I will argue that there is an alternative role attributed to reasoning in these approaches, and that this role alters the locus, but not the importance of reasoning. The bottom-up perspectives leave room for the intuitions and reasons to merge in moral judgement. By solely focussing on reason as the cause, the moral philosopher limits her scope of the full range that constitutes human morality.

I hope to show that moral philosophy can use the expanded knowledge of bottom-up perspectives not only as a means to redefine and put into context the theories of its own discipline, but also as an instrument to better analyse, reflect on and respond to contemporary ethical debates.

# Chapter one: The evolutionary roots of morality

*If the origins of morality are not biological, then what is the source of strength that enabled humanity to escape from its own nature and implement moral systems?* – Flack & de Waal (2000)

In this chapter I will use the work of primatologist Frans de Waal to explore two notions that, from an evolutionary biologist perspective, are considered the pillars of morality. Both *empathy* and *reciprocity* are seen as the necessary building blocks of human morality, and the presence of these capacities in our closest relatives suggests that these building blocks of morality are rooted in and shaped by evolution. Thus, in this line of argument, biology has to a certain degree influenced the development of human moral systems. The idea that morality stems from evolution instead of other, more top-down sources such as religion or the human ability to reason, was contested in the sciences and humanities for a long time. Instead of being rooted in biology, morality was seen as a way to supress an anti-social, aggressive nature which sole aim was self-perseverance. Morality functioned as the set of rules that restrained us from 'doing what we want', a point of view reflected in Thomas Huxley's *veneer theory*: the idea that morality is a thin layer of veneer that protects us from an essentially evil nature. In philosophy, thinkers such as Thomas Hobbes built their political theories based on this view. Hobbes, who wrote his magnum opus the *Leviathan* during the civil war in England, described life in the state of nature as 'solitary, poor, nasty, brutish and short'. His social contract theory relies on the idea that no one, in their right mind, would want to return to this brutal state in which man is a wolf to man and life is marked by a war of all against all.

But when Frans de Waal, at the beginning of his academical career, studied a chimpanzee colony in the Arnhem Zoo, he found that chimpanzees reconcile after fights. But why would chimpanzees reconcile after fights when they were essentially anti-social, aggressive, and programmed to be selfish? These questions where the starting point of years of extensive research into manifestations of prosocial behaviour in animals focussing on the capacities that underlie such behaviour.

Every organism strives to achieve certain goals or outcomes in life. Amongst them are survival and reproduction, and social animals are for a large part depended on groups for the maintenance of these goals. The chances to survival and reproduction depend on one's ability to show prosocial behaviour and fare well within the group, because both are needed to reap the benefits of cooperation (foraging advantages, protection from predators) and to be prepared for competition (dominance strategies, conflict resolution). De Waal (2006) argues that therefore, it is in the social domain that the highest cognitive achievements in social animals are expected. These achievements

are assisted by mechanisms favoured by evolution that make it possible to quickly detect and react to the emotional state of others and to measure whether one actually does reap the benefits of group membership.

In order for social animals to live and thrive together they need to have the capacity for *reciprocity* and a sense of *other-orientation*: the former requires a system that compares investments with payoffs and the latter has a cognitive and emotional component and is known as empathy.

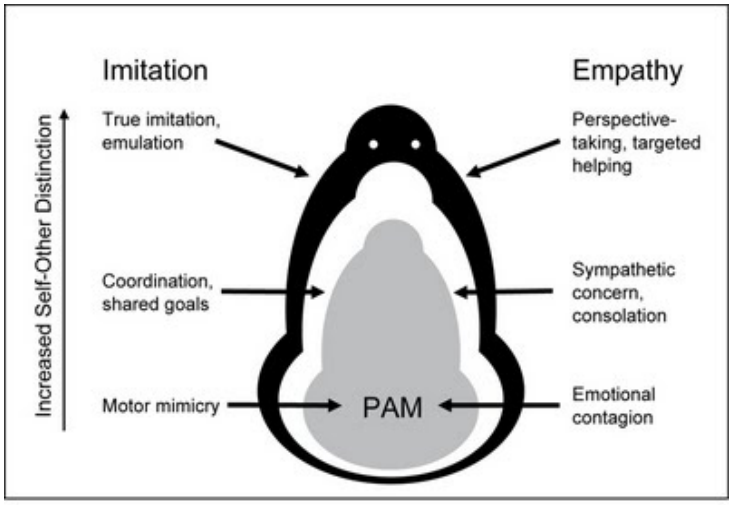
## *Empathy*

Instead of looking at the capacity for empathy as an 'all-or-nothing' concept, the biologist prefers a Russian doll model perspective in which evolutionary 'old' hardwired capacities remain present in 'new' cognitively advanced capacities (de Waal, 2006, p.21).

In the broadest, bottom-up definition of empathy it is referred to as the sensitivity to the emotions of others (de Waal, 2008). The capacity for empathy enables an individual to quickly relate to the state of the other, which is an essential ingredient for social regulation that most likely came into existence in the context of parental care (de Waal, 2008). Being receptive to the signaling of needs of infants is necessary for all species who's offspring relies on parental care in the early stages of life. Parents that had the capacity to quickly detect and react to the emotional state of their young signaling the need for food, warmth, or protection, had a selective advantage over parents who were indifferent to such signaling. This resulted in a rapid evolution of the ability for emotional connectedness, which later was applied to social relations outside of parent/child bonds. This very broad definition of empathy can be refined by adding cognitive capacities that create a more layered understanding of empathy that is shared across species.

### *The Russian doll model of empathy and imitation*

Frans de Waal argued for such a layered understanding of empathy when he introduced *the Russian doll model of empathy and imitation* (2008). This model explains empathy as a capacity that has a hard-wired basis and several layers that build upon it which require higher cognitive capacities limited to a few species, including the taxonomical group both primates and humans




de Waal FBM. 2008
Annu. Rev. Psychol. 59:279–300

belong to. This taxonomical group, *the Hominidae*, consists of four species including three types of orang-oetangs, two types of gorilla's, the bonobo, the chimpanzee and homo sapiens.

At the core of this model lies the ability to adopt the emotional state of the other. This adaption, either in whole or in part, is called (1) *emotional contagion*. As this is widely recognized in the human species, and as it evolved so rapidly, evolutionary biologists argue that it would be very strange if emotional contagion showed no continuity with other species (de Waal, 2008).
Emotional contagion is in fact recognizable in numerous other species and relies on a mechanism that provides the subject access to the objects' emotional state through motor mimicry. This so-called perception action mechanism (PAM) is an automatic reaction that mimics the state of the object through the subject's own neural and bodily representations, for example through mirror neurons (Preston & de Waal, 2002). By perceiving the state of the object, a similar state is automatically and subconsciously activated in the subject, who then physically shares the emotional state of the object. Farmers in the Netherlands effectively put this mechanism to use in defeating rat plagues: the agonized screams of a rat that is burned alive induces such terror in other rats that it is said to scare them off for quite some time.
This motor mimicry is an involuntary process that lies at the core of empathic processes. This basic relation to the emotions of others is established subconsciously. For example, in research conducted by Dimberg et all (2011), participants were shown photos of human faces expressing different emotions. Without being aware of it, all subjects showed facial muscle activity that simulated the expression of the emotional state shown in the picture.
In contrast, the use of facial botox treatment blocks all these forms of miniscule muscular activity, resulting in a diminished capacity for emotional contagion and the inducement of empathy in others (Prochazkova & Kret, 2017).
The very basic form of empathy as a subconscious, emotional experience, enables one to quickly and automatically relate to the state of others. It is a capacity that enables altruistic behavior which is defined in biology as "*behavior that increases the recipient's fitness at a cost for the performer*" (de Waal, 2008). This system is self-rewarding like other systems in nature that are needed for basic survival such as eating, sex and nurturing. Because emotional contagion activates a similar state in the object, the object searches to reduce the distress in the subject because of the negative vicarious emotional arousal that is activated in herself. Trying to relieve the distress in the subject reduces the object's own distress.

The second evolutionary layer in the model adds (2) *sympathetic concern* to the core process of emotional state matching. Sympathetic concern means to have a concern for others and a contextual understanding of what caused their emotions, and requires that the subject has the ability to differentiate between internally and externally generated emotions.

Sympathetic concern differs from mere emotional contagion because it involves other-orientation: An affected party that is not capable of sympathetic concern will only seek to relief the personal negative effects induced by automatic state matching. The reaction to the emotional state of the other without sympathetic concern is thus selfish and aimed at the aversion of negative vicarious arousal, whilst sympathetic concern involves an understanding of what caused the other's emotion and an attempt to alleviate the others' distress for the sake of the other.

A well-documented example of sympathetic concern is *consolation*. This behavior, defined as the reassuring and comforting attempts at the display of a distressed party, is very common in humans and apes (de Waal, 2008). Monkeys do not share the ability for consoling behavior with the Hominoidea. For example, macaque mothers fail to comfort their own children after fights (Schino et al. 2004).

The contrast in capacity on this point between monkeys and apes supports the *coemergence hypothesis* (Gallup, 1982): the prediction that advanced expressions of empathy and mirror self-recognition (MSR) appear together in development and phylogeny.

Advanced expressions of empathy require a heightened sense of self from the individual. This heightened sense of self is for example expressed by *theory of mind*, the capacity to attribute mental states (beliefs, desires, intentions, knowledge) to others and an understanding that these states can differ from one's own mental states. It is hypothesized (Gallup, 1983) that this heightened sense of self is related to the capacity of recognizing oneself in the mirror. Mirror-self recognition has thus far only been attested in humans, apes, dolphins and elephants.

In de Waal's Russian doll model of empathy (2008), this heightened sense of self marks the difference between more primitive forms of empathy and *cognitive* empathy.

The next layer in the Russian doll model that correlates with the ability of cognitive empathy is (3) *perspective taking*. Perspective taking is the capacity to understand the situation and needs of the other and knowing that it is separated from one's own situation and needs. Combined with automatic, subconscious emotional engagement, it is called *empathic perspective taking*. Only in this context psychologists speak of empathy, thus focusing on the cognitive affair rather than the automatic processes that underlie it, a point of view that explains the skepticism about nonhuman empathy (de Waal, 2008).

An expression of empathic perspective taking that is well documented in primates is *targeted helping*: help that is specifically aimed at the needs of the other through an understanding of their situation. This requires a heightened self-identity because one needs to be able to attribute mental states to the other, have an understanding of what caused the emotional state of the other and what is needed to ameliorate that situation. Plus, the induced state in the subject must be correctly attributed to the source, i.e. the state of the object instead of to the subject itself.
Examples of targeted helping, which is a form of altruism, is well documented in other species. Based on the capacity for empathic perspective taking, many animals including dolphins, elephants and apes show behavior that is aimed at helping others reach their goals (de Waal, 2008).

The involuntary process of state matching that lies at the core of all empathic, prosocial behavior, is biased. Empathic activation tends to be stronger when there is more familiarity and similarity between object and subject: it increases the accuracy of state matching. In humans, research showed (Singer et al. 2006) that subjects empathized more with the people they had a cooperative relation with compared to the people they had a competitive relationship with. The quality and nature of the relation influences the empathic or antipathic response in humans: when shown pictures of a cooperative ally in pain, pain-related brain areas were activated in male research subjects, whereas seeing a non-cooperative or unfair confederate in pain activated reward-related areas. So, having empathy in part relies on one's familiarity to the other and whether one has a positive relationship with the other, which emphasizes the importance of the second pillar of morality: *reciprocity*.

## *Reciprocity*

In evolutionary biology, reciprocity is referred to as the mechanisms that favour altruistic and cooperative behaviour over self-interested behaviour, because it establishes a positive reassurance for future interactions between individuals. Associated with reciprocity is a sense of justice and fairness: the ability to determine whether both cooperating parties profit equally over time. Cooperation could not have evolved without such a mechanism that compares pay offs with investments (Brosnan & de Waal, 2014). This comparison requires a sensitivity to (un)fairness which is expressed by a negative response to receiving less than a partner for a similar task, a reaction known as *inequity aversion* (IA). Inequity aversion is subdivided in (1) *disadvantageous IA*, a negative reaction to inequity at the cost of the actor and (2) *advantageous IA*, a reaction to inequity that benefits the actor. The former, a passive or active reaction to receiving less than a partner for the same task, is widespread in species that cooperate outside kin ties and mating bonds and probably evolved because those who did not react to unequal reward had a disadvantage over those who did

because they received less. Reactions to unequal reward are measured relatively easy through empirical investigation and have been found in monkeys, apes, dogs and birds. When there is a reaction to unequal pay, this indicates the ability to judge and respond to value (de Waal, 2006). Disadvantageous inequity aversion is a reaction to receiving more than a partner for the same task, and amounts to a full sense of fairness that allows an individual to turn down an immediate benefit because of the negative effect it might have on the relation in the future. Disadvantageous IA has so far only been observed in humans, and to a certain degree in apes. Disadvantageous IA most likely is an attempt to secure future cooperation options by anticipating inequity aversion in the partner that is receiving less. By reacting to unfairness even though the direct cost of inequity merits the receiver, the receiver is aware that the immediate reward does not outweigh the benefits of continued cooperation with the other party. This requires the ability to attribute mental states to others and the ability to plan for the future, and transcends the 'egocentric' sense of fairness that reacts to the disappointment in expectation of how one is treated, instead of reacting to expectations of how everyone should be treated. The latter reaction concerns multiple others and groups of others as a whole and is called 'community concern'.

Brosnan and de Waal (2014) hypothesise that the evolution of advantageous IA combined with advanced cognitive abilities, such as the capacity for complex language systems, allowed for the development of a complete sense of fairness in humans. Secondly, community concern, the concern for the stability of the group as a whole opposed to the primary focus on the welfare of the individual and its kin, marks another difference between human and primate morality. The ability to focus on shared interests, common goals and the greater good, is needed in the light of human societies as they bring forth much more complex situations and moral issues compared to primate societies, situations in which we aim to reach consensus through moral discourse, deliberation, and reasoning. The ability to use language attributed to these more advanced systems of morality, as communication on a large scale helped to tackle some of the major complexities of life in large groups.

In the next chapter, I will illustrate why evolutionary biologists think language played such an important role in the transcendence of the limits of primate societies. Then I will illustrate how Jonathan Haidt departs from this evolutionary function of language and morality to explain his social intuitionist model of moral judgement.

# Chapter two: Sociality and language

Human beings are amongst the most ultrasocial species of mammals that inhabit the earth. Only the naked mole-rats of East Africa live in such huge and highly cooperative groups, but they, like ants and bees, are all siblings and reap the benefits of kinship altruism (Haidt & Bjorklund, 2006). Human beings do not have this advantage and yet, they still maintain social and cooperative bonds that extend around the globe. Human societies, like societies of other members of the Hominidae, are built upon advanced forms of social cognition. But what sets human societies apart is that they exceed the number of members that inhibit their groups. Compared to the groups of our primate relatives, consisting of a maximum of 80 individuals, natural groups in humans usually have around 150 members. What aspects enabled human societies to push beyond this boundary and with that, to transcend the characteristics typical of primate societies? What is it that bridges the gap between animal and human morality besides cognitively advanced forms of empathy and advantageous inequity aversion? To answer these questions, I will first look at the premises that constitute primate societies and how apes and monkeys maintain group stability, to define what systems helped the evolution of human morality. Then I will illustrate how this evolutionary story about morality takes up a central position in *the social intuitionist model of moral judgement* (SIM-model) proposed by moral psychologist Jonathan Haidt (2001). This model poses a challenge to the rationalist models that have dominated research in moral psychology, by answering the question: *what causes moral judgement?* With an elaborately funded six stage model theory that favours an explanation based on the inherent social nature of humans and the strength of our affective system over theories that give precedence to the unique human ability to reason.

## *Primate societies*

Expressions of social behaviour, underpinned by different stages of cognitive competence, are vital to the species of the Hominidae (Dunbar, 2004). Because natural selection is a neutral process, simply favouring the individuals that are best at surviving and reproducing, there is nothing inherently 'good' about being social and functioning well within a group, it just offers major selective advantages over life in solitude. One of the most important benefits of living together in a group is that formation offers individual primates protection from predators, which is vital to survival as a species' ecological habits expose it to the great risk of being hunted by others (Dunbar, 2004). A big disadvantage of group life on the other hand is that personal preferences need to be sacrificed on the altar of the greater good. This means that individual, short-term goals must be compromised in order to achieve long-term risk reduction. The solution of primates to deal with the disadvantages of

life in large groups is to form smaller alliances that help and protect each other. These alliances are often, but not always, kin related and are maintained by grooming. This important social activity, also considered 'the glue of primate societies' (Henzi & Barett, 2005, p.1865), releases high amounts of endorphins in the receiver and brings it into a state of relaxation, increasing the willingness to cooperate with the actor. The strength of the alliance is dependent on how much time is devoted to the activity of grooming, because it increases feelings of trust and commitment between the members. On an average, 1/5$^{th}$ of the waking time of primates is devoted to it (Dunbar, 2004). The bigger the group, the more need there is for strong alliances as competition over resources and stress over reproduction increases and attacks by others are more common. But to form strong alliances, primates need to devote a significant amount of time to grooming. Ultimately, there is a limit to the number of members a group can have while maintaining the option for its members to commit 20% of their energy to this form of social bonding, and that number lies at around 80 individuals.

When our ancestors started to explore more predator risky territory, like the open terrestrial habitats of the savannah, it required a growth in size to secure the safety of the group. And a growth in group size beyond 80 required more effective ways of social bonding, because more social relationships needed to be maintained by the members in the same amount of time. The only way humans where able to push beyond the group size boundary and explore more riskier habitats than forests, was to have an alternative method of social bonding that had roughly the same effect as grooming, but dealt more efficiently with the time devoted to it so the number of social relationships could increase.

Evolutionary biologist Robin Dunbar (2004) argues that gossip, devoid of its negative character but more specifically defined as the exchange of social information, has been crucial to this extension of group size in hominid evolution. Language, its facilitator, is said to mark the difference between us and our primate relatives (De Waal, 2018, p.248).

The growth of the neocortex of the human brain facilitated the capacity for language, and gave our ancestors the ability to establish social efficiency through conversation (Dunbar, 2004, p.102). Where grooming is a time-consuming one-on-one activity that leaves no space for multitasking, in conversation, multiple listeners can be addressed and it offers the possibility to tend to other tasks while doing it. When talking to someone, I do not only express social commitment- *'I am talking to you specifically and not to them'*-, I also have the option to exchange information about the status of members of my social network, the reputation of others and the norms and rules that exist within my group. Monkeys and apes are limited to what they see and only have knowledge available to them that is gained by first-hand experiences with others.

## *The evolutionary function of gossip*

Dunbar argues that gossip derived its bad name from the consideration that it is a rather unsophisticated use of language compared to all the other, more elaborate ways in which humans use language, for example for the exchange of valuable technical information or the accumulation of human knowledge about the world. However, without the possibility to exchange information about our social network, the formation of large groups that eventually produced all that knowledge would not have been possible in the first place. It is true that gossip is often used to comment on the behaviour of others, but this, according to Dunbar, is vital to the existence of large groups as gossiping is an effective way to single out freeriders who pose a serious threat to group stability.

Freeriders are non-cooperating individuals who profit from the help offered by others without paying the costs of offering help themselves (Enquist & Leimar, 1993). The number of individuals that freeride grows as groups grow: freeriders become less visible and there are more 'naïve' members from whom they can take advantage. For cooperating individuals, it is very costly to be profited from because they carry the costs of cooperation and get nothing in return. It is also very time costly to constantly be suspicious of everyone in the community, and it also undermines group stability. To establish and reap the benefits of wide spread sociality there must be a level of trust between members, but it is exactly this level of trust that facilitates freeriding.

Thomas Hobbes (1651), convinced of the egoistic, rational nature of human beings, designed his social contract theory to tackle this specific problem. The premises of this theory still permeate large parts of the law, economics and political theories of Western civilization (de Waal, 2006, p. 3). If human beings are, as Hobbes argues, merely driven by self-interest, being a freerider is the inevitable obligatory outcome for everyone because a freerider receives the benefits of cooperation without carrying the costs. However, being driven by self-interest and knowing that everyone else is, creates the impossibility of group forming. In such a world, cooperation is the suboptimal choice: no one wants to pay the costs of cooperation without the assurance of earning back their investments. Because Hobbes perceived of humans as calculating, self-interested rationalists, he believed that nobody in their right mind would naturally choose to collaborate, turning cooperation and the benefits of joint effort into an unreachable goal.
Hobbes used this view of human nature to design a self-interested escape route out of the state of nature. The state of nature is Hobbes idea of how the original, pre-civilized world looked like, where there were no social restrictions, no rules and no norms. Because he was convinced of the evil, self-

interested nature of humans, Hobbes believed that this world was marked by a continuous war of all against all, in which there was " *no Knowledge of the face of the Earth, no account of Time, no Arts; no Letters; no Society; and which is worst of all, continuall feare, and danger of violent death: and the life of man, solitary, poore, nasty, brutish, and short*" ( 1651/2009, p. 179).
To get out of this eerie state, Hobbes' citizens hypothetically sign a contract with one another in which they declare to give up their right to individual freedom and hand it over to a Leviathan in order to establish a commonwealth. The Leviathan's main task is to guard the contracts, and anyone who is caught breaking it by being a freerider becomes an outlaw and is sent back to the state of nature. The rational obligatory option for Hobbes' citizens is to abide by the contract and curb their natural freeriding tendencies. Being self-interested also means to have the wish for continued survival, something that is possible in the commonwealth but not in the state of nature.

If we accept Hobbes' first premises, that humans are inherently selfish and rational, such contract-like situations are needed to constitute sociality. It is the glue that unites beings who are naturally inclined to act on what is best for themselves and in doing so will be wolfs to each other.
In Hobbes' worldview, the essence of human nature leads to the inability of cooperation and only through contracts and well-ordered societies, living together becomes a possibility. For Hobbes, these contracts are a matter of pure rational necessity, a vital condition for continued life amongst free and equal but also selfish and rational human beings. The social tendencies that are constituted via contracts are not natural, but artificial phenomena, and rules and reasoned judgement are needed to supress our inherently evil and anti-social nature.

Frans de Waal, on the other hand, argues for an opposite view of human nature based on his extensive research and observations of primate behaviour. The ancestral state from which human beings sprang, the apes and monkeys belonging to the subgroup of primates called the catarrhines, is ultra-social and based on different stages of social cognition. Human morality, in this view, is a cognitively advanced extension of forms of prosocial primate behaviour. What underlies these prosocial tendencies are the capacities for empathy and reciprocity, and the behaviour by which they are expressed are targeted helping, consolation, reconciliation and inequity aversion.
In moral psychology, Jonathan Haidt defends the position that our moral judgements are guarded by the moral emotions, which in turn are all derived from five foundations that have been relevant in the light of hominid evolution. The emotions motivated by these foundations create quick flashes of moral evaluations, a process akin to perception, and only in hindsight do people use reason to explain the nature of those judgements to others.

Both views are an expression of human moral behaviour as coming from within rather than installed from outside. They defend bottom-up views of human morality that locate the natural tendency of altruistic behaviour in phylogenetically old, automatic processes. Over the course of evolution, different modules allowed for a cognitive sophistication of these processes and so it was moulded into the elaborate systems of morality we apply today. Because sociality and group-orientation are crucial to survival of the lineage of beings from whom the human species evolved, and human morality is an advanced expression of the module to maintain stability within the group, "*if any decision to establish societies was made, therefore, credits should go to Mother Nature rather than to ourselves*" (de Waal, 2006, p.4).

So are we social all the way? Do people never take precedence in their own interests or the interests of their kin? Well, not entirely. Not when it comes to collective action. Evolutionary biologists Enquist and Leimar (1993) established through modelling that the strategy of being a freerider is evolutionary stable, which means that freeriders will eventually, if they are able to remain invisible, out reproduce honest co-operators in a group. The size and mobility of a group of individuals constrains the evolution of cooperation because it creates the opportunity for its members to rapidly switch between partners, and freeriders are able to move invisibly through a group of cooperating individuals. Enquist & Leimar hypothesized that a community of collaborators with a single 'mutant' freerider is driven to extinction within just a few generations. Hobbes was not entirely wrong when he assumed that collective action will inevitably lead to freeriding when there are no forms of social control. And a community of freeriders is no community at all. So, if the credits for establishing human societies should go to Mother Nature rather than to human rationality, then how did she get us out of the freerider loophole without hypothetical contracts or something of the like?
Unless our human ancestors had means to share information about the norms and rules within the group and the reputation of others, they could have never formed a society. Somehow, they had to curb the bandwidth of freeriders and maintain a dominant number of co-operators in the group. Enquist and Leimar (1993) refer to this crucial process as 'gossiping', a form of social control that takes away the invisibility of free riders. The only way to maintain stability in a group that exceeds the number of members that can rely on time-consuming one-on-one relations and first-hand knowledge about the reputation of others, is to have means to share information with one another. The exchange of social information forms an external motivation to curb the intention to both satisfy short term egoistic goals and profit from the benefit of group collaboration. When others can share information about your lack of cooperation, your bad actions or what you did to others, you can gain a reputation that can seriously decrease your options for future cooperation. The ability to gossip works demotivating for norm opposing behaviour and freeriding.

The ability of learning and assessing complex language systems enabled the human species to transcend characteristics typical of primate societies, such as being subject to power hierarchies and the cooperative limits of dyadic and triadic relationships. A combination of full theory of mind, the evolution of language and the ability to engage in reflective reasoning, enabled our species to define group norms and values, reflect on one's own reputation and the reputation of others, and apply social control in large groups. The amount of time, around 65%, that is devoted to gossip in average human conversations for example (Dunbar, 2004, p.100), but also the billions of cultural folk tales and myths, the development of philosophical ethics that is so extensive and started so early in antiquity, and the abundant number of religious belief systems that exist in the world, can be seen as examples of the basic human need for social evaluation and group norm enforcement.
According to Haidt, it is sociality that both shapes the set of foundations that are highlighted in our moral lives and it is where reasoning commonly takes place in the process moral judgement. That is why in Haidt's model, the definition of a moral judgement departs from the behavioural fact about human beings that in every society, people evaluate each other's actions and that these evaluations have consequences for future actions (Haidt, 2001, p. 817). Applying a biologist's perspective through a bottom-up method, Haidt turns to the *arche* of behaviour to find a starting point for the evaluation of the process of moral judgment.

## *The social intuitionist model of moral judgement*

The central claim of the social intuitionist model, a model that consists of four principal and two additional links, is that moral judgement is caused by intuitions and is followed by ex post facto reasoning. This ex post fact reasoning occurs almost exclusively in a social context. With this model, that to a great extent builds on empirical research from the fields of primatology, evolutionary biology and psychology, Jonathan Haidt offers an alternative to rationalist explanations of moral judgement in psychology and philosophy.

**Moral judgements:**
*'Evaluations of a person's character or actions that are made with respect to a set of virtues held to be obligatory by a culture or subculture'* - (Haidt, 2001)

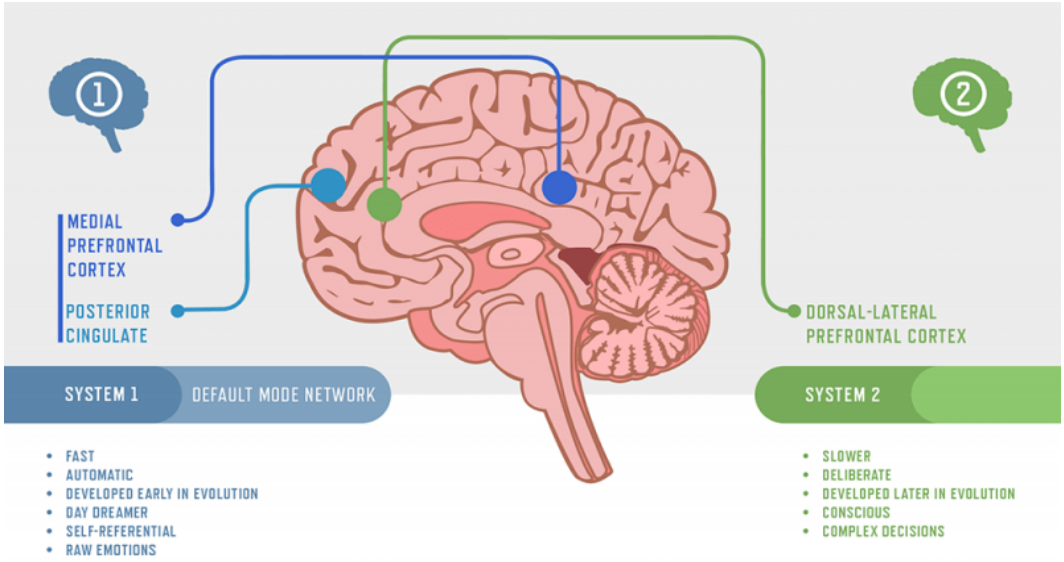
One of the central questions addressed in moral psychology is what it is that causes moral judgement. Rationalist approaches (Piaget, 1932; Kohlberg, 1969; Turiel, 1983), which have ruled the debate in psychology since the cognitive revolution that took place in the late 20$^{th}$ century, argue that moral judgement is the end product of a conscious, deliberate mental process in which a person weighs issues of harm, justice and fairness to reach a conclusion about a certain state of affairs.
Haidt's scepticism about this causality of reason in moral judgement was, inter alia, motivated by his research on harmless taboo violations (Haidt, Koller, and Dias, 1993) in which subjects were

presented with stories that caused an immediate affective reaction, but involved no harmful consequences for the parties involved in the story.
Affective reactions such as disgust, aversion or fear caused subjects to morally reject the actions of the parties involved in the stories, while the reasons that are traditionally seen as matters of moral consideration such as harm, justice and fairness did not apply to them. Even with the latter being explicit, the subjects hardly changed their initial condemnations and would, upon further interrogation about their reasons, often display embarrassment and start to stutter or laugh, making statements such as *'I know it's wrong, but I just can't come up with a reason why*'. Haidt dubbed this phenomenon 'moral dumbfounding' and in other, similar experiments (Haidt et al. 2000) the same thing would occur when subjects were unable to find supporting reasons for their judgements.

These types of experiments can be criticised for their lack of *ecological validity* (Haidt, 2001, p.829), the extent to which such hypothetical situations can predict behaviour in the actual world. However, the way in which people reacted to these hypothetical situations gave reason to think that (1) there are other sets of values that people can apply in immediate moral reasoning, apart from whether an action causes harm, is unjust or unfair, but also (2) that there might be different forms of cognition other than conscious reasoning at play, forms of cognition that might be less accessible to the conscious mind, but run parallel and are of equal importance in the process of making a moral judgement.

*Dual process systems*

In recent years, there has been a growing consensus in social and cognitive psychology that there are two processing systems at work in the human brain when problems are solved or judgements are made: an ancient 'hot' system that runs fast and automatically, and a phylogenetically newer, 'cool' system that involves slow deliberation and relies heavily on verbal thinking (Chaiken & Trope, 1999). The clumps of neural tissue that make up the human brain constantly evaluate everything they hear and see along a like-o-meter, aimed at answering the fundamental question of whether to approach or to avoid: a question even one-celled organisms must answer (Haidt & Bjorklund, 2006). The size of the human brain enabled for the capacity to slowly evaluate and analyse information, giving us the ability to come up with more detailed, deliberated answers to this fundamental question. However, at the core of this capacity lies the inescapability of our brain's affective nature that will always, in a split of a second, judge whether something is good or bad (Haidt, 2001).




**'Our Two Brains, Mindfulness and Decision Making' (2019) mindsciences.com**

The social intuitionist model takes the inescapability of the affective mind as the outset to defend that not only general judgements, but moral judgements too, are primarily based on quick and effortless processes that draw information from moral foundations or categories to form a judgement in a split of a second. This produces a certain gut feeling, something similar to aesthetic judgement, and is a form of cognition described as 'intuition'. It occurs immediately upon confrontation and is described in the model as the (1) *intuitive judgement link.* After the initial judgement is made by this process, that is inaccessible to the conscious mind, the individual can engage in reasoning. This process is described in the model by the (2) *post hoc reasoning link.* Reasoning is described as an effortful, conscious process, but one that is only engaged in to find supportive arguments for already arrived-at judgments.

It is this very link that explicitly opposes rationalist theories of moral judgement, because it states that the unique human ability to engage in conscious reasoning and evaluate positions, weigh moral principles, apply a-priori thoughts and logically deduce conclusions, is a secondary process, which is subject to direct, automatic processes that are inaccessible to the conscious mind. Critics argue that reasoning thus becomes a matter of mere confabulation because it loses all of its decisive strength in Haidt's model (Horgan & Timmons, 2007, p.282), and this, in turn, is hard to digest for those who believe firmly in the power of reasoning. The alternative function that Haidt attributes to reasoning will be elaborated upon later in this thesis.

To continue our discussion of the model, the third link, the (3) *reasoned persuasion link*, together with the (4) *social persuasion link,* emphasize the social function of morality that is suggested by the work of Frans de Waal and the evolutionary function of gossip that is suggested by Robin Dunbar.

The third link namely describes how the reasons made up in link (2) are presented to others to justify the judgement made in link (1).

The need to convince others of held convictions is a means to enhance or defend one's own reputation, which makes sense if we accept the view that the driving force behind the development of language in hominid evolution was not the option to find eternal 'truth', but rather the option to discuss behaviour of others and by doing so creating the opportunity to apply social control on a large scale. Those who are successful in reaching consensus on normative issues in the community they belong to, increase their own fitness as they will be more likely to reap the benefits of cooperation than those who fail to do so (Haidt & Bjorklund, 2006, p. 10).

So, when important others do not agree with the judgements made in (1), even after the reasons made in (2) are presented, this might lead to the need for reasoned persuasion (3) in order to get the motivation of the actor across. And this, in turn, is most effectively done when new intuitions are triggered in the listener.

Therefore, giving reasons in Haidt's model does not refer to offering logically compelling reasons, but rather to reasons that appeal to the feelings of the listener, with the aim to evoke new intuitions that support the claim of the actor. This, according to Haidt, is far more effective than offering rationally sound arguments. The (4) *social persuasion link* then, describes the human sensitivity to group norms and the direct influence of others in moral judgement: the fact that an ally, friend or acquaintance holds a moral judgement can account for the same judgement in the individual, even without reasoned persuasion. This link once again stresses the ultra-social nature of humans. The same psychological accounts that enable us to cooperate intensively with non-kin makse us extremely sensitive to the adaption of group norms and the need to fit in. Even when rational arguments fail to do their job, the desire to be part of a group and adapt to its norms can hardly be trumped because belonging is so vital to the survival of our species (Baumeister & Leary, 1995).

Finally, Haidt argues that the addition of the last two links makes the social intuitionist model only anti-rationalist in a limited sense (Haidt, 2001, p. 815). The (5) *reasoned judgement link* and the (6) *private reflection link* are the links in Haidt's model that can indeed be seen as his most important concessions to defenders of rationalism (Sauer, 2011, p.714). These links describe that (5) reasoning may happen in a private context and override the initial intuitions made in (1), but only when there was already more of a dual attitude towards the triggered judgement from the beginning, and (6) describes that one can be motivated to take on new positions through the process of role-taking: by having an inner dialogue in which one empathizes mentally with the position of another, new intuitions can be triggered based on the shifting perspectives this role-taking brings forth. Both links however are said to occur very rarely, and often only happen spontaneously in those who have had extensive training in modes of unnatural thought, like philosophers (Haidt & Bjorklund, 2006, p. 14).

Haidt argues that in our everyday moral reasoning we are so biased by the search for arguments that support our initial judgments, that objective, a priori and logical reasons hardly stand a chance to convince us of other positions.
In favour of the SIM model, it needs to be stressed that the model does not deny a role for reasoning in the process of moral deliberation (Haidt & Bjorklund, 2006, p.8), it only denies that reasons possess the decisive strength and objective character that rationalists attribute to it. In Haidt's argument, reasoning, as opposed to intuitive judgement, is an effortful, conscious process one engages in to justify their beliefs to others. Because it is a phenomenon occurring in a social rather than in a private context, it is more important that reasons are compelling socially rather than logically.
Even though the structure and motivation of these reasons might not match what rationalists' value to be important, because they are not based on a-priori principles or logic validity and do not play a primary, causal role, the capacity to give these socially compelling reasons is of utmost importance in the model and requires similar mental effort. The ability to justify our intuitions and compel others serves a social goal: it is the 'glue' that serves our fitness.

However, if we want to compel others, we do need shared moral values or certain ethical frameworks we can refer to. Moreover, the intuitions that are triggered in (1) must come from somewhere, there must be a 'beginning' that motivates us to evaluate certain states of affairs in a moral way, while denying the need to evaluate other situations in the same manner. What is it that triggers our moral reflection? And if this shared common ground is not found in universal truths about the world or a-priori reasons appearing suddenly in consciousness, then what is it that triggers our moral evaluations and has shaped the different moral systems of this world?
In the next chapter I will discuss Haidt's answer to this question in the form of his foundations theory (2006), which describes the evolutionary prepared modules or basic 'taste buds' that constitute all varieties of human morality.

# Chapter three: The first draft of the moral mind

Where the work of Frans de Waal focusses on the roots of altruistic behaviour in primates, Jonathan Haidt's research focusses on what causes human moral judgement and what foundations lie at its roots. Both make use of a bottom-up perspective to morality, de Waal by studying morality as a continuous phenomenon that is shared across species and Haidt by searching for what is written on the first draft of the moral mind and referring to evolutionary old systems that influence our judgements. In this chapter, I will illustrate why Haidt argues that morality is constituted by specific five domains and how the human mind is, through these innate domains, 'prepared in advance of moral experiences'. Just like the necessity for social behaviour is innate, the foundations that underlie our advanced systems of human morality are innate too. These moral frameworks function as the motivational basis for the concepts that subconsciously guide our ethical reflection, and are the places from where we draw the motivation for our moral judgements.

## *Quandary ethics*

The idea that human morality contains several domains and is thus not limited to issues of harm (/care) and fairness (/reciprocity/justice) contests the premises of *quandary ethics*, the generalized version of morality that prevails in modern Western philosophy and is widely applied in research in the social sciences. Quandary ethics, a term coined by Edmund Pincoffs (1986), is the modern consensus that ethics is about rationally solving dilemmas between individuals as opposed to the idea that ethics is about developing a good character and the process of becoming a virtuous person. The latter, better known as virtue ethics, involves a form of moral education that was more common in the ancient cultures of the West (Greece) and the East (India and China). Virtue ethics prescribes how the intuitions and emotions ought to be shaped through slow practice and reflection, whereby the goal is to slowly mature towards psychological perfection and moral virtuosity. In theories of virtue, this virtuosity is often exemplified by role models, such as Homer in Greece or the Mahabharata in India. This development of virtues through practice was widely regarded as morality until the 18$^{th}$ century came and brought the gestalt switch of the Enlightenment (Haidt & Joseph, 2006).

And even though philosophers like Alistair MacIntyre (1981) have passionately argued for a return of this approach to ethics in moral philosophy, a point on which I will elaborate later in this thesis, quandary ethics remains a popular approach to morality in academia. The popularity of these type of theories and viewpoints is for example reflected in the master courses of Applied Ethics at the University of Utrecht.

During the Enlightenment, objectivity, verifiability and empirical truths became strongly favoured over religious assumptions and metaphysical beliefs, which where even scrutinized. Enlightenment philosophers such as Hobbes, Kant, Rawls and Locke thus wanted to move away from these virtue-based moralities which embodied substantive and often religious ideas about human nature and society.

In this time, philosophers searched for ways to ground moral guidance not in metaphysical ideas about the world or human nature, but in abstract, logically verifiable rules one could physically point at. The ethical theories the philosophers of the Enlightenment gave birth to can be divided in *formalist* and *consequentialist* theories of morality (Haidt & Joseph, 2006). These theories have in common that they take hypothetical, contract-like relations between agents as the outset to explain moral rules, that both theories rely heavily on the human ability to reason, and both theories favour abstraction and universal verifiability over contingency and the particular. The difference between the two theories is that formalist theories in general design moral judgements according to logically valid principles, such as Kant's categorical imperative or Hobbes social contract theory, and consequentialist theories like utilitarianism look specifically at what action brings about the most favourable consequences and value the best outcome as morally right.

So formalist and consequentialist theories make it possible to either calculate or logically deduce the morally right thing to do, limiting the scope of origination to dilemmas caused by the competing interests of agents, making moral practice a reasoned practice confined to conflicts between individuals. The outcomes of these theories have a lot of practical advantages: moral rules can be unified, they are honest, clear, explicit and workable. Because of the commonalities and the applicability of these theories it is not strange that quandary ethics became widely applied in the social sciences and humanities.

For example, the pioneering work of psychologist Lawrence Kohlberg (1969) used quandary ethics to measure the process of moral development in young children. He argued that children gradually get better at applying abstract moral rules to moral dilemmas, until they reach the ultimate stage of his 'six stages model of moral development', when they are able to draw judgement from post-conventional ethical principles, such as justice and universalizability.

According to Haidt, this view of morality as being limited to solving dilemmas using values related to harm and justice, is parsimonious. It does not capture the full human moral domain and what other motives people globally apply in their moral judgements. Instead, quandary ethics focusses on what modern Western cultures prioritize, like ethics that value and protect the autonomy of individuals,

while in other cultures issues of religious sanctity or in-group loyalty might be of far greater importance.

The idea that morality is about much more than issues of harm and justice was inspired by the results of research on harmless taboo violations conducted by Haidt, Koller, and Dias (1993). In this research, subjects were interviewed about a set of stories that evoked an immediate reaction, but where not related to issues of harm and justice. For example, '*A family eats their pet dog after it was run over by a car*' or *'a woman cuts up an old flag and uses it to clean her toilet with'*. The group of research subjects consisted of adults and children from Brazil and the United States, half of high social class, half of low social class. Haidt et al. found that only the Western high social class adults did not perceive of these stories as 'morally wrong', they found them maybe strange or disgusting, but not reprehensible from a moral point of view. The other groups however, treated the harmless taboo violations as moral violations that where wrong and universally wrong, giving reasons that pertained to disgust, disrespect, and the violation of norms and rules. They seemed to apply a much broader moral domain than what is included in ideas of quandary ethics.

**Innate**: *'Prepared in advance of experience'* - Gary Marcus (2004)

What other foundations are part of the moral domain according to Haidt and Joseph (2006)? And what do they mean when they argue that these domains are innate? And how do they explain the innateness of these different domains in the light of cultural differences?

Lawrence Kohlberg believed that the only innate moral capacity that is present in the human mind from the outset is the outer layer of de Waal's Russian doll model. Empathy, specified as 'perspective taking' accumulated in the early stages of human life and enabled a child to adapt and learn from her environment. The capacity to differentiate between like and dislike helped her to do so: what I find pleasurable is probably pleasurable for another person too. It was through active participation Kohlberg believed that the child learned rules, the value of those rules and what external goal they served; a view of moral development known as *constructivism*.

Others, like Haidt and Joseph, believe that there are more concepts written on the first draft of the moral mind and that these concepts relate to certain adaptive advantages they gave to our ancestors. These concepts, which they call *intuitions*, connect certain patterns in the social environment to evaluations and moral emotions in the brain, resulting in quick flashes of mental output that strongly influence moral judgement (Haidt & Joseph, 2006). This process, which is not fully controllable, is extended and shaped by cultural learning. The authors view moral development of children as a gradually growing recognition of a large set of input patterns, to which their brain quickly and automatically learns to respond to with output, output that is prepared in specific domains of moral concern that have been valuable in the light of evolutionary advantages. So, there are innate, involuntary intuitions about what matters morally, which are shaped and moulded by

cultural beliefs. According to this theory, there are five domains that become moralized during development: they argue there might be more, but they claim that at least, social issues that cannot be related to one of these five categories are hard to learn or to motivate people to care about.

## *Haidt's five innate moral foundations*

The foundations are present in what they call 'the first draft of the moral mind', either as learning modules or as a form of evolutionary preparedness, and make up the psychological foundations of human morality. These five intuitions, or psychological 'building blocks' of morality, are (1) Harm/Care (2) Fairness/Reciprocity (3) Ingroup/Loyalty (4) Authority/Respect and (5) Purity/Sanctity. These foundations are innate in the sense that they are related to the adaptive challenges that ultrasocial animals have faced over the course of evolution, giving those who recognized and adapted more efficiently to difficulties by drawing behaviour from these categories a selective advantage over those who did not. These five intuitions are considered 'the taste buds' that lie at the root of all our moral judgements. Even though the five foundations of human morality serve evolutionary adaptive goals in essence, they have all been generalized and modified into 'modern' considerations. The figure below, drawn from Haidt & Joseph (2006) gives an overview of what adaptive challenges the foundations served, how they are triggered and what characteristic emotions, virtues and vices fall in their category.

| | **Harm/Care** | **Fairness/ Reciprocity** | **Ingroup/ Loyalty** | **Authority/ Respect** | **Purity/ Sanctity** |
|---|---|---|---|---|---|
| **Adaptive challenge** | Protect and care for young, vulnerable, or injured kin | Reap benefits of dyadic cooperation with non-kin | Reap benefits of group cooperation | Negotiate hierarchy, defer selectively | Avoid microbes and parasites |
| **Proper domain (adaptive triggers)** | Suffering, distress, or threat to one's kin | Cheating, cooperation, deception | Threat or challenge to group | Signs of dominance and submission | Waste products, diseased people |
| **Actual domain (the set of all triggers)** | Baby seals, cartoon characters | Marital fidelity, broken vending machines | Sports teams one roots for | Bosses, respected professionals | Taboo ideas (communism, racism) |
| **Characteristic emotions** | Compassion | anger, gratitude, guilt | Group pride, belongingness; rage at traitors | Respect, fear | Disgust |
| **Relevant virtues [and vices]** | Caring, kindness, [cruelty] | fairness, justice, honesty, trustworthiness [dishonesty] | Loyalty, patriotism, self-sacrifice [treason, cowardice] | Obedience, deference [disobedience, uppitiness] | Temperance, chastity, piety, cleanliness [lust, intemperance] |

The first two, (1) Harm/Care and (2) Fairness/Reciprocity, are what Frans de Waal considers to be the pillars of morality that serve as the foundation of altruistic behaviour for all cooperative mammals and explain the evolution of prosocial behaviour. The next two, (3) Ingroup/Loyalty and (4) Authority/Respect, are meaningful in the context of the large social groups primates and humans have lived in, and serve goals like group stability and the sustainment of dominance hierarchies. These four domains of morality developed in the context of sociality, but (5) Purity/Sanctity, which is thought to be a consideration unique to the human species opposed to the other four, probably developed in the context of nutritional adaptive challenges. The diet of our human ancestors transferred from fruit and plant-based to heavily meat-based only recently compared to the long background of primate evolution (1-3 million years ago) and this occurrence coincided with the growth of the frontal cortex of the human brain. Because a meat-based diet increases the risk of food containing microbes and parasites, and because at the same time the human frontal cortex, the 'control panel' of our brain, started to grow rapidly, it is thought that humans and only humans have developed the emotion of disgust (Rozin et al, 2000), which helped to avoid disease transmission by causing aversive feelings towards anything that could cause it. Gradually, disgust became a social emotion and the virtues connected to it – cleanliness and purity, got reflected in spiritual values, such as chastity and piety.

## *The ever-growing abyss*

The five foundations can be seen as separate psychological preparations from which we (unconsciously) draw our moral judgements. Based on the culture one is born into and the values one learns to appreciate during development, some of the foundations will be of more individual importance than others. To illustrate how different sets of foundations influence different ideas about morality, it is helpful to take a look at how Haidt & Graham (2007) use the moral foundations theory to explain the 'moral gap' between conservatives and liberals in North America.

According to Haidt and Graham, what divides liberals and conservatives in their political standpoints is that in their moral reasoning, liberals apply a 'narrow' version of the moral foundations, exclusively relating to issues of (1) Harm/Care and (2) Fairness/Reciprocity, whilst conservatives apply the full range of foundations in their moral reasoning, in which (1) and (2) may matter, but only as $2/5^{th}$ . Richard Shweder (1990) called the liberal version of morality 'the ethics of autonomy', in which moral regulation is aimed at the protection of the individual, and moral goods such as freedom, fairness, rights and justice are there to promote and protect the autonomy of individuals. This type of ethics, reflected in theories of quandary ethics, is often the only type of moral consideration that

is visible for liberals. It can be distinguished from a more conservative approach in which 'the ethics of community' and 'the ethics of divinity' are of equal relevance. 'The ethics of community', matches with foundations (3) Ingroup/Loyalty and (4) Authority/Respect and does not perceive the world as a collection of individuals but rather as a collection of groups that have their own separate identities, and morality functions as a way to protect the integrity and the stability of the group by thrusting great value upon institutions and traditions and the purpose of virtues such as duty, loyalty and respect. 'The ethics of divinity', is reflected in foundation (5) Purity/Sanctity, and builds upon the premise that God exists and that every soul is a gift from God. Because of the divinity of the human soul, individuals are not allowed to treat their bodies in any way they like. Rather, moral values are there to protect the spiritual essence of human beings and keep them from sin and spiritual pollution.

Haidt et al (1993), and Haidt and Hersh (2001) demonstrated that 'ethics of community' and 'ethics of divinity' are applied in moral reasoning in most cultures and Western subcultures, while only educated, secular Westerners focus heavily on Shweders' 'ethics of autonomy'. Because the part of the academical world that studies social justice is composed mainly of educated, secular Westerners, it is not strange that the consensus exists that the ethics that protects individual freedom is the only 'right' form of ethics, and that individuality, freedom and autonomy always prevail over values of group loyalty, tradition, authority and divinity. The latter values are often seen as repressive, as threats to individual freedom that need to be overcome (Haidt & Graham, 2007).

The legalisation of gay marriage is one of the main political issues that creates an abyss between liberals and conservatives. Conservatives view same-sex marriage as morally wrong because it is supposed to be detrimental to society, referring to arguments that relate to (3) Ingroup/Loyalty 'same sex marriage is a threat to traditional family constructs' (4) Authority/Respect: 'same sex marriage undermines the authority and the meaning of the institution of marriage' and (5) Purity/Sanctity: 'same sex marriage, and homosexuality in general, is not part of 'Gods plan' for humanity, it is a sin'. Liberals oppose the rejecting attitude of conservatives towards homosexuality on grounds of (1) Harm/Care: 'it is harmful and unjust to reject individual preferences and resist diversity based on religious prejudice' and (2) Fairness/Reciprocity: 'prohibiting same sex marriage creates inequality and social injustice because not all individuals are treated in the same way'.

Because liberals do not recognize the categories conservatives apply in their reasoning, the motivation behind conservative reasoning remains invisible to them. They are unable to see that the other categories could have moral relevance, and that these categories take precedence over high

values such as individual freedom and justice. Haidt argues that when liberals condemn conservative ideas for being delusional, or coming from a place of fear for change, they risk becoming 'politicocentric' i.e., liberals violate their own values such as the right to individual opinion, and it is 'a route to irrelevance' (Haidt & Graham, 2007) because it does not take down the wall between conservatives and liberals, it makes it even thicker and more opaque. To slowly change the attitude between liberals and conservatives, Haidt urges social justice researchers and activists in particular to use the foundations theory to step outside of their moral comfort zone and to try to imagine how the other foundations could make sense in an alternate narrative, so as to create a more open dialogue with those who directly oppose their viewpoints.

However, stepping outside one's own personal moral matrix, as Haidt urges academical liberals to do, is hard: imagine having to argue for positions opposite of what you hold. Would you be able to honestly examine the arguments made against gay marriage and abortion? And take seriously the arguments that are based on Ingroup/Loyalty or Purity/Sanctity? And give similar weight to these foundations?
According to Haidt (2006, p.14), the unease that comes with examining positions opposite of what we hold, are our 'moral emotions' taking the stage: "*It is as though our moral deliberations are structured by the sorts of invisible fences that keep suburban dogs from straying over property lines, giving them an electric shock each time they get too near a border*". This reaction of our affective mind is there to evaluate positions against the backdrop of an inner sense of justice, and in line with his intuitionist theory, Haidt believes that our opinions will eventually be settled based on this inner sense of justice rather than on a deduction of some kind (Haidt, 2006, p. 13).

Even though rationality is celebrated by the disciplines of moral psychology and philosophy, Frans de Waal agrees with Haidt by stating that "*when the push comes to the shove, we assign it little weight*" (de Waal, 2006, p.56). Because human morality is argued to have its roots in mammalian sociality, social emotions that are triggered by situations in the moral domain serve as compass, and these feelings are reflected in our eventual judgements.

With his foundation theory, Haidt suggests that there is a possibility to break down the invisible wall between people with different moral worlds. He urges liberals to look beyond their 'moral comfort zone' (Haidt & Graham, 2007, p. 111) that causes their beliefs, and to neutrally examine the categories that are valued by liberals to understand their points of view.

However, if our moral judgements are infused by strong emotions and intuitions, making sure our judgments stay within the structures of the domains we have learned to attribute value to, then how does Haidt assume it is possible for academic liberals to step outside this domain?

To recognize conservatives' concerns as moral concerns, rather than motivated social cognitions, they would need to train and cultivate link (6) in Haidt's social intuitionist model of moral judgement. This link, the *private reflection link*, can help them to 'dispassionately' evaluate their ideas about the importance of justice, equality and compassion and empathize with conservatives' values such as group cohesion, integrity or divinity. This is again connected to the outer layer of de Waal's model of empathy, as it requires the cognitively advanced capacity of perspective-taking: by actively empathising with others, rather than through active discussion, there is a possibility to learn to understand their points of view and sound arguments. Trying to compel someone to care about something in the same way as you do will requires much more than, for example, referring to justice and autonomy.

The main point of critique to Haidt's theories is the lack of importance that is attributed to reasoning, as intuitions are said to take precedence in making moral judgements (Horgan & Timmons, 2007; Pizarro & Bloom, 2003). But if Haidt himself urges others to take a position against their intuitions, then at least there must be some role attributed to intuition-altering processes in his theories. These processes must at the very least be marked as effortful as they go against one's initial incentives, and should involve active empathizing with others and conscious deliberation, either in an interpersonal or private context. In the next chapter I will illustrate how Haidt accommodates these types of reasoning in his theory.

# Chapter four: Rationalism

*"Apparently, admitting ignorance in response to a question, rather than being an indication of glibness and a low level of function, is a high-level cognitive ability, one that confabulators have lost"*
– William Hirstein (2004)

## *Reasons*

In the psychology of moral judgement, those who are considered rationalists make the empirical claim that people engage in deliberate cognitive processes to come up with (good) reasons, and that these reasons have a causal effect on the moral judgements they make. Hanno Sauer (2011) argues that the term 'reasons' can be subdivided in *motivating* reasons and *normative* reasons, the former being the type of reasons that bring a person to make a judgment and the latter the type of reasons that will ultimately justify that judgment (p.709).
Let's say I hold the judgement: 'eating meat is bad'. The reason that *motivated* me to hold this judgement is that the horrific circumstances in intensive farming repulse me and because I feel protective of these animals who cannot fairly defend themselves against such cruelty. The reason I will put forth in a discussion with a fellow philosophy student however, might be that eating meat is bad because intensive farming produces significant levels of greenhouse gas emissions, and that this has a detrimental effect on the climate. The latter argument will do a better job at *justifying* my statement because it does not merely depend on my personal convictions.
If there is a difference between these types of reasons, we can ask ourselves: are we motivated to hold certain beliefs because of the normative reasons that justify them, or do normative reasons play no such role and are our motivating reasons motivating ''all the way''? Intuitionists will affirm the latter and rationalists claim the former, arguing that our normative reasons have the potency to at least sometimes have a causal effect on the statements we hold.
This idea is reflected in the Effectiveness-Thesis proposed by Sauer (2011): "*The justifying (moral) reasons we have for our (moral) judgements figure in true causal explanations for why we hold these judgements.*" (p. 709), and is the most central claim of rationalist accounts of moral judgment in psychology (Kohlberg 1969; Piaget 1965), corresponding in philosophy to the idea of quandary ethics.

Jonathan Haidt's theory of quick flashes of intuition being the primary source of moral judgement diametrically opposes the Effectiveness-Thesis, and many of Haidt's critics stumble over the idea that conscious deliberation seems to be irrelevant to the cause of moral judgements in the social intuitionist model. (Pizarro & Bloom, 2003; Saltzstein and Kasachkoff 2004; Horgan & Timmons

2007). Taking it a step further, Haidt's critics argue that in this way, reasoning becomes a matter of post-hoc confabulation.

Confabulation is a clinical, pathological symptom that commonly occurs for example in patients with Korsakoffs' syndrome. When patients confabulate, they do not intend to deceive or lie but somehow have vivid memories of states of affairs that never actually occurred, and can be incredibly stubborn in their belief and in the defence of these previous events. According to Horgan and Timmons (2007) the same thing happens to the concept of reasoning in Haidt's theory: when intuition causes judgement and reasoning only occurs post hoc, the reasons people give for their judgements could have never played a causal role. In presenting these reasons to others, people confabulate about the strength of the decisive nature of these reasons. Because the quick, automatic process of forming a judgement is not accessible to the conscious mind and only its product appears in consciousness, people falsely believe that the normative reasons they make up afterwards to explain their judgments where also the reasons that motivated it, creating the illusion of conscious reasoning (Sauer, 2011, p.713). Just like patients with Korsakoff's syndrome, people who think they reached their judgement through reasoned steps are convinced of a certain chain of events that never really occurred.

According to Saltztein and Kasachkoff (2004), by denying an active role for conscious reasoning in moral judgement, Haidt treats the mind as "*solely a passive recipient of the social and physical environment"* (p.274). And if the mind is treated as a mere recipient, it is legitimate that Pizarro and Bloom raise questions about whether intuition alone is sufficient enough to solve complexities faced by modern humans (Pizarro & Bloom, 2006). Moreover, there is very little proof of which process takes precedence in moral judgment, the intuitions or prior reasoning, because research on the cause and rational quality of intuitions has just started (Sauer, 2011, p.718). So, as the social intuitionist model is not 'proven' all the way, there is room for improvement of the model, a fact that Haidt himself does not deny (Haidt, 2001; Haidt & Bjorklund, 2006).

The concept of reasoning in Haidt's theory differs from the concept of reasoning in rationalist theories, but does take on a more active role than the one that is presented by critics such as Saltzstein & Kasachkoff (2004). Rather than in the privately reflected, causal context rationalists position it in, reasoning in Haidt's theory either takes place in a social explanatory context, and only in specific cases in a private reflective context.

In our daily lives, we make 'simple' moral decisions based on intuition, our inner moral compass that is shaped by the moral emotions, which in turn are connected to the set of moral foundations that we have learned to value during development. Only when the moral compass is pointing in different

directions because it is set off by conflicting intuitions, we briefly become the rationalist judge searching for the truth amongst those intuitions (Haidt, 2001). If people do not agree with the decisions we made based on our intuitions, it is rather the triggering of new intuitions in the listener than presenting them with the product of deduction that persuades them to think otherwise. By learning the full range of foundations that are intuitively triggered in others when they make moral judgements, we can learn to dissect our own and reflect on the importance we attribute to our own values as opposed to others. These specific situations require active deliberation as is described by the (6) private reflection link in the Social Intuitionist Model.

So, rather than a discussion about the locus of reasoning in the process of moral judgement, the discussion between intuitionists and rationalists seems to be a matter of the weight that is attributed to reasoning as a decisive strength and how often we face situations in which this effortful, conscious mental endeavour is required.

In Haidt's theory, this is only the case when situations are very complex and multiple intuitions are triggered, or when we try to cognitively empathise with others that have intuitions we ourselves do not recognize or to which we attribute less significance.

Jonathan Haidt has worked with recent findings and theories from different disciplines such as primatology, developmental psychology and neuroscience to support his social intuitionist model, but does not claim that a similar case could not be made for rationalist' models. To combat fragmentation in the sciences and work towards a more complete understanding of the processes involved in human moral deliberation, Haidt (2001, p. 380) encourages academics to engage in interdisciplinary research:

"*The debate between rationalism and intuitionism, now over 200 years old, is not just a debate between specific models; it is a debate between perspectives on the human mind. All of the disciplines that study the mind should contribute to the debate*".

As philosophy has explored to great extend what the human mind is capable of in terms of cultivating the mind to ask and respond to fundamental questions, how can philosophy contribute to the debate and what does the debate contribute to philosophy? In the next and final chapter, I will answer these questions in my main argument, which is the answer to the central question of this research project.

# Chapter five: conclusion

## *Preface*

As part of completing my master studies in applied ethics, I took an elective in 2018 on primate morality. The course, offered by the biology department of the UU, consisted of three main lectures given by prof. dr. Frans de Waal. The first chapter of this thesis can be seen as an overview of the most important scientific results that where explored during the course. By reading additional articles and popular scientific literature by Frans de Waal, I learned how years of extensive research on the social nature of primates had led to his advocacy of a bottom-up approach to morality.
I can safely state that this course was the main source of inspiration for this thesis, as it led me on a path of exploring interdisciplinary research of a phenomenon which I had, until then, only actively learned to see through the lens of philosophy. The scientific results and empirical observations that indicated a continuation of capacities with other species sparked my long-held enthusiasm to put into perspective the assumptions of human uniqueness in moral philosophy, in particular because of the arguments it accommodates for those who wish to defend our extensive use of other animals or who deny any form of moral consideration towards other species. However, the inquiry into this research on morality gave me much more than yet another motivation to get on my animal ethics horse, as I experienced the intellectual benefits of the cross fertilization that is brought about by studying the same phenomenon from multiple disciplinary perspectives.

## *Answer to research question*

The aim of this thesis was to answer the question: *What does an interdisciplinary, bottom-up approach to morality attribute to the toolbox of the moral philosopher?*

What follows is a short overview of what has been discussed in the previous chapters and how this leads to a bottom-up approach to morality.
I argue that this view of morality is conducive to the applied ethicist because it (1) puts into perspective the methods and theories of her own discipline (2) gives insight into the processes involved in actual moral judgment (3) motivates to take on novel perspectives in meta-ethical discussions, and (4) emphasises the value of the philosopher's perspective in interdisciplinary research on morality.

What do we talk about when we talk about morality? The Stanford encyclopaedia of philosophy explains morality as "*certain codes of conduct put forward by a society or a group (such as a religion), or accepted by an individual for her own behavior*" (Bernard and Joshua, 2017) , Flack & de Waal (2000) describe it as "*the human sense of right and wrong used by society to promote pro-social behaviour"* .

But how and why did it come into existence? What does it rely on? An attempt to answer these questions is similar to what ancient Greek philosophers did when the searched for the 'arche' of things: Arche is the fundamental principle of a phenomenon, its original formation, and the search of what accounts for its continuing existence. This process can help to get a more complete understanding of its function.

It is probable that human morality does not have a single, unified account, because it is the result of specific adaptions through time that has varied between groups and environments, and relies on different social consensuses. However, just like primates, humans are social animals. Social animals live together in colonies, groups and societies and profit from collective action and collaboration. Group life also brings about conflict and distrust between members. Without a means to solve conflicts and trust issues, groups will eventually fall apart. Morality, existing in different forms of cognitive sophistication depending on the species, is, from a biologist's perspective, the function that creates the possibility for continued social cohesion.

Human morality is distinguished from primate morality by the capacity for language and reasoning, which introduced a top-down element to the bottom-up process, one that made moral discourse possible and, with that, an extension of group size beyond the scope of primate societies.

Haidt & Joseph (2006) studied the work of Frans de Waal to design the foundations theory for intuitive ethics, a theory that states that there are sets of basic moral intuitions that are shared across all moral systems. The human mind is sensitive to the emotions that trigger these foundations as they each served a specific evolutionary goal. Depending on which cultural framework we learn to appreciate during development, we automatically draw upon this moral information in later stages of life when we find ourselves in situations that initiate the process of moral judgement.

This process of making a judgement is described by the social intuitionist model, which central claim is that moral judgement is the result of quick flashes of these intuitions, and that reasoning occurs post-hoc either in a social context or in a private context, with the aim to defend those previously made judgements.

The main point of critique by rationalists on the model is that reasoning becomes a matter of post-hoc confabulation because conscious reasoning does not play the causal role that has long been attributed to it. This causal importance of reasoning is central to ideas of quandary ethics, a way of

looking at morality from a top-down perspective that, because of its practical applicability, has been popular in the sciences as a definition of morality.
However, instead of focussing on which process comes first, which is yet to be indicated by further scientific research, the debate between rationalists and intuitionists could also focus on the interplay between the intuitions and the reasons by approaching the discussion from a bottom-up perspective.

Rather than taking morality to be a human construct depending on reason only, scientific research can contribute to a more layered understanding of morality which, in turn, can be helpful to reach more of a common ground in ethical discussions that have led to increased polarization in the US and the Netherlands. Finding common ground and consensus amongst group members is yet another factor that sets human morality apart from primate morality (de Waal & Sherblom, 2018, p.255).

As our species evolved and learned how to apply and reflect on moral principles that in essence serve our need to be together, we can take the roots of empathic perspective taking as a means to understand moral complexities from multiple perspectives. Especially philosophers, who are trained to engage in modes of unnatural thought, are the perfect candidates to develop such a quality.
Virtue ethics, aimed at reaching a state of moral maturation that can be applied to the fluctuating moral situations of modern life, can assist in the development of mental flexibility by taking distance from primary judgements. These capacities are in turn necessary for empathic perspective taking.
In this way, primatology and biology can help to explain why morality is the way it is, and applying the 'ought implies can' argument, this creates the possibility to design moral frameworks and concepts that can be applied to in the real world, because: *'Is' and 'ought' are like the yin and yang of morality. We have both, we need both, they are not the same, yet they cannot be completely disentangled. They complement each other."* (de Waal 2014, p.186).
Moreover, psychological research can give the philosopher insight into the mental processes involved in moral deliberation. And as reason alone cannot motivate to act (Haidt, 2001) and intuition alone is insufficient to solve modern day complexities (Pizarro & Bloom, 2006), they need each other as well.

However, without the richness that an interdisciplinary perspective brings to the concept of morality and without the virtue/care ethical approach it advocates, the philosopher might thrust great value upon reason and logically compelling arguments in ethical debates, risking (1) a biased search in favour of arguments and values that support already held judgements within themselves or within academia, and (2) it might lead to loss of the compelling strength of arguments when they are applied in real life, as the persuasive strength of arguments might be less in reasoned statements and

more in empathically accessed convictions. Learning to see the range, effect and origination of moral beliefs and the processes involved in moral judgment by studying empirical scientific research on morality from multiple disciplines, the applied ethicist can create clarity in her philosophical toolbox by revising traditional, and adding new tools to it.

If, for example, the applied ethicist would learn during her studies how to recognize the foundations that underlie the judgements of others, she can obtain a more tailored approach to actual moral problems, acknowledging moral reasons that she might not even have recognised as 'moral' reasons from her own perspective.

Or, if for example the applied ethicist is aware of the affective importance of the intuitions and the moral emotions, she can more specifically structure an argument and try to incorporate both logically and intuitively compelling arguments.

Being aware of the social function of morality, it makes sense to cultivate one's character by means of virtue ethics and non-western ethical approaches. If morality is not only about rationally solving dilemma's between the competing interests of individuals, but about much more, namely: the strength of forming allies, being able to reconcile after fights, relying on each other for help and support, cooperating fairly and working together, than its function serves that sociality rather than the protection of individuality and autonomy.

In the 1980's, Alisdair MacIntyre's book *After Virtue* (1984) advocated such types of practice by arguing for a revival of Aristotelean ethics in contemporary moral discourse.

Virtue ethics' founding fathers, Aristotle and Plato, argued that moral rules were derived from virtues which in turn were derived from an understanding of *telos,* the final cause or goal in life. Being a morally 'good' person was to have a virtuous character, a complex, ideal mindset that allowed one to accept specific considerations such as courage, honesty or temperance, and have these as restraints or motivations for behaviour. According to Aristotle, a virtuous character could only be obtained over time, through training, practice, and life experience. Moral education involved an undergoing of life's difficult choices and accepting that there are no universal moral rules that apply to every situation at hand. Rather, through slow practice and maturation one gradually learned to cultivate a virtuous character from which good actions would flow naturally.

According to MacIntyre (1984, p.107) this type of ethics is much more solid and useful than the types of ethics our modern ethical discourse is built on. The failure of what he called '*the Enlightenment project*' (MacIntyre, 1984, p.46), the attempt to ground morality in abstract principles in order to disengage ethics from religion, resulted in incoherent moral vocabulary which had doomed the moral theories that developed during the Enlightenment from the start. Replacing ideas about a human

*telos* with an admiration of human rationality had caused a shift in focus: rather than moving closer to an ideal state like virtue ethics prescribed, morality began to rely on subjectively applied rules and ideas. The individual got granted all moral authority by philosophers, without a teleological character to fall back on. This gradually turned moral discourse into a subjective and highly irrational practice (MacIntyre, 1984, p. 67).

A bottom-up perspective to morality offers ideas about other things being of equal or more importance than values such as justice, autonomy and impartiality, and methods such as following universal rules, creating social contracts and calculating outcomes. Because academia has focussed so heavily on the latter, and research on the former has just started, it seems as though we can solve polarisation with sound reasons, but, viewed from an interdisciplinary perspective, this is not the case in the actual world: many other things like peer socialization, moral foundations, and quick intuitions influence what we think and why we think that.
If it is in our nature to be social rather than to be individual and autonomous, rule-based ethical theories might not be as helpful to cultivate morality as virtue-based theories are. Training in virtue ethics, by means of actual practice, can add to the cognitive flexibility of the philosopher and her openness to new and shifting perspectives.

The master in applied ethics offered by the philosophy department at the University of Utrecht, the master from which I aim to graduate with this thesis, offers an elaborate overview of the ethical theories most applied in relation to practical moral dilemmas. However, these theories often have a top-down character relying heavily on methods of quandary ethics and focussing on values of impartiality, autonomy and justice. A novel course that treats the concept of morality from multiple disciplines and offers knowledge of ethical traditions that follow from a bottom-up perspective to morality, like non-western/virtue and care ethical theories, I believe would both be helpful to the future ethicists as to academia as a whole. Combining approaches to ethics and broadening the scope of interdisciplinary research possibilities will lead to richer theories in all the fields that study the premises of morality.

# Message to take home

When do people change their minds?

In Louis Theroux' documentary "*Surviving America's most hated family* (2019) a portrait of the radical Westboro Baptist church known for their radical religious ideas and provocative demonstrations, he talks to one of the girls who has left the church and now lives a completely different life. As he asks her what made her make that decision to leave the church, she refers to twitter, and the people who, instead of reacting with hate, the reaction the religious movement thrives on, tirelessly took the effort to make her see different perspectives. There was some factual persuasion involved as they pointed out to her the logical fallacies of the church's beliefs, but without compassion and empathy those people would have never been able to release the tight grip the church had on her. This is an extreme cause, but fighting the 'hate-evokes hate mechanisms' and broadening the scope of what moral values people appreciate in their lives, whether we as the intellectual community approve of it or not, could be the remedy against the growing polarisation. By learning about Haidt's moral foundations theory I learned to realise that just because we do not recognize the reasons people give, or because in our world those reasons aren't valid or don't take precedence over reasons like the protection of autonomy and justice, it doesn't mean they are not there. And only by recognizing that they are there and examining them through active empathic perspective taking, we are able to get to the core of what motivates others, and by doing so we create the option for ourselves to design strategies to evoke novel intuitions in those who oppose our viewpoints.

If others, with their views, directly oppose my convictions in a way that might even be detrimental to the case, it becomes increasingly more challenging to cognitively empathize or '*trade places mentally*' (Flack & de Waal, 2000, p. 22) with them.

Because we react with hate so easily when hate is given to us, maybe because mirror neurons evoke similar mental states, the option to empathise with those who oppose our ideas becomes an impossibility. But by being aware of it, and through actively fighting that with intelligence, by nurturing our curiosity and openness to new experiences, we have a chance of overcoming the abyss. Social media is a brilliant place to do it: mirror neurons have no chance on screen as letters do not evoke the same emotional response as face to face interaction. And as we can shield behind the fence of our devices, we might be able to give our empathic intellect a chance to thrive.

Nina Marloes van den Berg
*Amsterdam*
Januari 2020

# Acknowledgements

For this thesis I am indebted to my supervisor *dr. Frans Stafleu* for the conversations that inspired the main topics of this thesis and for the endless support and trust I received, which was of inestimable value during the writing process, and to my good friend and philosophical ally *Harm-Jan Zondag* for his tireless motivation to engage in conversations about the topics of this thesis and his helpful comments and corrections to my first draft, and also to *prof. dr. Christopher Sinha*, for his constructive suggestions to the first version of this thesis, and the effort to give intelligible comments from a different discipline.

I thank my second reader *dr. Hanno Sauer* in advance for his time and effort to read and comment on this thesis.

# Literature


Bentham, J. (1999). *An Introduction to the Principles of Morals and Legislation*, Batoche Books. 1789.

Chaiken, S., & Trope, Y. (Eds.) (1999). *Dual process theories in social psychology*. New York: Guilford

Dekker, p. Den Ridder, J., & SCP. (2019). Burgersperspectieven 2019. 1. *Continu Onderzoek Burgerperspectieven.*

Descartes, R. (2009). *A Discourse on the Method of Rightly Conducting One's Reason and Seeking Truth in the Sciences.* The Floating Press.

Dimberg, U., Andréasson, P., & Thunberg, M. (2011). Emotional Empathy and Facial Reactions to Facial Expressions. *Journal of Psychophysiology*, *25*(1), 26–31.

Dunbar, R. I. M. (2004). Gossip in Evolutionary Perspective. *Review of General Psychology*, *8*(2), 100–110.

Hirstein, W. (2004). What is confabulation? In: *Brain fiction: Self-deception and the riddle of confabulation*. Retrieved from https://ebookcentral.proquest.com

Hobbes, T. (2009). *Leviathan: The matter, forme, and power of a common-wealth ecclesiastical and civill*. Retrieved from https://ebookcentral.proquest.com

Flack, J.C., and de Waal, F. B. M. (2000). 'Any Animal Whatever' Darwinian Building Blocks of Morality in Monkeys and Apes. *Journal of Consciousness Studies* ,7, No. 1-2, 2000, pp. 1-29

Gallup, GG. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology*, No 2, pp. 237-248.

Gallup, GG. (1983). Toward a comparative psychology of mind. *Animal Cognition and Behavior*, ed. RL. Mellgren, pp. 473–510.

Gert, Bernard and Gert, Joshua (2017) 'The Definition of Morality' *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.).

Gilligan, C. (1982). *In a different voice: psychological theory and women's development.* Cambridge, Massachusetts. Harvard University Press.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory. *Advances in Experimental Social Psychology*, 55–130.

Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? Journal of Personality and Social Psychology, 65, 613-628.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and reasons of liberals and conservatives. Journal of Applied Social Psychology, 31, 191-221.

Haidt, J., & Joseph, C. (2006). The moral mind: How 5 sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In: P. Carruthers, S. Laurence & S. Stich (Eds.), *The Innate Mind*, Vol. 3.

Haidt, J., & Graham, J. (2007). When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals may not Recognize. *Social Justice Research, 20*, 98-116.

Haidt, J. (2003). The emotional dog does learn new tricks: A reply to Pizarro and Bloom (2003). *Psychological Review*, *110* (1), 197–198.

Horgan, T., & Timmons, M. (2007). Morphological Rationalism and the Psychology of Moral Judgment. *Ethical Theory and Moral Practice* 10 (3):279-295.

Kant, I. (1996). *Anthropology From A Pragmatic Point of View.* Carbondale, III: Southern Illinois University Press. 1798.

Kant, I. (1998). *Groundwork of the metaphysics of morals*. Cambridge, U.K: Cambridge University Press. 1785.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago: Rand McNally

Norlock, Kathryn. (2019) Feminist Ethics. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.)

MacIntyre, A. C. (1984). *After virtue: A study in moral theory*. Notre Dame, Ind: University of Notre Dame Press.

Pincoffs, E. L. (1986). *Quandaries and virtues: Against reductivism in ethics*. Lawrence, Kansas: University of Kansas.

Pizarro, D. (2000). Nothing more than feelings? The role of emotions in moral judgment. *Journal for the Theory of Social Behaviour* 30 (4):355–375.

Preston SD, de Waal FBM. (2002). Empathy: its ultimate and proximate bases. *Behavioral Brain Sciences*. 25:1–72

Prochazkova, E., & Kret, M. E. (2017). Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neuroscience & Biobehavioral Reviews, 80,* pp. 99-114.

Rawls, J. (1971). *A theory of justice*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Ross, D. (1995). *Aristotle*. Retrieved from https://ebookcentral.proquest.com

Rozin, P., Haidt, J., & McCauley, C. R. (2000). Disgust. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions*, 2nd ed. (pp. 637-653). New York: Guilford Press.

Saltzstein, H. D., & Kasachkoff, T. (2004). Haidt's Moral Intuitionist Theory: A Psychological and Philosophical Critique. *Review of General Psychology*, *8*(4), 273–282.

Sauer, H. (2011). Social Intuitionism and the Psychology of Moral Reasoning. *Philosophy Compass* 6 (10), 708-721.

Shweder, R. A. (1990). Cultural psychology: What is it? In J. W. Stigler, R. A. Shweder & G. Herdt (Eds.), *Cultural psychology: Essays on comparative human development*, 1-43. New York: Cambridge University Press.

Shafer-Landau, R. (2012). *The fundamentals of ethics*. New York : Oxford University Press

Singer T, Seymour B, O'Doherty JP, Stephan KE, Dolan RJ, Frith CD. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature,* 439:466–69.

de Waal, F.B.M. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.

de Waal, F.B.M. (2006). Morally evolved: Primate social instincts, human morality and the rise and fall of "veneer theory." *Primates and Philosophers: How Morality Evolved.,* pp. 2-80.

de Waal, F. B. M. (2008). Putting the Altruism Back into Altruism: The Evolution of Empathy. *Annual Review of Psychology*, *59*(1), 279–300.

de Waal, F.B. M. (2014). Natural normativity: The 'is' and 'ought' of animal behavior. *Behaviour,* 151 (2-3), 185-204.

de Waal, F. B. M. & Sherblom, Stephen A. (2018). Bottom-up morality: The basis of human morality in our primate nature. *Journal of Moral Education, 47 (2), 248-258*

Webb, C. E., Woodford, P., & Huchard, E. (2019). Animal Ethics and Behavioral Science: An Overdue Discussion. *BioScience*, *69*(10), 778–788.