Moral Values, Behaviour, and the Self

An empirical and conceptual analysis

van den Berg, T.G.C.

DOI
10.4233/uuid:8cabdf81-f503-463f-89d7-7a2874d3f876

Publication date
2023

Document Version
Final published version

Citation (APA)
van den Berg, T. G. C. (2023). *Moral Values, Behaviour, and the Self: An empirical and conceptual analysis*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:8cabdf81-f503-463f-89d7-7a2874d3f876

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Moral Values, Behaviour, and the Self

An empirical and conceptual analysis

Moral Values, Behaviour, and the Self

An empirical and conceptual analysis

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus, prof. dr. ir. T. H. J. J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
dinsdag 9 mei 2023 om 12:30 uur

door

Tomas Gerardus Constantijn VAN DEN BERG
Master of Arts in Philosophy of Law and Governance
Master of Science in International Crimes and Criminology
Vrije Universiteit, Nederland
geboren te Beverwijk, Nederland

Dit proefschrift is goedgekeurd door de promotoren.

Samenstelling promotiecommissie bestaat uit:

Rector magnificus	voorzitter
Prof. dr. ir. C. G. Chorus	Technische Universiteit Delft, promotor
Dr. ir. M. Kroesen	Technische Universiteit Delft, promotor
Dr. mr. L. D. A. Corrias	Vrije Universiteit Amsterdam, copromotor

onafhankelijke leden:

Prof. dr. S. Roeser	Technische Universiteit Delft
Prof. dr. G. P. van Wee	Technische Universiteit Delft
Prof. dr. A.L. Smeulers	Rijksuniversiteit Groningen
Prof. dr. J. M. Halsema	Universiteit Leiden

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 724431)

Keywords: Moral Values, Moral Behaviour, Moral Self, Narrative Identity, Moral Foundations Theory, Moral Psychology, Phenomenology, Ricoeur, International Crimes

ISBN: 978-94-6483-127-6

Front and back cover drawings by Frits Smid

Cover layout by Ridderprint
Printed in the Netherlands by Ridderprint, www.ridderprint.nl

Contents

1. Introduction

Prelude

The Egg[1]

I bought an egg
The farmer said:
'This one has just been laid
I even left home a little late
To take it with me
Here is a young life,
About 16 cents a deal
And on behalf of the parents:
Have a good meal!'
.....
When the egg said
When the egg said
'Think that it is a boy that you are about to fry
Think that it is your brother who is sizzling in the pan
Think that paralyzed with fear he tries to grab the edge
But that he slips back into the butter
What then? What then?'
.....

Instead of frying the egg, the character in the above song by Dutch composer and singer Jaap Fischer decides to keep the egg and hatch it. Though this story sounds quite absurd, there is also an aspect to which we can quite easily relate, namely that the character is moved by moral considerations. Called to account by the other, he can no longer simply crack the egg, fry it and eat it. He reconsiders his initial plan out of consideration for another being and does what he thinks is the right thing to do. *It is exactly this idea, namely that people's moral considerations matter for their behaviour, that is the main topic of this thesis.*

[1] Fragment from "Het ei" (1961) by Jaap Fischer. Translated from Dutch to English by the author of this thesis.

1.1 Background of the research

1.1.1 The relevance of morality for decision-making and behaviour

In our daily life experience, it regularly seems the case that the consideration that *something is the right or wrong thing to do* plays an important role in making a certain decision. Moral considerations can play a role in quite simple daily life decision situations, such as standing up for an older person on the train, taking a Covid-19 rapid test before going to a party, calling a friend or family member who is going through a difficult time, or helping out a stranger who has lost their way. Decisions in which moral considerations play a role can also have a more structural character, such as the decision to stop eating meat or reduce one's flying frequency out of concern for the well-being of animals or the planet (Zeiske, 2021). Furthermore, societal upheaval and discussions in the media frequently attest to the idea that moral choices matter and that people can and should be moved by moral considerations. Think of the recent MeToo-cases on the work floor highlighted by the Dutch TV-programme BOOS ("Na onthullingen BOOS", 2022), or the moral upheaval around the so-called "Mondkapjes-affaire", where three Dutch entrepreneurs became millionaires through a face mask-deal with the Dutch government while maintaining towards the general public that the face masks were delivered on a non-profit basis for the benefit of 'our heroes of the health care system' (Volkskrant, 2022).

Also, with regard to the major crises we are facing today (e.g., a war within Europe, climate change, Covid-19 pandemic), we see in governmental policy strategies, as well as in strategies of non-governmental organizations, the presupposition that people's morality is relevant for behaviour. Both of these strategies, namely make a moral appeal to citizens to change their behaviour. For instance, Dutch prime minister Rutte has often referred to the health threat for grandmas and grandpas in an attempt to convince the Dutch youth to stick to the Corona measures (e.g., "Rutte tegen jongeren", 2020); Dutch animal well-being organization 'Wakker Dier' aims at behavioural change by confronting the public with horrific tales and pictures of the bio-industry (Wakker Dier, 2022); and citizens have been asked to turn down their heater, in order to become less dependent on Russian gas and not support Russia's aggressive and illegal war against Ukraine (Harvey, 2022; Rijksoverheid, 2022).

1.1.2 Raising empirical questions on morality and behaviour and their societal relevance

It should be clear that in our daily life choices and our institutions it is generally supposed that people's moral considerations matter for decision-making, which can range from simply frying an egg to adhering to a complete lockdown for weeks. Next to the classical normative and ethical questions on how we should behave that have been systematically studied within ethics, *this raises numerous empirical questions on the relation between morality and behaviour*. For instance: what is the influence of morality on behaviour? How does morality influence behaviour? What do people consider to be

part of their morality? What are the differences between people's individual moralities[2]? How can we measure people's individual morality and its influence on behaviour? What other (non-moral) factors play a role in our moral decision-making?

From a societal point of view, these empirical questions on morality and its role in decision-making and behaviour are particularly relevant to those who aim at behavioural change and the promotion of socially desirable behaviour, such as policymakers and interest groups. Many of these behaviours, if not all, have a strong moral dimension in the sense that moral considerations can play an important role in these behavioural choices. For policymakers and non-governmental organizations, triggering people's moral motivation can be a potentially fruitful source to move people towards enduring behavioural change (e.g., Steg, 2016). Given the major crises that the world is facing and the fact that many of these can only be solved by structural behavioural change, the need for a good understanding of how people's morality moves behaviour is rather pressing.

1.1.3 The empirical study of morality and behaviour: Kohlberg and the judgment-action gap

Given the relevance of moral considerations for many of our choices and their potential importance for behavioural influence, it is not surprising that, in the last few decades, these empirical questions about morality and behaviour have gained an increasing interest within the behavioural sciences (Chorus, 2015). A growing number of disciplines aiming at a better understanding of people's decision-making are taking moral factors into account, ranging from fields such as economics (Harsanyi, 1982; Andreoni & Miller, 2002) to consumer behaviour (Andorfer & Liebe, 2012; Arvola et al., 2008) to transport (Matthies et al., 2006; Chorus et al., 2018). Among these, the discipline that has most systematically empirically studied morality and moral behaviour, is psychology and, in particular, its subfield of *moral psychology* (see Alfano (2015) for an overview).

Here, Kohlberg's theory of moral development (Kohlberg, 1969, 1984) can be considered one of the major starting points. Kohlberg's theory is one of the first attempts to develop a systematic theory and measurement tool to empirically measure people's individual moral differences. It maintains that people differ in the extent to which their way of moral reasoning is developed. Accordingly, its method is directed at classifying people in one of the six increasingly elaborate reasoning stages that are identified. Though this theory received wide acclaim and has instigated a wide body of research, especially on the cognitive-developmental approach to morality (Lapsley, 2006, 1996), it has also received substantial criticism. It has decreased in influence in the last two to three decades (Lapsley and Narvaez, 2005). One of the problems that the theory faces is its link to behaviour. Though some evidence has been found for this connection (Rest & Narvaez, 1994), the moral reasoning stage measurements generally

[2] By individual morality, I understand the moral aspects that can be attributed to the individual, such as one's moral considerations, moral beliefs, moral values, moral norms, etc. These can be the object of empirical study, the measurements of which are used to understand, explain and predict certain attitudes and behaviour.

turn out to be rather inconsistent and weak predictors of morally relevant behaviours (Blasi, 1980). This issue has become known as the "*judgment-action gap"* (Walker, 2004).

Problems with Kohlberg's approach motivated new conceptualizations and measures of individual morality. One direction of research aiming to tackle the "judgment-action gap" has focused on the conceptualization and measurement of how important morality is to a person, named one's "moral identity" (Blasi, 1983; Lapsley & Narvaez, 2004; Jennings et al., 2015). Moral identity is within moral psychology defined as the extent to which moral notions such as moral considerations and moral values are part of one's personal identity or self-concept (Hardy & Carlo, 2011). An empirical measure for this construct has been developed by Aquino and Reed II (2002) and can be regarded as a general measure of how morally motivated people are. Though this concept of moral identity is generally seen as useful and an improvement with regard to only looking at moral reasoning to grasp the complexities of the moral person (Frimer & Walker, 2008), a recent meta-analysis shows that its predictive effect on moral behaviour is still rather weak and is 'not much stronger than the effects of other constructs, notably moral judgment or moral emotions.' (Hertz & Krettenauer, 2016, p. 136).

1.1.4 General moral values as a way of conceptualizing and measuring individual morality

Another way of conceptualizing and measuring individual morality that has become important in reaction to the Kohlbergian approach is defining a person's morality in terms of a limited set of general moral values. I will call these "*empirical moral value theories*". Conceptualizing and measuring morality in terms of people's moral values is not new, of course (see, for instance, Pittel and Mendelsohn (1966) for an early review), but it has relatively recently received new impetus with the development of Moral Foundations Theory (MFT) (Graham et al., 2013; Haidt & Joseph, 2004). *It is the conceptualization and measurement of individual morality in terms of general moral values and their relation to behaviour that is specifically investigated in this dissertation.*

Moral Foundations Theory (MFT) positions itself opposite to Kohlberg's rational and monist model of morality. Instead, MFT claims that our morality consists of a plurality of universal general moral values[3], the so-called moral foundations, which intuitively influence our decision-making. The five moral foundations that have been identified are: care/ harm, fairness/ cheating, loyalty/ betrayal, authority/ subversion, and sanctity/ degradation[4]. These universal moral foundations have an evolutionary origin and each develops as part of a person's morality to a certain extent during one's life, depending on a person's cultural and social upbringing and environment. The extent to which a moral foundation is developed reflects a sensitivity to situations,

[3] See for further elaboration on using the concept of general moral values regarding MFT Chapter 4, section 4.2.1, footnote 33.

[4] The authors state that they are confident about the existence of these five but that more may be discovered. Possible candidates are 'liberty/ oppression', equity/ undeservingness, and 'honesty/ lying' (Graham et al., 2018).

concepts, principles, beliefs, etc., belonging to a certain moral domain, resulting in intuitive moral judgments[5] (Graham et al., 2013).

The level of development of a moral foundation within an individual's morality is measured with the Moral Foundations Questionnaire (MFQ) (Graham et al., 2011; MoralFoundations.org, 2013). MFQ is a generic questionnaire that asks about the relevance of general moral considerations when making a moral judgment and one's agreeableness to certain moral statements, tapping into specific foundations, without referring to a specific context. Similar to moral identity measures, MFQ can be regarded as measuring the importance of moral notions for an individual, but now disaggregated into different moral domains, represented by a general moral value[6].

MFT has also given rise to further theorizing and scale development on general moral values. For instance, a recently developed empirical moral value theory is Morality As Cooperation theory (MAC) (Curry et al., 2019). MAC specifically grounds its theory in the idea that moral values have evolved as a set of solutions to problems of social cooperation. Accordingly, it identifies a different set of general moral values than MFT. These can be individually measured with the generic Morality As Cooperation Questionnaire (MAC-Q) [7].

1.1.5 The gap: the general moral value-moral behaviour relationship

Measuring people's general moral values has become a popular way to identify people's differing individual moralities and use them to investigate their influence on other variables -inside and outside of moral psychology. Moral foundations have been most extensively related to other attitudes, especially people's political orientation (e.g., Graham et al., 2009; Haidt & Graham, 2007; Koleva et al., 2012; Kim et al., 2012; Kivikangas et al., 2021) and socio-political attitudes (e.g., Dickinson et al., 2016; Kugler et al., 2014). Furthermore, general moral value measures, including MFQ, have been used in a variety of contexts to measure the influence of moral values on certain forms of behaviour. For example: behaviour in the workplace (Cohen et al., 2014), charity donations (Nilsson et al., 2016), sustainability behaviour (Vainio & Mäkiniemi, 2016), delinquency (Tarry & Emler, 2007), prosocial and antisocial behaviour in sport (Shields et al., 2018), covid-19 related behaviour (Díaz & Cova, 2022; Presti et al., 2021; Qian & Yahara, 2020), Facebook language (Kennedy et al., 2021), and creativity (Kapoor & Kaufman, 2022).

Although researchers use general moral value scales, such as MFQ, in their models as quick and easy instruments to investigate to what extent certain behaviours

[5] See for a more extensive and detailed elaboration and discussion of MFT and its Moral Foundations Questionnaire (MFQ) Chapter 3, sections 3.2 and 3.5, and Chapter 4, section 4.2.

[6] Paradoxically, though MFT builds upon Haidt's (2001) Social-Intuitionist model of moral judgment and, accordingly, explicitly states that moral judgments are formed through emotional and intuitive reactions, while it regards deliberate moral reasoning as a form of post-hoc confabulation, MFQ measures the development of these intuitions through a form of deliberate and conscious self-report. This implies that MFQ is not confined to intuitionist theories of morality, but can also be used to measure people's general moral values in the context of (more) rational theories, i.e., ones that do reserve a place for deliberate reasoning in the constitution of moral judgments.

[7] See for a description of this theory and scale Chapter 3, section 3.2 and 3.5.

can be attributed to someone's morality, *a systematic and broader insight into the relationship between general moral values and behaviour is lacking*. The research efforts with regard to behaviour have been rather scattered. The greater part of research on moral values has confined itself to mapping out people's priorization of moral values and, in addition, linking these to other attitudes (Graham, 2012). As Ellemers (2019) states, most empirical studies on individual morality investigate what is part of people's moral thinking rather than how these moral beliefs actually influence behaviour. Though people's moral thinking is interesting in itself, it becomes a lot more interesting if we also know how it affects behaviour. However, this link itself remains understudied. Concerning general moral values, it is at this point largely unknown to what extent the theory and methods of empirical moral value theories are suitable to conceptualize and measure people's individual morality and its influence on actual moral behaviours and decisions. This is problematic as the general moral value measures are promoted as useful ways and valid tools to capture people's morality. They are used accordingly in a broad range of disciplines to test morality's influence on different forms of behaviour.

1.2 Research objectives

1.2.1 The main objectives of this thesis

This thesis' main objective is *to better understand the link between general moral values and behaviour on a conceptual and empirical level*. *Secondly, a subsidiary aim is to make methodological suggestions following from my analysis to better capture individual morality and its influence on behaviour.* To reach these objectives, I chose an approach in which I selected the next study based on the results found and conclusions drawn from the former study. For my first study, I decided to investigate the link between general moral values and behaviour within the specific context of aggressive driving.

1.2.2 Chapter 2: general moral values and behaviour in the context of aggressive driving

In this first study, I explore the empirical and conceptual link between general moral values and aggressive driving behaviour. Driving behaviour can be regarded as a common and everyday form of behaviour with a clear moral component, as aggressive driving can have severe consequences for others. It, therefore, is a proper context to investigate the relationship between general moral values and everyday life decision-making. Furthermore, risky and aggressive driving behaviour has been linked to a number of psychological constructs, but the relationship with people's general moral values has not been investigated.

Conceptually, I extend Rest's (1986) and Schwartz's (2016) model of Ethical Decision-making to articulate the relationship between general moral values and aggressive driving. I argue that for a generally endorsed moral value to influence behaviour, it needs to go through several stages of the moral decision-making process (i.e., moral awareness, moral judgment, moral intention, and moral behaviour), each of which is influenced by many situational and individual factors. From this analysis I derive *that whether a certain moral value influences behaviour is rather contextual and*

this makes the empirical relationship between general moral values and aggressive driving indeterminate. The *empirical study* confirms this expected relationship. Here, I tested to what extent general moral values, operationalized by the Moral Foundations Questionnaire, are predictive of respondents' self-reported aggressive driving behaviour. *The results show only a few weak relationships between people's generally endorsed moral values and aggressive driving behaviour.*

Chapter 2 then concludes that the empirical relationship between general moral values and aggressive driving behaviour is very weak at best. It suggests that a possible explanation is the contextuality of the moral decision-making process. However, in this study, I only tested the empirical relationship for one form of behaviour and only used one sort of moral value scale. Also, the conceptual analysis was mainly applied to the specific context of aggressive driving. In the next study, I continue my investigation of the relationship between general moral values and behaviour by extending my approach.

1.2.3 Chapter 3: a broader empirical and conceptual investigation -general moral values are poor predictors of concrete behaviour

In Chapter 3, I address the above three knowledge gaps by broadening my investigation on the relation between general moral values and moral behaviour. In three empirical studies, *I used two different moral value scales* to measure people's general moral values (MFQ and MAC-Q) and related these *to a broad set of everyday life concrete moral behaviours* (i.e., voluntary work, informal care, conforming to Covid-19 measures, and eating meat). Also, in the third empirical study, I included more *specific moral values* tailored to the specific context of eating meat.

Furthermore, I elaborate on the conceptual framework linking general moral values to moral behaviour and deepen the understanding of the contextual aspects of the moral decision-making process. For instance, I make clear that to become aware of a certain moral value within a specific decision situation, it is crucial what specific meaning is given to the situation. Also, regarding the moral judgment and moral intention stage, it is rather hard to indicate the relative importance of moral values vis á vis each other and vis á vis other kinds of values in any general contextless terms. *Therefore, I expected that this contextuality of the moral decision-making process renders the predictive effect of general moral values on behaviour rather indeterminate*.

This expectation was confirmed in the three empirical studies. Here, *I found for both moral value scales across all studied forms of behaviour only weak to very weak effects*. The c context-specific moral value proved to be a somewhat better predictor than its more general counterpart and explained away the latter's initial effect on behaviour. The fact that minimal effects were found for general moral values and somewhat larger effects for more specific moral values, together with the articulated contextual aspects of moral decision-making, suggests that people's moral beliefs can influence decision-making but that context matters and that this needs to be reflected in the measuring method. Following my conceptual analysis, the influence of morality on behaviour seems to be more context-dependent and dynamic than general measures,

such as MFQ, can detect. I, therefore, propose that people's moral values should be measured within a delineated context of behaviour.

1.2.4 Moving towards a more fundamental level of analysis: the moral self

The empirical findings in chapters two and three indicate that general moral values are poor predictors of concrete moral behaviours. One possible improvement that follows from the presented conceptual model of moral decision-making is to include additional general individual and situational variables in the predictive model, e.g., general moral awareness or moral motivation measures. Though an interesting route to explore, I do not think that this will bring us much further to closing the gap between individual morality and behaviour, nor in getting a better understanding of the influence of morality on behaviour. At least two reasons can be given for this.

The first reason is that, given the generally weak relationships between general moral constructs and moral behaviour, I do not expect that adding more general moral constructs will lead to much better predictions. The second -and most important- reason is that my conceptual analysis in chapter 3 actually suggests that the problem of indeterminacy between general moral values and behaviour lies deeper than simply adding more general predictors can solve. As argued in the discussion of Chapter 3, my conceptual analysis points towards the fundamental contextuality of the moral decision-making process. For instance, in the awareness stage, the *specific meaning* given to a decision situation is crucial. With regard to moral judgment and moral intention, I argue that different general moral values can become important, which seems to depend on the *specific meanings* these receive in a certain context; this directly relates to another point, brought forward in chapter 3, namely the fact that a general moral value can lead to several specific interpretations within a decision situation, and thus to several differing specific moral judgments and behaviours. *These specific meanings influencing our decision-making are hard to grasp by general variables.*

This analysis then has important implications for the conception of our relationship with our moral values and, thus, for the relationship between moral values and behaviour. *It suggests that the relationship between moral values and behaviour is characterized by interpretation and is more dynamic than is presupposed by empirical moral value theories* with their method of capturing individual morality through a general questionnaire.

This idea, which is derived from my conceptual analysis based on and grounded in moral psychology, needs further articulation and substantiation, and this needs to be done on a more fundamental level. It poses the question of what our individual morality consists of when the idea of a stable cross-situational set of moral values determining our behaviour is challenged. So, how can we make sense of a moral person and their moral dispositions, such as one's general moral values, when these differ across contexts? And what does this imply for the empirical measurement of individual morality? *These questions directly relate to the question of moral identity or moral self*[8]. Notice that this concept is not primarily used in the sense of a concept of moral

[8] Throughout the thesis, I use the concepts of moral self and moral identity interchangeably.

motivation, as it is often referred to within moral psychology, but *rather as the way in which one understands oneself and others in terms of one and the same moral person*. The central question becomes what constitutes this unity and how it can incorporate the supposed interpretive and dynamic relationship with our general moral values?

1.2.5 Chapter 4: empirical moral value theories and the narrative moral self

In Chapter 4, I take up the above questions. I use insights *from the philosophical field of phenomenology* to propose a concept of moral identity that articulates and substantiates on a fundamental level the interpretive and dynamic relationship with moral values that was signalled in Chapter 3. The phenomenological approach is characterized by taking the first-person perspective of the experiencing and meaning giving subject, and, therefore, seems well equipped to bring forward this fundamental moral aspect of the human being.

More in particular, I turn to *Paul Ricoeur's phenomenological analysis of the moral person*, as expressed in his notion of *narrative identity* (Ricoeur, 1992). In a head-on confrontation with Moral Foundations Theory, I present Ricoeur's narrative identity as an improved concept of moral identity for the empirical study of morality. First, through an analysis of MFT's theory and questionnaire, I show that it assumes an essentialist concept of the moral self, consisting of stable moral traits. Then, I argue that this concept is unable to grasp the dynamical and context-sensitive aspects of the self, aspects that MFT's developers themselves do seem to assume in a more fundamental theoretical exposition of the theory. I submit that Ricoeur's narrative notion of identity, a self that reinterprets itself in every decision situation through self-narrative, is a viable alternative. It is able to incorporate the context-sensitive and dynamic relationship with one's moral values while maintaining a persisting moral identity. Finally, I suggest that accepting this narrative concept of moral self has implications for measuring individual morality with regard to predicting behaviour. It suggests that taking an explorative approach, including a qualitative research phase, within a delineated context is needed to grasp morality's influence on behaviour better.

1.2.6 Chapter 5: the narrative moral self in the context of international crimes

I conclude from the discussion in Chapter 4 on Ricoeur's narrative concept of moral self in relation to the moral self assumed by MFT, that the former turns out to be the superior concept when it comes to explaining moral decision-making. With this concept of moral self, I am able to fully expose the contextual aspect of moral decision-making, already signalled in the first two empirical chapters. It also leads to methodological suggestions for incorporating the contextual aspect of individual morality and its influence on behaviour. However, the discussion in Chapter 4 remains somewhat abstract and is confined to showing the usefulness of this concept of moral self to one specific sort of theory –empirical moral value theories. The question remains whether this notion of moral self can be further validated as a useful fundament for the empirical study of individual morality and behaviour by applying it to research fields studying specific forms of moral behaviour.

In Chapter 5, I further theoretically validate the narrative concept of moral self as a fundament for the relation between individual morality and behaviour within empirical study. Here, I investigate its relevance for a specific context of moral behaviour, namely the perpetration of international crimes[9]. I chose this extreme context because situational factors are regarded as particularly relevant for explaining this behaviour. Within this field, this generally dominance of situational explanations has led to a vivid discussion on the influence of personal dispositions versus situational factors. This discussion has constantly been lingering in the background in the earlier studies of this thesis, when aiming to measure the moral aspects of the individual while incorporating their contextual dependence. In this study, I show that the narrative concept of moral identity deepens our understanding of the transformation process by which ordinary men turn into perpetrators by exposing the narrative structure and the role of self-understanding within this transformation. I argue that it functions as a useful fundamental conception of the moral person for an interactionist approach towards explaining behaviour within the context of international crimes that transcends the simple product of general personal dispositions and situational factors. Finally, I suggest methodological directions specified for the study of this form of behaviour, following my analysis.

1.3 Outline of the thesis

The thesis follows the four studies described. In the discussion, I will present and reflect on the main conclusions and insights that can be derived from my investigations, which will be followed by methodological as well as policy implications that follow from these.

A final word on "The egg"

Now let me briefly return to "The egg", with which I started this introduction, to give a glimpse of my conclusion. At the end of the investigations in this thesis, we can see that the character's refusal to fry the egg should be primarily understood as a *refusal to understand himself* as someone killing a valuable living being for a meal. In this decision, the moral values he calls his own play a valuable role -*not* in directly determining a decision- but in guiding the recognition of himself as a saviour of this life in this situation. His conception of the bond with his brother, which developed during his life, may also have played an important role here. Unfortunately for the chick, the character is less forgiving towards its subsequent mockeries once hatched –i.e., when the chick starts laughing at the character for believing the "ridiculous" pleas made when still an egg. The following day, the chick ends up in a bowl of rice. Speaking of a dynamic morality…

[9] International crimes are crimes for which an individual can be held individually criminally responsible under *international* criminal law. These are genocide, crimes against humanity, war crimes, and the crime of aggression (Shaw, 2008).

References

Alfano, M. (2016). *Moral psychology: An introduction*. John Wiley & Sons.

Andorfer, V. A., & Liebe, U. (2012). Research on fair trade consumption—A review. *Journal of business ethics*, *106*(4), 415-435.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.

Arvola, A., Vassallo, M., Dean, M., Lampila, P., Saba, A., Lähteenmäki, L., & Shepherd, R. (2008). Predicting intentions to purchase organic food: The role of affective and moral attitudes in the Theory of Planned Behaviour. *Appetite*, *50*, 443–454.

Aquino, K., & Reed ll, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, *88*, 1–45.

Blasi, A. (1983). Moral cognition and moral action: A theoretical perspective. *Developmental Review*, *3*, 178–210.

Chorus, C. G. (2015). Models of moral decision making: Literature review and research agenda for discrete choice analysis. *Journal of Choice Modelling*, *16*, 69–85

Chorus, C. G., Pudāne, B., Mouter, N., & Campbell, D. (2018). Taboo trade-off aversion: A discrete choice model and empirical analysis. Journal of choice modelling, 27, 37-49.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology*, *107*(5), 943–963. https://doi.org/10.1037/a0037245

Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008

Díaz, R., & Cova, F. (2022). Reactance, morality, and disgust: the relationship between affective dispositions and compliance with official health recommendations during the COVID-19 pandemic. *Cognition and Emotion*, *36*(1), 120–136. https://doi.org/10.1080/02699931.2021.1941783

Dickinson, J. L., McLeod, P., Bloomfield, R., & Allred, S. (2016). Which Moral Foundations Predict Willingness to Make Lifestyle Changes to Avert Climate Change in the USA? *PLOS ONE*, *11*(10), 1–11. https://doi.org/10.1371/journal.pone.0163852

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017. *Personality and Social Psychology Review*, *23*(4), 332–366. https://doi.org/10.1177/1088868318811759

Fischer, J. (1961). Het ei [Song]. On *Om je geld.* Studenten Grammofoonplaten Industrie.

Frimer, J. A., & Walker, L. J. (2008). Towards a new paradigm of moral personhood. *Journal of Moral Education*, *37*(3), 333–356. https://doi.org/10.1080/03057240802227494

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental Social Psychology, 47*, 55-130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4407236-7.00002-4

Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitch, C., & Mooijman, M. (2018). Moral foundations theory. *Atlas of moral psychology*, 211-222.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and Conservatives Rely on Different Sets of Moral Foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Meindl, P., & Beall, E. (2012). Integrating the Streams of Morality Research: The Case of Political Ideology. *Current Directions in Psychological Science*, *21*(6), 373–377. https://doi.org/10.1177/0963721412456842

Graham, J., Nosek, B., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385.

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., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55-66.

Hardy, S. A., & Carlo, G. (2011). Moral identity: What is it, how does it develop, and is it linked to moral action?. *Child development perspectives*, *5*(3), 212-218.

Harsanyi, J. C. (1982). Utilitarianism and beyond. In A. Sen & B. Williams (Eds.), *Morality and the theory of rational behaviour* (pp. 39-62). Cambridge: Cambridge University Press.

Harvey, F. (2022, March 3). Turn down heating by 1c to reduce need for Russians import, Europeans told. *The Guardian*. https://www.theguardian.com/environment/2022/mar/03/turn-down-heating-reduce-need-russian-imports-europeans-told

Hertz, S. G., & Krettenauer, T. (2016). Does moral identity effectively predict moral behavior?: A meta-analysis. *Review of General Psychology*, *20*(2), 129-140.

Jennings, P. L., Mitchell, M. S., & Hannah, S. T. (2015). The moral self: A review and integration of the literature. *Journal of Organizational Behavior*, *36*(S1), S104-S168.

Kapoor, H., & Kaufman, J. C. (2022). Unbound: The relationship among creativity, moral foundations, and dark personality. *The Journal of Creative Behavior*, *56*(2), 182-193.

Kennedy, B., Atari, M., Davani, A. M., Hoover, J., Omrani, A., Graham, J., & Dehghani, M. (2021). Moral concerns are differentially observable in language. *Cognition*, *212*, 104696.

Kim, K. R., Kang, J. S., & Yun, S. (2012). Moral intuitions and political orientation: Similarities and differences between Korea and the United States. *Psychological Reports*, *111*(1), 173–185.

Kivikangas, J. M., Fernández-Castilla, B., Järvelä, S., Ravaja, N., & Lönnqvist, J.-E. (2021). Moral foundations and political orientation: Systematic review and meta-analysis. *Psychological Bulletin, 147*(1), 55–94. https://doi.org/10.1037/bul0000308

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In Goslin, D. A. (ed.), *Handbook of Socialization Theory and Research (*pp. 347-480). Chicago, IL, Rand McNally.

Kohlberg, L. (1984). *The Psychology of Moral Development*. Cambridge, MA: Harper and Row.

Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the threads: How five moral concerns (especially purity) help explain culture war attitudes. *Journal of Research in Personality*, *46*(2), 184–194.

Kugler, M., Jost, J. T., & Noorbaloochi, S. (2014). Another Look at Moral Foundations Theory: Do Authoritarianism and Social Dominance Orientation Explain Liberal-Conservative Differences in "Moral" Intuitions? *Social Justice Research*, *27*(4), 413–431. https://doi.org/10.1007/s11211-014-0223-5

Lapsley, D. K. (1996). *Moral psychology*. Boulder, CO: Westview Press.

Lapsley, D. K. (2006). Moral stage theory. In M. Killen & J.G. Smetana (Eds.), *Handbook of moral development* (pp. 55-84). Psychology Press.

Lapsley, D., & Narvaez, D. (2004). *Moral development, self, and identity*. Mahwah, NJ: LEA.

Lapsley, D. K., & Narvaez, D. (2005). Moral psychology at the crossroads. *Character psychology and character education*, 18-35.

Matthies, E., Klöckner, C. A., & Preissner, C. L. (2006). Applying a Modified Moral Decision Making Model to Change Habitual Car Use: How Can Commitment be Effective? *Applied Psychology*, *55*(1), 91–106.

MoralFoundations.org. (2013). Retrieved December 20, 2023, from https://moralfoundations.org

Na onthullingen BOOS: 'Alsof er een leger is opgestaan'. (2022, January 22). Nieuwsuur. https://nos.nl/nieuwsuur/artikel/2414010-na-onthullingen-boos-alsof-er-een-leger-is-opgestaan

Nilsson, A., Erlandsson, A., & Västfjäll, D. (2016). The congruency between moral foundations and intentions to donate, self-reported donations, and actual donations to charity. *Journal of Research in Personality*, *65*, 22–29. https://doi.org/10.1016/j.jrp.2016.07.001

Pittel, S. M., & Mendelsohn, G. A. (1966). Measurement of moral values: A review and critique. *Psychological Bulletin*, *66*(1), 22.

Presti, S. Lo, Mattavelli, G., Canessa, N., & Gianelli, C. (2021). Psychological precursors of individual differences in COVID-19 lockdown adherence: Moderated-moderation by personality and moral cognition measures. *Personality and Individual Differences*, *182*, 111090.

Qian, K., & Yahara, T. (2020). Mentality and behavior in COVID-19 emergency status in Japan: Influence of personality, morality and ideology. *PloS One*, *15*(7), e0235883.

Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.

Rest, J., & Narvaez, D. (Eds.). (1994). *Moral development in the professions: Psychology and applied ethics*. Hillsdale, NJ: Erlbaum.

Ricoeur, P. (1992). *Oneself as another*. Chicago: The University of Chicago Press.

Rijksoverheid (2022, December 17). *Wat kunnen huishoudens en bedrijven doen om minder energie te gebruiken?* https://www.rijksoverheid.nl/onderwerpen/gas/minder-energie-gebruiken-voor-huishoudens-en-bedrijven

Rutte tegen jongeren: Ga niet naar ouderen als je je niet aan regels houdt. (2020, August 6). Nu.nl. https://www.nu.nl/coronavirus/6069241/rutte-tegen-jongeren-ga-niet-naar-ouderen-als-je-je-niet-aan-regels-houdt.html

Schwartz, M. S. (2016). Ethical Decision-Making Theory: An Integrated Approach. *Journal of Business Ethics*, *139*(4), 755–776. https://doi.org/10.1007/s10551-015-2886-8

Shields, D. L., Funk, C. D., & Bredemeier, B. L. (2018). Relationships among moral and contesting variables and prosocial and antisocial behavior in sport. *Journal of Moral Education*, *47*(1), 17-33.

Steg, L. (2016). Values, norms, and intrinsic motivation to act proenvironmentally. *Annual Review of Environment and Resources*, *41*, 277-292.

Tarry, H., & Emler, N. (2007). Attitudes, values and moral reasoning as predictors of delinquency.

British Journal of Developmental Psychology, *25*(2), 169–183. https://doi.org/10.1348/026151006X113671

Vainio, A., & Mäkiniemi, J. P. (2016). How Are Moral Foundations Associated with Climate-Friendly Consumption? *Journal of Agricultural and Environmental Ethics*, *29*(2), 265–283. https://doi.org/10.1007/s10806-016-9601-3

Volkskrant (2022, December 12). *Sywert van Lienden en de Mondkapjesdeal: al ons nieuws en achtergronden rond de mondkapjesdeal van Sywert van Lienden en de overheid*. https://www.volkskrant.nl/dossier/sywert-van-lienden-en-de-modkapjesdeal

Zeiske, N. (2021). The Intrinsic Route to Pro-Environmental Behaviour. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. https://doi.org/10.33612/diss.180931035

Wakker Dier. (2022, December 17). Wakkerdier.nl

Walker, L. (2004). Gus in the gap: Bridging the judgment–action gap in moral functioning. In D. Lapsley, & D. Narvaez (Eds.), *Moral development, self and identity* (pp. 1–20). Mahwah: LEA.

2. Does morality predict aggressive driving?

A Conceptual Analysis and Exploratory Empirical Investigation

This chapter has been published as a journal article: van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2020). Does morality predict aggressive driving? A conceptual analysis and exploratory empirical investigation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *74*, 259–271. https://doi.org/10.1016/j.trf.2020.08.017

Abstract

Risky and aggressive driving is an important cause of traffic casualties and as such a major health and cost problem to society. Given the consequences for others, risky and aggressive driving has a clear moral component. Surprisingly, however, there has been little research on the relation between morality and risky and aggressive driving behaviour. In this study we aim at addressing this gap. First, we present a conceptual analysis of the relationship between moral values and aggressive driving behaviour. For this purpose, we extend Schwartz's Integrated model of Ethical Decision Making and apply it to the context of aggressive driving. This conceptual analysis shows that moral decision-making processes consist of several stages, like moral awareness, moral judgment and moral intent, each of which are influenced by individual and situational factors and all of which need to materialize before someone's generally endorsed moral value affects concrete behaviour. This suggests that the moral value-aggressive driving relationship is rather indeterminate. This conceptual picture is confirmed by our empirical investigation, which tests to what extent respondents' moral values, measured through the Moral Foundations Questionnaire, are predictive of respondents' aggressive driving behaviour, as measured through an aggressive driving behaviour scale. Our results show few and rather weak empirical relationships between moral values and committed aggressive driving behaviours, as was expected in light of our conceptual analysis. We derive several policy implications from these results.

Keywords: **aggressive and risky driving; moral values; Integrated Model of Ethical Decision Making; Moral Foundations Theory; Aggressive Driving Behavior Scale**

2.1 Introduction

Risky and aggressive driving behaviour[10] causes major health and cost problems to society. Worldwide 1.35 million people die from road accidents annually. Moreover, it is the number one cause of death for children and young adults (WHO, 2018). Also, in Europe traffic accidents are still a major problem and reach much higher numbers than targeted by EU-policy. After a sharp decline in traffic fatalities since the beginning of the century, the decrease has effectively stagnated during the past five years, rendering the EU target for reducing traffic deaths by 50% between 2010 and 2020 far out of reach (Adminaité-Fodor et al., 2019). In the Netherlands, after a few decades of decline, the number of traffic victims has even gone up again in the last couple of years, counting well above 600 deaths per year (CBS, 2019a). Research shows that the behaviour of the driver plays a significant role in the vast majority of car accidents (Evans, 1993). In particular, risky driving behaviour, like ignoring basic travel rules and speeding, is linked to involvement in car accidents (Fergusson et al., 2003; Iversen & Rundmo, 2002). As such, a better understanding of this harmful behaviour is of vital importance for the design of policies and measures to prevent accidents and their harmful consequences.

In the last two decades a vast body of literature within psychology and travel behaviour research has taken up the task of better understanding risky and aggressive driving behaviour, by studying individual factors that affect this type of behaviour. Two streams of literature have been dominant within this research (Ulleberg & Rundmo, 2003). The first focuses on personality traits as predictors of risky and aggressive driving behaviour. This research has revealed weak but stable relations between some personality traits, like sensation seeking, anger and normlessness, and potentially harmful driving behaviour (Yang et al., 2013; Nordfjaern et al., 2010; Iversen & Rundmo, 2002). Personality traits are therefore generally considered as distal influencers of this kind of behaviour (Mallia et al., 2015; Ulleberg & Rundmo, 2003).

A second stream of literature has directed its attention towards so-called social cognitive factors, like perceptions, beliefs, and attitudes, which presumably influence risky and aggressive driving behaviour more directly (Ulleberg & Rundmo, 2003). To capture the influences of these factors a frequently used theoretical framework is the theory of planned behaviour (Ajzen, 1988), which posits that a person's behaviour is determined by one's intention to perform the behaviour. In turn, someone's intention is determined by one's attitude towards the behaviour, one's perception of control over the behaviour, and one's perception of the social norms involved with the behaviour (Ajzen, 1988). Several empirical studies have established significant relationships between these three components of the theory of planned behaviour and the intention to perform risky driving behaviour (Parker et al., 1996).

[10] Following Houston et al. (2003, p. 270) their definition, we define risky and aggressive driving behaviour in this study as *'a pattern of unsafe driving behavior that puts the driver and/or others at risk... and that others perceive as potentially aggressive and harmful'*. Note that this definition does not include an intentional or cognitive element, i.e. it is not required that the driver him or herself also had the intention to drive aggressively or harmful for the behaviour to fall within this definition and the scope of this study.

Inspired by both approaches Ulleberg and Rundmo (2003) integrated personality factors and social cognitive factors into a single model, in which personality influences risky driving behaviour directly as well as indirectly through its effect on attitudes. This integrated model has consequently been adopted to explain the risky driving behaviour of different particular groups of drivers, like young drivers (Machin & Sankey, 2008) and professional bus drivers (Mallia et al., 2015).

Though much work has been done in psychology and travel behaviour research to explain individual differences in risky and aggressive driving behaviour, an important area that seems particularly relevant to get a better understanding of this kind of behaviour has been largely overlooked: *moral* psychology. The field of moral psychology looks into the moral dimension of the human being, like the moral values and beliefs that people have, and their influence on people's attitudes and behaviour in moral contexts. The fact that moral considerations can play an important role in the constitution of many kinds of behaviour has a long pedigree and is in recent years increasingly being recognized by a growing number of disciplines that aim at a better understanding of social behaviour, ranging from economics (Harsanyi, 1982; Andreoni & Miller, 2002) to consumer behaviour (Andorfer & Liebe, 2012; Arvola et al., 2008) to transport (Matthies et al., 2006; Chorus et al., 2018).

(Unsafe) driving behaviour seems to be an obvious form of social behaviour in which moral considerations may play a role. Traffic violations and risky or aggressive driving behaviour have a clear moral component, as it can produce great harm to oneself, others, and society as a whole (Parker et al., 1995). Considerations of wrong and right thus potentially play a role in drivers' decision to (not) perform such kinds of risky behaviour. In this light it is surprising to see the earlier mentioned lack of attention from a moral psychological perspective.[11] Next to personality and social cognitive factors, people's moral values and attitudes seem to be a relevant factor to investigate in order to get a further understanding and explanation of aggressive driving behaviour. In this study we aim to contribute to addressing this gap by conducting a conceptual analysis of the relationship between moral values and aggressive driving behaviour and subsequently investigate this relationship empirically.

[11] Morality within a driving context has been studied increasingly in the last decade in relation to the moral questions and, more specifically, hypothetical moral dilemmas, that the development of autonomous vehicles raise (i.e. trolley problems) (e.g., Li et al., 2019; Awad et al., 2018). These investigations often have a more normative character (what is the right decision when an automated vehicle has to choose between hitting a granny or a child?) or are focused on people's decisions in such hypothetical situations, rather than the role of moral values in actual moral decision-making by drivers, which is the focus of this study.

2.2 Earlier work on morality and aggressive driving and specification of the knowledge gap

So far, to the best of our knowledge, only two studies have explicitly focused on the relation between morality and risky and aggressive driving. One study directly measured the relationship between people's general morality and driving behaviour (Bianchi & Summala, 2002). This study, conducted among Brazilian students, used a measure of moral judgment that reflected Kohlberg's (1969) stages of moral development. The results revealed no significant relationship between the development stage of moral judgment and aggressive driving violations. According to the authors, this result was probably due to the small variance of the moral judgment score in the sample.

Another study, conducted by Parker et al. (1995), did find a relationship between a moral concept and the intention to commit a driving violation. The study tested if an extended model of the theory of planned behaviour, including a fourth predictor variable of 'personal norm', would yield better predictions than the original model. Here, a relationship was found between a person's moral attitude towards a certain driving violation and the intention to commit such a violation. This is an interesting finding that at least suggests the relevance of moral convictions for explaining driving violations.

Since this earlier work on this subject, however, major developments have taken place within the field of moral psychology. Importantly, a new and highly influential theory of morality has been introduced, the so-called Moral Foundations Theory (MFT) (Graham et al., 2013; Haidt & Joseph, 2004), which is considered as a reaction to and alternative for Kohlberg's theory of moral development (Graham et al., 2013; Haidt, 2007). MFT maintains that human morality consists of at least five different universal and irreducible moral values that developed during human evolution and which thus all people possess to some extent, the extent to which depending on social and cultural influences. These values are considered our moral foundations, producing the moral intuitions by which we judge wrong from right. The five moral foundations that have so far been identified are: care/harm, fairness/cheating, loyalty/betrayal, authority/subversion and sanctity/degradation (Graham et al., 2013). For empirical analyses, MFT is operationalized in the Moral Foundations Questionnaire (MFQ) (see MoralFoundations.org, 2013), which is an extensively validated scale to measure each of the five foundations (Graham et al., 2011).

MFT and MFQ have become widely used in the last decade, also outside its field of origin, like in economics and the political sciences, to measure what people find morally important and to explain different phenomena in terms of people's moral values and beliefs. It has, for instance, been used to explain political and ideological orientation (Graham et al., 2009), climate friendliness (Vainio & Makiniemi, 2016), and charity donations (Nilsson et al., 2016). When investigating the influence of moral values and considerations on risky and aggressive driving behaviour MFT/MFQ therefore seems to be the obvious choice of measure.

The present study aims at getting a better understanding of the role that moral factors play in the constitution of risky and aggressive driving behaviour. Building upon the moral framework of MFT we test to what extent individual differences in morality can predict *committed* risky and aggressive driving behaviour. In order to do so, we proceed to answer the following research question:

'*To what extent is there a relationship between an individual's endorsement of the different moral foundations of MFT and her or his committed risky and aggressive driving behaviour?*'

However, before embarking on this empirical research question, it is important to conceptually clarify the assumed psychological processes through which moral values may influence aggressive driving behaviour; this will also enhance our understanding of the empirical findings to be discussed further on. For this conceptual analysis, we use and extent Schwartz's (2016) Integrated Ethical Decision-Making model.

2.3 A theoretical conceptualization of the moral value-aggressive driving behaviour relationship

2.3.1 Models of ethical decision-making

During the last three decades moral psychology has investigated the moral decision process and developed models of ethical decision-making[12]. Ethical decision-making models aim to clarify which behavioural and situational factors and processes play a role when an individual is confronted with a moral decision situation, which may be defined broadly as a situation in which moral considerations can potentially play a role in decision-making. As such, these models seem well equipped to explicate the relationship between moral values and concrete behaviour with a potential moral dimension, such as aggressive driving. Although much of the work done in this field is related to moral decision-making within an organizational environment, its findings are not confined to this context. In fact, one of the founders and main contributors to the field of ethical decision-making intended to develop a general model applying to a broad range of contexts (Rest, 1986).

The ethical decision-making literature can be divided into two main conflicting streams: one focused on developing rational models and the other on non-rational models (Schwartz, 2016). Rational models proceed from the idea that people's moral decision-making is essentially based on moral reasoning resulting in a moral judgment (e.g. Rest, 1986; Trevino, 1986; Jones 1991). Non-rational models, in contrast, assume that people's moral judgments are not the result of moral reasoning but are rather the results of intuitions and emotions that directly produce a moral judgment. Moral reasoning is then assumed to take place after the fact as an *ex post facto* process that has no direct influence on the moral judgment itself; it only functions as a

[12] Note that these are *descriptive* models of ethical decision-making rather than *normative* models determining how one *should* behave.

rationalization of the moral judgment that was already produced through intuitive and emotional processes (Haidt, 2001).

For our study it may seem logical to follow the non-rational model to further explicate the moral value-aggressive driving relationship in more detail. This is because the instrument that we use to measure moral values, the MFQ, is theoretically grounded in Haidt's non-rational social-intuitionist model (Graham et al., 2013). However, the use of MFQ itself and the questions that are asked do not seem to strictly exclude a rational model. It seems perfectly justifiable for researchers within both the rational and the non-rational stream to use MFQ as a measure of people's moral values and relate it to behaviour. Luckily, we do not have to choose. Schwartz (2016) developed an integrated model of ethical decision making (I-EDM), based on models such as those of Rest (1986), Jones (1991), Haidt (2001), and others, by incorporating these rational and non-rational perspectives. In the following, this integrated model is discussed and further elaborated on, in order to get a better conceptual understanding of the processes that play a role in moral decision-making and thereby explicate the moral value-aggressive driving relationship.

2.3.2 A conceptualization of the moral value-moral behaviour relationship in the context of aggressive driving

Figure 1 presents a modified version of Schwartz's (2016) I-EDM model, one which is simplified and focused on aggressive driving behaviour.

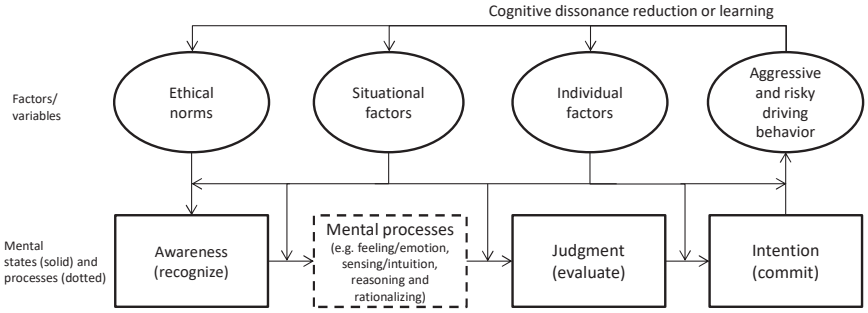

Figure 1. Modified version of Schwartz's (2016) Integrated Ethical Decision Making model

The process of moral decision-making[13] potentially starts off in a context in which an ethical issue can arise. As we are interested in explicating the individual moral value-moral behaviour relationship in the context of aggressive driving, we can define this starting point more explicitly. Our starting point then is an individual with a certain moral foundations make-up who, while driving a car, makes a decision (rationally or intuitively) about whether to perform an aggressive driving behaviour act or not. As moral considerations potentially play a role in this decision situation, the person could get involved in a moral decision-making process.

[13] Several authors within the ethical decision-making literature indicate to use the words 'ethical' and 'moral' interchangeably (Jones, 1991; Schwartz, 2016). The process of *ethical* decision-making that they describe can thus also be understood as the process of *moral* decision-making. We will mainly use the latter when referring to the decision process in which moral considerations potentially play a role.

Based on Schwartz's (2016) model it can be derived that, in order for a person's endorsed moral values to affect the decision whether to perform an aggressive driving behaviour act, the process of moral decision-making has to fully materialize, meaning that the person needs to go through the following four process stages[14]: for a person to actually get involved in a process of moral decision-making he or she first needs to become *aware (1)* of the fact that a moral value could be at stake and that a moral consideration plays a role in the decision situation. Once aware, the person makes a *moral judgment (2)* about the right course of action that should be taken, either through a rational or non-rational process. This is followed (or not) by establishing the *intent (3)* to follow up on this moral judgment, and finally this leads to according *behaviour (4)*. Schwartz (2016) states that the first three process stages are regarded as mental states, the fourth as an active conduct.

In Schwartz's (2016) model, each of these four stages within the process of moral decision-making are affected by two sorts of factors: individual and situational factors. Individual factors are captured by the broad construct of 'moral capacity'. It is one's level of ethical maturity, or moral development, as well as to what extent one follows through on one's moral judgment when faced with pressures to deviate from it. Other individual factors, like demographics or personality etc., are regarded as affecting someone's personal moral capacity and thus only indirectly affect the process of moral decision-making. Situational factors are characteristics of the situational context that influence the different stages of the decision-making process. These characteristics have to do with the specific issue that is dealt with, like its complexity or its moral intensity; with the (organizational) environment, like the ethical infrastructure and culture; and with someone's personal situation, like someone's changing ethical vulnerability (think of someone's changing financial situation) (Schwartz, 2016). In the following we will go deeper into the four different process stages and the influence of individual and situational factors. Since becoming aware of the relevant moral meaning of a decision situation is the crucial first step for the link between moral values and moral behaviour to be activated at all, we will first elaborate on this stage and then discuss the other stages more briefly.

Awareness

The first process stage of "awareness" represents the fact that before a moral value can influence the decision process it is first necessary that the person becomes aware of the fact that a moral value is possibly at stake within a decision situation. Moral awareness comes with the interpretation of the decision situation, which involves imagining the possible courses of actions *and* its consequences for the wellbeing of (oneself and) others and to determine possible violations of moral principles and duties (Rest, 1986). This leads to awareness of the *moral meaning* of a situation. As Jones (1991) states, when someone is not aware that the situation he or she is dealing with is a moral situation, the moral decision framework is not activated. Rather, this person may deal with the decision situation using a different framework, like one based on economic

[14] Here Schwartz follows Rest's (1986) four component model of ethical decision-making.

transactions. It should be understood though, that in reality people might not think about whether they are facing a 'moral situation' in such an abstract sense. Instead, people may or may not realize that one or more of their endorsed moral values, or more concrete moral standards based on these values, are in jeopardy within a situation they are confronted with and which they have interpreted in a certain way. In both the rational as well as the non-rational model this awareness stage is relevant and can be distinguished from the other stages for analytical purposes. But note that, for the non-rational models, it seems that this stage and the moral judgment stage practically arise simultaneously as it is assumed that the judgment is produced by a direct intuition.[15]

It may well be that a person does *not* become aware of the fact that a moral value that he or she finds important in general is potentially at stake within a certain decision situation. Though not mentioned by Schwartz (2016), it seems that at least three different general reasons can be given for not realizing a relevant moral meaning of the situation. First, a person may not know or does not believe the correct facts or lacks crucial information. For example, someone might not know or believe that speeding significantly increases the risk of making a collision (instead, the person might think that he or she has everything under control).

Secondly, one may know the material facts, yet somehow miss or not realize a possibly relevant moral meaning of the situation. It may be that a person interprets the social situation differently or that he or she simply does not think, or not very thoroughly, about the moral implications of a decision or action. The person then does not realize what the consequences are for him- or herself and others nor whether it violates a moral principle. For instance, experiments on 'moral framing effects' conducted by Wolsko et al. (2016) show that conservatives tend to shift their opinion on climate change when presented with a pro-environmental message framed in terms of conservative values like authority and sanctity instead of the usual care/ harm frame. This suggests that it was at first not realized that within the same situation or issue particular moral values were in jeopardy. In the context of aggressive driving it may be that one realizes that speeding increases the risk on a collision but that he or she does not take the second step of really thinking about how this could dramatically impact the lives of the people hurt in a potential accident.

Thirdly, it is possible that a person knows the material facts of the situation and realizes what the possible consequences are for one's and other's wellbeing, but that he or she places relevant others outside of one's moral concern. This often takes some form of dehumanization. The person may think of the other road users as annoying objects and 'forget' that there are actual human beings in those other cars and on those bikes. Delbosc et al. (2019) recently provided evidence for this train of thought.

Schwartz (2016) mentions several psychological mechanisms that can be linked to one or more of these possible reasons for not becoming aware of a relevant moral meaning of a situation. One is the notion of ethical blindness, which refers to the

[15] It should be noted though that also for the rational model this can be the case. At least in a very preliminary way, as becoming aware of the moral meaning of a decision situation implies a cognition at some level that a moral value is at stake, which implies an initial form of moral judgment.

temporal and non-conscious deviation from one's own norms as you do not see the moral dimension of a decision. Palazzo et al. (2012) explain that this often has its source in a framing of the specific situation by the individual that blocks a moral interpretation. So, in our context, someone is driving to work but is only focused on getting there on time for an important meeting. They may only think of how to be quicker without realizing that the reckless moves they perform in the rush put themselves and others at considerable risk. Another, related mechanism, is Bandura's (1999) notion of moral disengagement, which entails the ability to convince oneself that ethical standards do not apply to a particular situation, by using rationalizations such as disregarding the injurious effects for others or by dehumanizing others. This process of evading the moral meaning of the situation may be supported by using sanitizing or euphemistic language. When you refer to bikers as "annoying objects" they are no longer humans and you cutting them off may seem rather like a "fun" game than jeopardizing people's life. The moral dimension of the decision then fades away. Such mechanisms of rationalizations and using sanitizing language may also constitute or result in a lack of awareness retroactively: though you may have become aware of moral considerations at first, these are retroactively negated through these mechanisms ("if he drives on the pavement, I am allowed to do it"). Also, rationalizations or sanitizing language may in the long term become so ingrained that they prevent you from realizing a relevant moral meaning of a situation in the first place ("tourists do not know how to bike so they should just get out of my way").

Schwartz (2016) states that these psychological mechanisms are often related to someone's environment, and are thus predominantly influenced by situational factors. In our context: another person in the car may considerably instigate, support or shatter one's dehumanizing views and rationalizations. However, also individual factors play a role. One that influences moral awareness is, for instance, a lack of moral imagination that prevents one to think of what possible consequences an action can have for others. You are then less inclined to imagine the impact your reckless action may potentially have on other road users. Such situational and individual factors then determine whether the awareness stage commences and thus whether an endorsed moral value can potentially influence driving behaviour.

Moral judgment

In the case that a person does become aware that one or more moral values are at stake, then the next phase commences in which a moral judgment is constituted. A moral judgment is defined as the "determination of the most ethically appropriate course of action among the alternatives" (Schwartz, 2016, p. 767). It is here that endorsed moral values materialize into a concrete judgment about what is the right thing to do. Schwartz (2016) emphasizes that the constitution of a moral judgment involves an interplay of reason, intuition, emotion and rationalization. However, how this mechanism exactly works and what influence each has is still debated.

For our purpose of explicating the moral value-moral behaviour relationship, it should be noted that the endorsement of a particular moral value does not determine one particular moral judgment. First of all, a moral value can be outweighed by another

moral value that is considered more important within a concrete situation. Think of driving a friend to an important meeting that he cannot miss. Your loyalty towards him may require you to speed, overriding your possible concern for other road users. The importance and influence of a generally endorsed moral value can thus be said to be context depended and can therefore differ per situation. Also, general moral values are not specified towards a particular situation; the same general moral value can therefore result in different and sometimes even contrasting moral judgments within a specific context. Speeding your friend to a meeting is in line with your loyalty towards him, but may conflict with your possible loyalty towards society and its laws. In other words, though people may agree in embracing loyalty in an abstract sense, they may still differ on what loyalty demands in a particular situation.

Moral intention and behaviour

Once a moral judgment is formed on what would be the right course of action, this does not automatically lead to the according behaviour. In the third stage this moral judgment is weighed against other non-moral considerations, like egoistic preferences, before a final intention[16] or decision is formed regarding what to do. Say that you are aware that speeding will put others at considerable risk and you know that the right thing to do is to stick to the speed limit. However, you still decide to step on the gas because you really want to be home as soon as possible because your favourite TV-show is starting. In this case the moral value of caring for others that you endorse in general is overruled by an egoistic preference and as such does not affect the eventual behaviour.

In Schwartz's (2016) model, whether a moral judgment is overruled, again, depends on individual and situational factors. Individual factors that play a role at this stage are those that influence the moral integrity side of someone's moral capacity, like the level of one's moral identity (Aquino & Reed, 2002) and, as Schwartz mentions, moral courage (Hannah et al., 2011). Situational factors that can co-determine whether the moral judgment also becomes the behavioural intention are, for instance, phenomena like the influence of an authority (Milgram, 1963) or the above mentioned ethical vulnerability of a person. A situational factor that has been shown to be of considerable effect on risky and aggressive driving is peer pressure (Shepherd et al., 2011).

Note that Festinger's theory of cognitive dissonance (1957) suggests that such a discrepancy between a moral judgment and your final decision cannot be maintained without considerable stress. In the case that a person sticks to a decision that deviates from his or her moral judgment it is likely that the person will adjust the initial judgment through rationalizations. One may for instance downplay the risk and possible consequences of speeding, dehumanize potential victims, or come up with an alternative moral reason that supports the final decision. This then produces a back loop from decision to moral judgment or, as mentioned above, to moral awareness. Schwartz

[16] When the intention is in line with the moral judgment it is regarded as 'moral intention' (Schwartz, 2016; Rest, 1986).

(2016), though emphasizing the role of rationalizations in the moral decision process, does not make this back loop explicit in his model. This may be due to the fact that he discusses the moral judgment and intention stages simultaneously. Indeed, though analytically distinct, it is important to stress that the different process stages can commence virtually at the same time and that together with the mutual influence of process stages this makes the moral decision process not a strictly sequential one.

When the (moral) intention or decision is made this consequently leads to according behaviour. However, this does not have to be an automatism; it is still possible that the behaviour stage does not commence even though the intention has been formed. The intended action may be out of your reach or a physical barrier may be in the way of performing the intended action. For instance, if you intend to speed you may be stopped from performing this behaviour by road bumps or another car blocking your way.[17]

Together, the above elaboration gives a conception of the moral value-aggressive driving behaviour relationship and the 'route' along which moral values need to travel before having an effect on aggressive driving. When someone with a certain moral foundation becomes aware that an endorsed moral value is at stake in a certain situation, judges a certain course of action as the right one in light of this value and other values, and sticks to his or her judgment by forming the intention to act upon it, while not hindered by any other barriers, the endorsed moral value will affect the driving behaviour. Whether this relationship materializes for a particular person in a particular situation at a particular moment in time depends on a variety of individual and situational factors. It should thus be recognized that this conceptual chain reflects a rather uncertain and indeterminate route; the question appears legitimate to what extent this influences, i.e., weakens or even precludes, the empirical relationship between moral values and aggressive driving behaviour. In the following we will empirically explore this question by testing the effects of drivers' moral foundations on their aggressive driving behaviours.

2.4 Method

2.4.1 Measurements

To measure people's moral values and risky and aggressive driving behaviour we employed an online survey consisting of the Moral Foundations Questionnaire (MoralFoundations.org, 2013) and the Aggressive Driving Behavior Scale (ADBS) (Houston et al., 2003). The ADBS consists of 11 self-reported items related to aggressive and potentially harmful driving behaviour (see Table 1). Respondents were asked to indicate the frequency with which they have engaged with each of the 11 aggressive driving behaviours in the past six months, on a scale from one to six (1= never, 6= always). The scale has previously been validated in a study among American

[17] Note that physically blocking a *moral intention* in the context of aggressive driving is much harder to imagine.

undergraduate students (Houston et al., 2003), revealing two components, namely "speeding" and "conflict behaviour". As the members of the target population of the research are Dutch drivers, the items were translated from English into Dutch.

A Principal Component Analysis (PCA) was conducted to assess the dimensionality of the ADBS-items in our sample and summarize the data. Instead of the two-component solution found by Houston et al. (2003), our data produced a three-component solution that (after varimax rotation) led to a simple structure (all loadings >0.5 and all cross-loadings <0.5) which was easily interpretable (see Table 1). The components reflect three different aspects of aggressive driving, namely 'speeding/rushing', 'thwarting', and 'aggressive communication'. Based on the items with loadings above 0.5 sum scores were computed, which were used as dependent variables in the subsequent analyses.

Table 1. Rotated component matrix

Component	Item	Loadings
Speeding	1. Drive 20 miles per hour faster than the posted speed limit	0.738
	2. Follow a slower car at less than a car length	0.703
	3. Pass in front of a car at less than a car length	0.638
	4. Accelerate into an intersection when the traffic light is changing from yellow to red	0.537
	5. Merge into traffic even when another driver tries to close the gap between cars	0.517
Thwarting	6. Intentionally tap my brakes when another car follows too closely	0.660
	7. Follow another car in front of me closely to prevent another car from merging in front of me	0.611
	8. Speed up when another car tries to overtake me	0.607
Aggressive Communication	9. Flash my high beams at slower traffic so that it will get out of my way	0.802
	10. Honk when another driver does something inappropriate	0.695
	11. Make rude gestures at other drivers when they do something I don't like	0.548

Moral values were measured through the 20-item MFQ-scale (MFQ20) (Graham et al., 2011; MoralFoundations, 2013). MFQ20 consists of twenty items plus two control items and is divided into two parts. The first part contains ten considerations (two for each moral foundation) that may to various extents be relevant to one's thinking when deciding whether something is right or wrong, judged on scales from 1 (not at all relevant) to 6 (extremely relevant). For example, whether or not someone suffered emotionally (an item of the care/harm dimension) or whether or not some people were treated differently than others (an item of the fairness/cheating dimension). The second part consists of ten statements (again two for each foundation) to which respondents indicate their agreement on a scale from 1 (strongly disagree) to 6 (strongly agree). This part includes statements such as 'respect for authority is something all children need to learn' (an item of the authority/subversion foundation) and 'people should not do things that are disgusting, even if no one is harmed' (an item of the sanctity/degradation foundation). For the Dutch translation of MFQ20 we used the official one that is made available by its developers through the moral foundations website (MoralFoundations.org, 2013).

For the MFQ we used the existing five component structure that is given by the theoretical framework of MFT (see table 2). While this structure has been validated in several studies (Graham et al., 2011), other studies report that MFQ only gives rise to a two-factor model where care and fairness are grouped together as the "individual" foundations and loyalty, authority and purity as the "binding" foundations (Curry et al., 2019; Iurino & Saucier, 2019). This two-factor structure has also been recognized by the developers of MFT/MFQ and has, as such, also been used and referred to in several empirical studies (Smith et al., 2014; Graham et al., 2009). A PCA performed on our own dataset in fact generates a six-factor solution. In line with the fact that different datasets find a different number of factors, we found that on our data the scales of the five-factor structure stay below the reliability threshold of 0.7. In the discussion we will further elaborate and substantiate the choice for using the original structure of the MFQ as our measurement tool, despite a lower scale reliability than usually desired, and also elaborate on the extra analyses that were performed to ensure that this choice did not influence our results and conclusions.

Table 2. Descriptive statistics of the five moral foundations (based on MFQ20)

Moral foundation	Mean	SD	Reliability (Cronbach's α)
Care/ Harm	4.04	0.87	0.57
Fairness	4.60	0.71	0.49
Loyalty	3.05	0.90	0.63
Authority	3.33	0.83	0.52
Purity	3.86	0.89	0.55

Our survey additionally included several socio-demographic characteristics and driver's license ownership.

2.4.2 Data collection and sample composition

For data collection we made use of a convenience sample (N=281). The first author and a master student promoted the survey among their respective social networks. To ensure that respondents could potentially engage in the behaviour under investigation (i.e. aggressive driving), only respondents that were in possession of a driver's license were considered eligible and included in the sample. The demographic make-up of the sample is shown in Table 3. With respect to gender, the sample is more or less representative, but the distributions of age and education level were biased compared to the respective Dutch population distributions. In particular, higher educated and younger people were overrepresented. Although age and education level influence average levels of aggressive driving downwards and upwards respectively (see also our results further below), it should be noted here that, importantly, our study does not aim to make any claims about these average levels of aggressive driving per se. Rather, we focus on the relation between moral values and aggressive driving behaviour; clearly, there is no conceptual intuition or theoretical reason to expect that sample-bias in terms of age and education would have a (large) effect on these estimated relationships.

Finally, note that in our empirical analyses, we use age and education (as well as gender) as control variables, which was enabled by a sufficient range in these variables.

To prevent bias due to ordering effects, we randomly alternated the order of either asking the MFQ-items or the ADBS-items first. Half of the respondents were thus first asked about their aggressive driving behaviour and half first about their moral foundations.

2.4.3 Method of analysis

To test whether the endorsement of the different moral foundations predicts aggressive driving behaviour, multiple linear regression analyses were conducted. Six models were estimated; two for each type of aggressive driving where the latter was entered as the dependent variable. The first of each pair includes only the three control variables age, gender, and education, the second also the five moral foundations as predictors (see Tables 4-6). Using multiple regression analysis enabled us to estimate the individual effects of each moral foundation on the different types of driving behaviour. By estimating two models per aspect of aggressive driving in which the first model consisted of only the control variables and adding the moral foundation as predictor in the second of the pair, it was possible to discern the unique predictive value of the moral foundations framework as a whole on aggressive driving behaviour. Each model satisfied the assumptions for linear regression analysis. An inspection of the residuals showed that the errors approximate the normal distribution, also there was no multicollinearity among the independent variables (VIF all between 1 and 2).

Table 3. Sample distributions of socio-demographic characteristics

		Sample (%)	Population (%)[a]
Gender	Male	54.8	49.6
	Female	44.5	50.4
	other	0.7	-
Age	18-27	11.7	12.7
	28-37	34.6	12.4
	38-47	12.0	12.4
	48-57	15.2	14.8
	58-67	22.3	12.6
	>67	4.2	15.8
Education level	Lower	19.4	65.3
	Higher	80.6	34.7

[a] Retrieved from Statistics Netherlands (CBS, 2019b)

2.5 Results

The results show that for each aspect of aggressive driving only one foundation has a significant, although weak effect. For 'speeding/rushing' only the fairness foundation has a significant negative effect (model 2). This means that those who consider fairness to be an important value when making a moral judgment are less inclined to show speeding or rushing behaviour while driving. A high or low endorsement of the four other foundations does not have any effect on speeding and rushing based on this

model. Regarding aggressive communication, we only find an effect for the care/harm foundation, which is again negative (model 4). This indicates that people who consider caring for and not harming others to be important when making a moral judgment score lower on the aggressive communication scale. For the four other foundations we found no statistical association with the communication aspect of aggressive driving. Also, for thwarting we find only one moral foundation, loyalty, to have a significant effect (model 6). Interestingly, the found relationship between loyalty and thwarting is *positive*. This suggests that people who consider the value of loyalty important when making a moral judgment also tend to thwart other drivers on the road more. We will further go into and interpret this, at first sight, counter intuitive finding in the conclusion.

When considering the added explanatory power of the models (delta R squared) when the five moral foundations are included as predictors, it is seen that these do not explain significantly more variation in aggressive driving than the models consisting of only the control variables.

Table 4. Standardized parameter estimates of the regression models predicting speeding/ rushing

	Model 1		**Model 2**	
	Beta	***p*-value**	**Beta**	***p*-value**
Age	-0.234	0.000	-0.241	0.000
Female (ref.: male)	-0.228	0.000	-0.200	0.001
High level of education (ref.: low)	-0.020	0.727	0.000	0.999
Care/ Harm			-0.014	0.843
Fairness			-0.148	0.036
Loyalty			0.059	0.429
Authority			0.081	0.245
Purity			0.058	0.407
R square	0.102		0.133	
Delta R square			0.031	0.089

Table 5. Standardized parameter estimates of the regression models predicting aggressive communication

	Model 3		**Model 4**	
	Beta	***p*-value**	**Beta**	***p*-value**
Age	-0.297	0.000	-0.335	0.000
Female (ref.: male)	-0.267	0.000	-0.258	0.000
High level of education (ref.: low)	0.013	0.824	0.014	0.817
Care/ Harm			-0.142	0.041
Fairness			0.074	0.275
Loyalty			0.085	0.239
Authority			-0.061	0.370
Purity			0.128	0.061
R square	0.156		0.183	
Delta R square			0.027	0.114

Table 6. Standardized parameter estimates of the regression models predicting thwarting

	Model 5		Model 6	
	Beta	***p*-value**	**Beta**	***p*-value**
Age	-0.151	0.012	-0.164	0.008
Female (ref.: male)	-0.123	0.036	-0.068	0.274
High level of education (ref.: low)	0.101	0.092	0.144	0.022
Care/ Harm			0.014	0.850
Fairness			-0.128	0.077
Loyalty			0.174	0.023
Authority			0.051	0.476
Purity			-0.037	0.609
R square	0.052		0.084	
Delta R square			0.032	0.099

2.6 Conclusion and discussion

The conducted conceptual and empirical investigation into the relationship between moral values and aggressive driving behaviour gives us new insights into and a better understanding of the role that moral values play in the constitution of this kind of behaviour. The empirical results suggest that the relationship is very weak at best. Per aspect of aggressive driving we find only one moral foundation to have a predictive effect. When controlled for age, gender and education, the standardized effects are all well under 0.2 and must therefore be regarded as rather small. Though few and weak, the found effects are interpretable. The negative effects of fairness and care/ harm on speeding and aggressive communication respectively are as expected. Also, the found positive effect of the moral foundation of loyalty on thwarting is logically interpretable. Though it may seem counterintuitive at first to suggest that a higher endorsement of a moral value can lead to behaviour that is actually undesirable from a moral and societal point of view, this relation is intelligible from an in-group/out-group perspective. The loyalty foundation includes the idea of taking care of the in-group and protect them from outgroup threats (Smith et al., 2014; Graham et al., 2013). When on the road, other drivers may well be identified as an out-group potentially threatening you and your in-group (e.g. family) in your car. Getting other drivers off your tail by tapping the breaks, for instance, or not giving them leeway in order to improve your own position and that of the one's in your car seems logical for someone who endorses the value of loyalty.

Furthermore, the amount of extra variation explained by the models that, besides the three control variables, also include the five moral foundations does not reach statistical significance. This means that adding the five moral foundations to the model does not lead to a better prediction of aggressive driving behaviour. The endorsement level of the different moral foundations does then not have an effect on aggressive driving behaviour that cannot already be explained by age, gender and education. This also suggests that the individual moral foundations for which a significant effect was found mediate, for a very small part, the effects of the social demographics on aggressive driving. Model 2 for instance shows a slight reduction of the effect of gender on

speeding/ rushing in comparison with model 1 and a small significant effect for fairness. This indicates that the relation between gender and the tendency to speed may for a small part be explained by the value of fairness. Overall, however, our results imply a very limited association between the moral foundations framework and aggressive driving, and a very weak prediction of the latter by the former.

Considering the fact that aggressive driving has a clear moral component and that it therefore seems to be intuitive that people's moral values influence their (absence of) aggressive driving behaviour, our results raise important questions for further explanation. By building upon our conceptual analysis of the process of moral decision-making it is possible to interpret and further clarify the results and argue whether, based on this analysis, it is in fact in line with what should be expected. Before elaborating on the interpretation of our empirical results in terms of the process of moral decision-making we will first go into the limitations of our empirical study that may have influenced our results.

First of all, it should be noted that the analysis is based on a convenience sample whose composition does not fully reflect the Dutch population. As discussed in section 2.4.2, we believe that the influence of this on our results and conclusions is limited. Of course, to test the robustness of our empirical findings beyond the confines of our sample, further empirical research is needed. It can then also be tested if our findings can be replicated in different cultures and different countries.

Secondly, we are aware of some limitations regarding the used measurement tools concerning its reliability and validity. With respect to *reliability* it is important to be conscious of the fact that asking people about their moral beliefs and about behaviour that is generally found to be morally or socially objectionable is of course sensitive to socially desirable answering. Even the assurance of anonymity may not fully take away that effect. Although this is necessarily a limitation of our study, as it is a limitation of many studies dealing with socially deviant behaviour (and as such a common issue within fields like criminology), we believe that the effect on our results is limited. The main reason for this belief is that our study aims to discern the relationship between the two main variables instead of estimating variable levels within the population. As long as we can assume that respondents under report their level of aggressive driving (and, in the same vein, over report their level of moral value endorsement) to the same degree, so for instance all one point lower (or higher) than in reality –note there does not seem to be a pressing reason not to assume this– the variations used for predicting the dependent variable based on the independent variables can be assumed to be similar (though varying around a higher or lower mean than found in reality) and thus useful for analyzing their relationship.

A second concern regarding the reliability of the used measurement tools is whether people are able to make a good estimation of their actual aggressive driving behaviour of the last six months. The fallibility of memory and interpretation seems to play a role here. However, any errors resulting from this may be assumed to be random and as such will only affect the reliability of the parameter estimates (the standard errors) and not the estimates themselves.

A third issue concerning reliability is the selection of a particular constellation and number of factors and their internal consistency. Note, that for both the ADBS and the MFQ20 there is the choice to go for the original factor structure *or* the solution that is generated by our own data. Regarding the ADBS we chose for our own solution instead of the original one given by the developers of the questionnaire, as the PCA here functions as a data-reduction technique and we wanted to use the best summarization possible that is logically interpretable. This was the case for the found three factor structure of the ADBS that we used in the analysis. Note that two out of the three scales do not reach the 0.7 Cronbach's Alpha reliability threshold. However, because we use the dimensions here merely as a way to summarize the data and do not assume them to be existing latent variables this does not seem problematic.

In the case of MFQ20 we did decide to stick to the original structure, instead of the six-factor solution that appeared from our data. The reason for choosing the original structure here is that in this case the factors do represent genuine psychological latent variables. As it is our primary goal to use MFQ20 as a theoretically backed up empirical tool to measure people's endorsement of different moral values, using the original structure seems to be the right route. As reported in the method section though, the reliability of the original MFQ20 scales did not reach the usually required threshold.

To be sure that the latter as well as the other abovementioned choices concerning scale structure did not affect our results and consequent conclusions, we conducted (but will not report here for reasons of space limitations) some additional tests. First of all, it was tested whether using the above-mentioned alternative scales –the original ADBS structure, the two-factor MFQ structure (*see* method section) and the six-factor MFQ structure– render different results. Secondly, a series of eleven *ordinal* regression models were estimated, one for each single ADBS-item, which was entered as the dependent variable. With these latter analyses we checked whether the used summarization of the ADBS-data through the three-factor structure influenced our results. Also, conducting ordinal regression analyses enabled us to make sure that the possible limitation of entering originally ordinal variables, i.e., the three ADBS-factors, as dependent variables in the linear regression analyses -thereby assuming them to function as continuous variables- did not affect the results. None of the above additional analyses produced results that are substantively different from those presented in the results section.

With respect to the *validity* of the used measurement tools the question can be posed whether they measure what they intend to measure. A limitation of our research in this regard is that at least one important form of what generally may be regarded as risky and aggressive driving is not included in the ADBS-questionnaire: drunk driving. In future research, it would be interesting to take this form of aggressive driving into account. One approach would be to investigate the relation between one's moral values and the decision to drink and drive. Another interesting aspect is to look into how the moral decision process underlying aggressive driving behaviour is influenced by the intake of alcohol.

However, also when the mentioned limitations are taken into account, we have good reason to believe that we would still find the same few and weak relationships

between people's moral values and aggressive driving behaviour. An explanation for this result can namely be found within the process of moral decision-making itself. In light of the discussed conceptual model of moral decision-making, it becomes clear that for a moral value to have an effect on behaviour it needs to go through several process stages that are each influenced by individual and situational factors. This makes the relationship between moral values and aggressive driving behaviour indeterminate. First of all, as we have seen, a person driving in a car may not become aware that the choice for aggressive driving may violate (or is supported by, as seems to be the case for some persons endorsing the value of loyalty) an endorsed moral value. In the case awareness does arise, a judgment follows whose outcome is not just dependent on the level of endorsement of the moral value in a general non-specified context, as is the setting of MFQ, but also on the specific context the person is confronted with. For instance, when different endorsed moral values are in conflict it also depends on the context which one prevails. Furthermore, it can be unclear what choice a general moral value actually determines in a specific context. The found positive effect for loyalty on aggressive driving (thwarting) indicates this possible indeterminacy as it may conflict with an interpretation of loyalty that instigates conformity towards the rules of the community. Then, once a judgment is made, it may subsequently be pushed aside by a preference for certain egoistic considerations. In the case that the moral judgment does prevail and a moral intent is formed then there may be certain obstacles that prevent the intended behaviour to commence (though for the decision *not* to commit aggressive driving this seems hard to imagine).

Because a generally endorsed moral value is for its effect dependent on these process stages that are each influenced by individual and situational factors, the potential effect of a moral value on one's behaviour is easily annulled. This then can account for the found absence of and weak relationships between people's moral values and reported aggressive driving behaviour. Though people endorse a certain value in general, it does not take effect when driving in a car.

The above analysis has at least the following implications. First of all, it suggests that in order to get a better understanding of the phenomenon of aggressive driving and the role of morality in the constitution of this behaviour we need to zoom further in into the different process stages of the moral decision-making process within the specific context of aggressive driving and see what factors influence them. So, for instance, what context specific situational and individual factors play a role in (not) becoming aware of the moral implications that a choice for aggressive driving has? And, furthermore, what context specific egoistic considerations play a role that can overrule someone's moral judgment to not drive aggressively? What factors influence people their moral motivation and moral perseverance when they are driving in a car? This first point also suggests that the found empirical result does not imply that moral values necessarily have barely any effect on aggressive driving behaviour. It only means that at this moment it is quite weak, but also that it might be improved by, for example, raising awareness about the moral impact of people's choices while driving, or by improving people's moral motivation and moral perseverance to keep them from discarding their moral judgment in favour of self-interest while driving a car.

A second implication of our work regards the validity and scope of the MFQ. With MFQ we assume to measure aspects of morality that play an actual role in the decision-making process, but the question can be asked to what extent this is indeed the case. The questions asked in the questionnaire are quite general, not tailored to the context of aggressive driving, and could also just reflect what people find morally relevant in a very general sense. It may well be then, that these general moral statements made by people in a questionnaire do not reflect, or barely reflect, the moral dimensions that play a role when real life decisions are made in the context of aggressive driving nor in other particular contexts. Our conceptual discussion as well as our empirical results clearly suggest that this may be a fundamental problem associated with the MFQ. Although the MFQ is a well-established tool to measure people's beliefs and attitudes, the above-mentioned criticism should be taken seriously and asks for further investigation into its validation and its usefulness as a predictor of concrete and particular moral behaviours. A potential fruitful line of research could be to investigate the presence and measurement of more contextual moral values, for instance, to what extent persons care about vulnerable others, and their relationship to aggressive driving behaviour. In such an investigation one should be aware though of the possible inverse of the causal relationship between the context specific moral values and aggressive driving as a consequence of predicting behaviour from more specific attitudes (Kroesen & Chorus, 2018).

This brings us to the following policy implications. The first is that in order to increase traffic safety, developing safety policies only targeted at changing people's moral values does not seem the most effective way forward. Policies should have significant attention for the different process stages that play a role in the moral decision-making process and that form the route from moral value to behaviour. This means that attention needs be directed toward e.g. raising people's moral awareness when driving a car and inducing them to hold on to their own moral judgment in the face of internal and external pressures to choose otherwise. For awareness, our analysis at least suggests that people should become aware of the *facts* about aggressive driving, realize the *moral meaning* it has, by linking it to a moral principle or realizing the actual consequences it has for others, and thirdly, by acknowledging that this moral meaning is *applicable to other traffic users*, as on every bike and behind every car windscreen there is an actual human being. Which individual and situational factors exactly influence these aspects of awareness and people's moral intent within the context of aggressive driving is subject to further research. If our analysis is on the right track, this suggests that it may be possible to influence aggressive driving behaviour without necessarily raising the level of the endorsement of certain moral values, but rather by influencing the role of the process stages that block or give way to the influence of moral values that are already endorsed. However, further empirical research is needed on the effects of the intermediate process stages on aggressive driving behaviour before anything more conclusive can be said about this idea.

Secondly our empirical and conceptual analysis suggest that targeting 'soft' factors, like people's moral beliefs and attitudes towards aggressive driving, is quite a vulnerable route for preventing this kind of behaviour. This supports the idea that

physical interventions, located in between the intention and behaviour stage, could be the most effective ones. On roads one can of course think of speedbumps and other physical measures that prevent speeding. But it may also be possible to put restrictions on the car itself that make it impossible to show aggressive driving behaviour. Research on the effect of alcohol ignition interlock supports the above train of thought as it is shown that the device is only effective in reducing traffic violations as long as it is installed and physically prevents drunk driving. Once it is removed and the behaviour of drunk driving is only dependent on (moral) decision-making, there is no difference anymore between the ones who formerly were and were not subjected to this measure (Beck et al., 1999).

References

Adminaité-Fodor, D., Heilpern, C., & Jost, G. (2019). Ranking EU progress on road safety: 13th road safety performance index report. Retrieved from https://etsc.eu/wp-content/uploads/ETSC-PIN-AR_2019-Final-EMBARGO.pdf

Ajzen, I. (1988). *Attitudes, personality, and behavior*. Chicago: Dorsey Press.

Andorfer, V. A., & Liebe, U. (2012). Research on fair trade consumption—A review. *Journal of business ethics*, *106*(4), 415-435.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.

Aquino, K., & Reed, I. I. (2002). The self-importance of moral identity. *Journal of personality and social psychology*, *83*(6), 1423.

Arvola, A., Vassallo, M., Dean, M., Lampila, P., Saba, A., Lähteenmäki, L., & Shepherd, R. (2008). Predicting intentions to purchase organic food: The role of affective and moral attitudes in the Theory of Planned Behaviour. *Appetite*, *50*, 443–454.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J-F., & Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59-64.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*(3), 193–209.

Beck, K. H., Rauch, W.J., Baker, E.A., &Williams, A.F. (1999). Effects of Ignition interlock license restrictions on drivers with multiple alcohol offenses: a randomized trial in Maryland. *American Journal of Public Health*, *89*(11), 1696-1700.

Bianchi, A. and Summala, H. (2002). Moral Judgment and Drivers' Behavior among Brazilian Students. *Psychological Reports*, *91*(3), 759-766.

Brown, T.G., Ouimet, M.C., Eldeb, M., Tremblay, J., Vingilis, E., Nadeau, L., et al. (2016) Personality, Executive Control, and Neurobiological Characteristics Associated with Different Forms of Risky Driving. *PLoS ONE*, *11*(2), 1-18.

CBS (2019a). Death by traffic accidents. Retrieved from: https://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=81452ned&D1=0-27&D2=0&D3=0&D4=0,4,9,(l-9)-l&VW=T

CBS (2019b). Level of education and age of Dutch population. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82275NED/table?fromstatweb

Chorus, C. G., Pudāne, B., Mouter, N., & Campbell, D. (2018). Taboo trade-off aversion: A discrete choice model and empirical analysis. *Journal of choice modelling*, *27*, 37-49.

Curry, O. S., Chesters, M. J., Van Lissa, C.J. (2019). Mapping morality with a compass: Testing the theory of 'morality as cooperation' with a new questionnaire. *Journal of Research in Personality*, *78*, 106-124.

Delbosc, A., Naznin, F., Haslam, N., & Haworth, N. (2019). Dehumanization of cyclists predicts self-reported aggressive behaviour toward them: A pilot study. *Transportation Research Part F*, *62*, 681-689.

Evans, L. (1993) Comments on driver behavior and its role in traffic crashes. *Alcohol, Drugs and Driving*, *9*, 185-195.

Fergusson, D., Swain-Campbell, N., & Horwood, J. (2003). Risky driving behaviour in young people:

prevalence, personal characteristics and traffic accidents. *Australian and New Zealand journal of public health*, *27*(3), 337-342.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. (2013). Moral Foundations Theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55-130.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029–1046.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*, 366-385.

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

Haidt, J., & Joseph, L. (2004). Intuitive Ethics: how innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*, 55-66.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*, 998-1002.

Hannah, S. T., Avolio, B. J., & May, D. R. (2011). Moral maturation and moral conation: A capacity approach to explaining moral thought and action. *Academy of Management Review*, *36*(4), 663–685.

Harsanyi, J.C. (1982). Utilitarianism and beyond. In A. Sen & B. Williams (Eds.), *Morality and the theory of rational behaviour* (pp. 39-62). Cambridge: Cambridge University Press.

Houston, J.M., Harris, P.B. and Norman, M. (2003). The Aggressive Driving Behavior Scale: Developing a Self-Report Measure of Unsafe Driving Practices. *North American Journal of Psychology*, *5*, 269-278.

Iurino, K., & Saucier, G. (2019). Testing measurement invariance of the moral foundations questionnaire across 27 countries. *Assessment*, 1-8.

Iversen, H.H. & Rundmo, T. (2002) Personality, risk behaviour and accident involvement among Norwegian drivers. *Personality and Individual Differences*, *33*, 1251-1263.

Jones, M.T. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review*, 16(2), 366-395.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D.A. Goslin (Ed.), Handbook of socialization theory and research (pp. 347-480). Chicago: Rand McNally.

Kroesen, M., & Chorus, C.G. (2018). The role of general and specific attitudes in predicting travel behavior – A fatal dilemma? *Travel Behaviour and Society*, *10*, 33–41.

Li, S., Zhang, J., Li, P., Wang, Y., & Wang, Q. (2019). Influencing factors of driving decision-making under the moral dilemma. *IEEE Access*, *7*, 104132-104142.

Machin, M.A. & Sankey, K. S. (2008). Relationships between young drivers' personality characteristics, risk perceptions, and driving behaviour. *Accident Analysis & Prevention*, *40*(2), 541-547.

Matthies, E., Klöckner, C. A., & Preissner, C. L. (2006). Applying a Modified Moral Decision Making Model to Change Habitual Car Use: How Can Commitment be Effective? *Applied Psychology*, *55*(1), 91–106.

Mallia, L., Lazuras, L., Violani, C., & Lucidi, F. (2015). Crash risk and aberrant driving behaviors among busdrivers: the role of personality and attitudes towards traffic safety. *Accidents, Analysis & Prevention*, *79*, 145-151.

Milgram, S. (1963). Behavioral study of obedience. *The Journal of abnormal and social psychology*, *67*(4), 371-378.

MoralFoundations.org. (2013). Retrieved August 20, 2019, from https://moralfoundations.org

Nilsson, A., Erlandsson, A., & Västfjäll, D. (2016). The congruency between moral foundations and intentions to donate, self-reported donations, and actual donations to charity. *Journal of Research in Personality*, *65*, 22–29.

Nordfjærn, T., Jørgensen, S.H., & Rundmo, T. (2010) An investigation of driver attitudes and behavior in rural and urban areas in Norway. *Safety Science*, *48*(3), 348–356.

Palazzo, G., Krings, F., & Hoffrage, U. (2012). Ethical blindness. *Journal of business ethics*, *109*(3), 323-338.

Parker, D., Manstead, A.S.R., & Stradling, S.G. (1995). Extending the theory of planned behaviour: the role of personal norm. *British Journal of Social Psychology*, *34*, 127-138.

Parker, D., Stradling, S.G., & Manstead, A.S.R. (1996). Modifying beliefs and attitudes to exceeding the speed limit: an intervention study based on the theory of planned behavior. *Journal of Applied Social Psychology*, *26*, 1-19.

Rest, J.R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.

Schwartz, M.S. (2016). Ethical Decision-Making Theory: An Integrated Approach. Journal of Business Ethics, 139, 755-776.

Shepherd, J. L., Lane, D. J., Tapscott, R. L., & Gentile, D. A. (2011). Susceptible to Social Influence: Risky "Driving" in Response to Peer Pressure 1. *Journal of Applied Social Psychology*, *41*(4), 773-797.

Smith, I.H., Aquino, K., Koleva, S., & Graham, J. (2014). The moral ties that bind…even to out-groups: The interactive effect of moral identity and the binding moral foundations. *Psychological Science*, *25*(8), 1554–1562.

Trevino, L. K. (1986). Ethical decision making in organizations: A person-situation interactionist model. *Academy of Management Review*, *11*(3), 601–617.

Ulleberg, P., Rundmo, T. (2003). Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. *Safety Science*, *41* (5), 427–443.

Vainio, A. & Makiniemi, J. (2016). How are moral foundations associated with climate friendly consumption. *Journal of Agriculture and Environmental Ethics*, *29*, 265–283.

Wolsko, C., Ariceaga, H., & Seiden, J. (2016). Red, white, and blue enough to be green: Effects of moral framing on climate change attitudes and conservation behaviors. *Journal of Experimental Social Psychology*, *65*, 7-19.

World Health Organization (2018). *Global status report on road safety*. Retrieved from https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/

Yang, J., Du, F., Qu, W., Gong, Z., Sun, X., (2013). Effects of personality on risky driving behavior and accident involvement for Chinese drivers. *Traffic Inj. Prev.*, *14*(6), 565–571.

3. Why are General Moral Values Poor Predictors of Concrete Moral Behaviour in Everyday Life?

A Conceptual Analysis and Empirical Study

This chapter has been published as a journal article: van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2022). Why are general moral values poor predictors of concrete moral behavior in everyday life? A conceptual analysis and empirical study. *Frontiers in Psychology*, 3834. https://doi.org/10.3389/fpsyg.2022.817860

Abstract

Within moral psychology, theories focusing on the conceptualization and empirical measurement of people's morality in terms of general moral values –such as Moral Foundations Theory- (implicitly) assume general moral values to be relevant concepts for the explanation and prediction of behaviour in everyday life. However, a solid theoretical and empirical foundation for this idea remains work in progress. In this study we explore this relationship between general moral values and daily life behaviour through a conceptual analysis and an empirical study. Our conceptual analysis of the moral value-moral behaviour relationship suggests that the effect of a generally endorsed moral value on moral behaviour is highly context dependent. It requires the manifestation of several phases of moral decision-making, each influenced by many contextual factors. We expect that this renders the empirical relationship between generic moral values and people's concrete moral behaviour indeterminate. Subsequently, we empirically investigate this relationship in three different studies. We relate two different measures of general moral values -the Moral Foundations Questionnaire and the Morality As Cooperation Questionnaire- to a broad set of self-reported morally relevant daily life behaviours (including adherence to Covid-19 measures and participation in voluntary work). Our

empirical results are in line with the expectations derived from our conceptual analysis: the considered general moral values are poor predictors of the selected daily life behaviours. Furthermore, moral values that were tailored to the specific context of the behaviour showed to be somewhat stronger predictors. Together with the insights derived from our conceptual analysis, this indicates the relevance of the contextual nature of moral decision-making as a possible explanation for the poor predictive value of general moral values. Our findings suggest that the investigation of morality's influence on behaviour by expressing and measuring it in terms of general moral values may need revision.

Keywords: moral values, moral decision-making, moral behaviour, moral foundations theory, compliance with covid-19 measures, contextual aspects of moral decision-making

3.1 Introduction

Studies focusing on the empirical investigation of people's general moral values, implicitly or explicitly assume these to be relevant for the explanation and prediction of people's behaviour in everyday life. Theories within contemporary moral psychology that aim at the conceptualization and measurement of people's general moral values, such as Moral Foundations Theory (MFT) (Graham et al., 2013; Haidt & Graham, 2007) and Morality as Cooperation theory (MAC) (Curry et al., 2019), typically refer to an evolutionary explanation for the existence and content of moral values. Indeed, this only makes sense when the identified moral values also influence actual behaviour. Furthermore, one important reason to study people's general moral values seems to be its potential influence on decision-making and behaviour. Their relevance for (applied) researchers, as well as policy makers, would considerably diminish were it assumed that moral values do not affect nor predict acts performed in everyday life. The relationship between general moral values and the concrete behaviours people perform in everyday life[18] then seems to be an important underpinning of the empirical study of moral values.

However, despite forming an important assumption of moral value research, much is still unknown about the theoretical as well as the empirical aspects of this relationship. Theoretically, for instance, there is no general agreement on how moral decision-making exactly works and how different presumably relevant phenomena, like moral values, moral judgments, empathy, emotions, intuitions and reasoning etc., interact in moral decision-making and consequent behaviours (Hoover et al., 2019; M. S. Schwartz, 2016). Exemplary in this regard is the debate between intuitionists and rationalists on moral reasoning and moral judgment (Bucciarelli et al., 2008; Musschenga, 2009). Intuitionists claim that moral values feature within an emotional and intuitive process of moral judgment formation and regard any deliberate reasoning as post-hoc confabulation. Conscious reasoning involving moral values and moral principles is thereby virtually excluded from having any direct influence on moral

[18] This study focuses on the relationship between general moral values and concrete moral behaviours in everyday life (we use the phrases 'everyday life' and 'daily life' interchangeably in this paper). By this we mean behaviours that we encounter in, or which are often part of, our everyday endeavours. Unique or heroic deeds are not the focus of this study, neither are behaviours performed in an experimental set-up.

judgment (Haidt, 2001). Rationalists give a more prominent place to the process of conscious reasoning in forming moral judgment, featuring moral values and moral principles (Kasachkoff & Saltzstein, 2008; Kennett & Fine, 2009; Kohlberg, 1984; Rest, 1986).

Empirically, on the other hand, research investigating the link between general moral values and concrete moral behaviour in daily life is relatively scarce (Ellemers et al., 2019; O'Grady et al., 2019). As Graham et al. (2012) state, most research on individual differences in morality has concentrated on people's prioritization of values and less on how these differences influence people's moral behaviour in real life. Likewise, Ellemers et al. (2019) conclude in a major review of psychological studies on morality since 1940 that, although authors commonly express that their main interest in studying a certain aspect of morality lies in the explanation and prediction of moral behaviour, the vast majority of studies concentrates on how people *think* about morality instead of how such moral beliefs and attitudes influence actual moral behaviour. As Ellemers et al. (2019) state, the assumed relationship between the studied moral constructs and behaviour remains thereby mainly hypothetical. In line with these observations, general moral values have been more commonly related to attitudinal variables and other general dispositions (Graham et al., 2011; O'Grady, 2019). MFT has, for instance, been extensively related to people's political ideology (Graham et al., 2009; Haidt & Graham, 2007) and socio-political attitudes (e.g., Dickinson et al., 2016; Kugler et al., 2014). Yet, the question whether someone who scores higher on a moral value scale also shows more moral behaviour (related to that dimension) in everyday life has not been satisfactorily answered (Graham et al., 2012; Hoover et al., 2019).

However, this does not mean we are completely in the dark about the empirical aspects of the relationship between general moral values and concrete moral behaviour. First, to get a better grasp of this relationship, it is insightful to turn to the more general field of value research, which focuses on the broader concept of basic or personal values (Sagiv et al., 2017; S. H. Schwartz, 1992). This field has extensively and more systematically studied the empirical relationship between values and behaviour than has so far been done in the moral domain. In particular, the vast amount of literature that builds on Schwartz's theory of basic values (S. H. Schwartz, 1992; S. H. Schwartz et al., 2012), typically defining values as desirable trans-situational goals or abstract ideals guiding people's life, has linked values to a broad set of behaviours (e.g., Bardi & Schwartz, 2003; S. H. Schwartz et al., 2017; S. H. Schwartz & Butenko, 2014). Here, values are usually regarded as important notions to predict and explain how people think, decide, and act within value-relevant situations (Maio et al., 2006; Miles, 2015; Rohan, 2000; Sagiv & Roccas, 2021; S. H. Schwartz et al., 2017). However, at the same time, it is acknowledged that empirical studies generally find only weak to moderate effects of values on behaviour (Cieciuch, 2017; Lee et al., 2021). Studies which consider the morally relevant basic values of 'benevolence' and 'universalism'[19] in connection to moral or pro-social behaviour, find similarly sized effects (e.g., Bardi & Schwartz, 2003; Caprara et al., 2012; Miles, 2015).

The generally found weak to moderate effects between values and behaviour may even be overestimations. Boyd et al., (2015) note that the self-report measures of

[19] In short, 'benevolence' is defined as the enhancement of the welfare of people with whom one is in frequent personal contact; 'universalism' entails the appreciation and protection of the welfare of all people (S. H. Schwartz, 1992).

behaviour in these studies often have a rather general level of abstraction[20], which risks relating only different aspects of one's self-concept (one's values to one's conception of one's broad behavioural or personality traits) instead of one's values to actual concrete behaviour. This suggests that the effects of values on actual concrete forms of behaviour may be weaker than found in those studies. Such weak associations between general values and concrete behaviour are also in line with what has been generally found between general attitudes and more concrete or specific forms of behaviour in other fields, like transportation (Kroesen & Chorus, 2018). These findings, at least, serve as an indication of what can be expected of the effect of general *moral* values on concrete *moral* behaviour, namely that these may be rather weak.

Secondly, we can look into studies that have related moral values to specific forms of moral behaviour. While not specifically focusing on a systematic investigation of the general moral value – moral behaviour relationship (though see O'Grady et al. (2019) for a recent more relevant effort with regards to MFT), a number of studies from different fields do take general moral value measures into account (as part of their models) to explain specific forms of morally relevant behaviour (e.g., Cohen et al., 2014; Nilsson et al., 2016; Reynolds & Ceranic, 2007; Tarry & Emler, 2007; Vainio & Mäkiniemi, 2016 (*study 1*); van den Berg et al., 2020; Díaz & Cova, 2022; Presti et al., 2021; Qian & Yahara, 2020). When inspecting their results, we find, in line with the above, only small effects. While often concluding that the measured general moral values are important predictors of the considered moral behaviours, the reported correlations, when significant, do not exceed the 0.2-0.3 range. This means that at best about 4-9% of the variance in the studied behaviour is explained by people's general moral values. It may even be expected that the actual number of studies finding very weak to no empirical relationships is higher due to publication bias (Rothstein et al., 2006).[21] Vainio and Makiniemi (2016, *study 2*) find somewhat higher effects by making the moral values specific to the context of the behaviour that was studied (climate-friendly consumption). This result suggests that contextual factors may play an important role in the influence of morality on behaviour, possibly undermining the attractive idea of having a limited set of fundamental moral values that can represent one's morality and is able to predict behaviour in different contexts.

In this study, we want to further contribute to the investigation and understanding of the relationship between general moral values and concrete moral behaviour, theoretically as well as empirically. Theoretically, we offer new insights, without solving the difficult debate concerning the exact process of moral decision-making, through a conceptual characterization of the general moral value-concrete moral behaviour relationship that should be acceptable for both intuitionists as well as rationalists. To articulate the relationship, we build on and extend a well-established

[20] Used behavioural items are for instance: 'take it easy and relax', 'do unconventional things', 'make sure everyone I know receives equal treatment' (Bardi & Schwartz, 2003) and 'I try to help others' (Caprara et al., 2012).

[21] Note that there are also studies investigating the moral value-moral behaviour relationship in a game-like or lab settings. As our focus is on the influence of moral values on behaviour in everyday life, we leave these studies out of our main reflections. Though these are important studies to investigate certain dynamics of moral behaviour (Graham et al., 2012), doubt has been raised to what extent lab findings can be translated to behaviour in real life (Graham et al., 2012; Levitt & List, 2007). Either way, findings in these studies do seem to be in line with the above, as only quite weak effects are found (e.g., Clark et al., 2017; Duc Huynh, 2020; Schier et al., 2016).

ethical decision-making model (Rest, 1986; M. S. Schwartz, 2016). From this conceptual characterization we derive the notion that general moral values are expected to be poor predictors of concrete moral behaviour in everyday life. A reason for this is the contextuality[22] of the moral decision-making process that arises from our analysis. In the subsequent empirical part of the study, we look into the empirical side of the relationship between general moral values and concrete moral behaviour. In three separate studies, we relate measures of general moral values to different kinds of self-reported morally relevant behaviours in daily life. In the last study, we also take more specific moral values into account to investigate whether moral values become stronger predictors when they are tailored to the context of the behaviour that they are to predict.

3.2 Specifying general moral values in line with MFT and MAC

Before embarking on the conceptual analysis and subsequent empirical investigation, we start by specifying in some more detail what we refer to by the term 'general moral values'. With general moral values we mean trans-situational moral ideals which guide our moral judgment, i.e., our judgment in terms of 'right' and 'wrong'. These trans-situational moral ideals are the focus of theories such as Moral Foundations Theory and Morality As Cooperation theory. The extent to which individuals endorse these different moral ideals or general moral values is empirically measured with their accompanying questionnaires. Within these theories, general moral values get the character of general moral dispositions. We will explicate this in a bit more detail in the following.

Considering the MFT, what can be viewed as an individual's general moral values is one's endorsement of the five moral foundations (Care/ harm, fairness/ cheating, loyalty/ betrayal, authority/ subversion, sanctity/ degradation). According to MFT, these universal moral foundations have an evolutionary origin, functioning as fitness enhancing solutions to distinct problems of (group) survival; as such, they exist as innate modules on which every human being's morality is built. To what extent each moral foundation is developed into an individual's actual morality depends on factors like cultural influences, upbringing and individual experience. The development of a moral foundation reflects a sensitivity to situations, concepts, principles, beliefs etc. that belong to a certain moral domain, and results in intuitive moral judgments (Graham et al., 2013). The level of development of each moral foundation within an individual's morality can be measured with the Moral Foundations Questionnaire (MFQ) (MoralFoundations.org, 2013). The MFQ-items ask about the relevance of several general moral considerations when making a moral judgment (e.g., 'whether or not someone suffered emotionally'- care/ harm foundation), referred to as the relevance-items, and one's agreement with several general moral statements (e.g., 'I am proud of my country's history'- loyalty/ betrayal foundation), referred to as the judgment-items.

[22]In this study we take context or 'contextuality' to come in degrees, moving on a line between the endpoints of 'in general' – where there is no context, i.e., there is no articulation of a (kind of) situation as there is, indeed, abstracted from such specificity- to the specific decision situation an individual finds oneself in, on a given moment in time. In between, one can find the broader context in which certain social practices take place. This level has a more general character than the actual decision situation itself, but is more specific than the most abstract endpoint. Think of the context of 'driving in traffic', 'the workplace' or 'sustainable food choice'.

Each item belongs to a specific moral foundation. The score per foundation is then taken to be the individual's level of endorsement of this general moral value.

MAC's theory and questionnaire offer a similar structure as MFT. However, as MAC explicitly starts from the theoretical premise that morality evolved as a biologically and culturally developed set of solutions to problems of *cooperation in social life*, it identifies a different -and theoretically stronger substantiated- set of general moral values. According to MAC, each problem of social cooperation and each solution –which are basically those of zero-sum games– gives rise to a distinct moral domain and an accompanying general moral value. MAC identifies seven of those: family values, group loyalty, reciprocity, bravery, respect, fairness, and property rights (Curry et al., 2019). Again, the development of each moral value differentiates between individuals and can be measured for every individual by a questionnaire -MAC-Q- which has a similar structure as MFQ. It asks about the relevance of general moral considerations for the respondent and about the respondents' level of agreement with general moral statements.

In sum, what can be regarded as people's general moral values in light of these empirical moral value theories, may be described as fairly stable psychological dispositions that latently exist within the individual. Also described as moral traits (Haidt & Joseph, 2007), they reflect a trans-situational sensitivity to a moral domain, influencing decision-making across contexts. Accordingly, these general moral values can be measured outside of a specific contextual situation by asking about general moral considerations and principles.

Identifying such moral dispositions, is of course particularly valuable when these have an effect on behaviour; this would help enable the prediction –and possibly also the influencing– of behavioural patterns. As such, it would, for instance, greatly contribute to MFT's claim to 'pragmatic validity' (Graham et al., 2011, 2013), a validity of the theory based on the extent to which it produces further and new forms of understanding of morality and behaviour, especially with regards to people's actual behaviour within everyday life (Rozin, 2006). In the following, a conceptual characterization of the general moral value-concrete moral behaviour relationship is explicated to investigate whether such a relationship is to be expected on theoretical grounds.

3.3 A minimal characterization of the general moral value-concrete moral behaviour relationship

For a general moral value, considered as an individual moral disposition, to perform as a predictor of concrete moral behaviour, its presence within an individual would have to (regularly) lead to this behaviour. To get a better understanding of the possible effect of someone's general moral values on one's behaviour in daily life, it is fruitful to take the perspective of an individual making a moral decision. Even though there is not a universally accepted theory of how individual moral decision-making exactly works, nor how moral values exactly feature within it, it does seem possible to identify basic elements within this process that at a minimum need to become manifest before one's general moral value, regarded as an individual moral disposition, can have an effect on actual behaviour. Rest's (1986) basic Four Component Model of ethical decision-making is a useful starting point here. Rest (1994, p. 23) states that his Four Component Model

is an answer to the question: 'What must we suppose happens psychologically in order for moral behaviour to take place?'. As an answer, the model depicts four different process stages an individual has to go through in order to make a moral decision and perform the corresponding moral behaviour: *moral awareness*, *moral judgment*, *moral intention* and *moral behaviour*.

Though initially developed as a rational model of moral decision-making, a recent extension of the model by M. S. Schwartz (2016) shows that, besides rational, it can also accommodate intuitive/ emotional and mixed conceptions of moral decision-making. As it can incorporate different conceptions of moral decision-making, we can apply the model to answer the question that is relevant to our study, posed in the same vein as Rest: 'At a minimum, what processes need to take place before an individual's endorsed general moral value can affect one's behaviour?'. Although the four-component model provides only a minimal characterization of the general moral value-concrete moral behaviour relationship, it does give insight into what can be expected from the empirical prediction of concrete moral behaviour from general moral values and why. In short, it brings forward that this relationship can be expected to be indeterminate due to the contextuality of the moral decision-making process. In the following, this is explicated by discussing the four stages that an individual's generally endorsed moral value needs go through within a specific decision situation before it can have an effect on behaviour.

3.4 Four necessary process stages for linking general moral values to concrete moral behaviour

3.4.1 Moral awareness

To start with, for someone's general moral value to become effective in decision-making at all, a person needs to become *aware*[23] that within a certain concrete decision situation this moral value plays a role. Overlooking the relevance of an endorsed moral value should not be considered as a rare exception. First of all, social situations need to be interpreted and these can be ambiguous to the decision maker (Latane & Darley, 1970). Is a girl screaming from laughter and fun as she is being teased by friends or from fear as she is harassed by bullies (Thornberg et al., 2018)? Secondly, the decision maker needs to become aware of the actions possible in the situation and of their factual consequences and effects on others. For instance, one may simply not realize that an action breaks a promise or that it has certain harmful consequences for a (group of) person(s). It may be part of one's customary behaviour to which no further (critical) thought is given. Also, some behaviours or situations, though relevant to one's endorsed general moral value, may be less typical exemplars or instantiations of this value and therefore not as easily linked to it by the individual (Hanel et al., 2017). Third, several psychological processes, which Bandura et al. (1996) grouped under the name of processes of 'moral disengagement', make it possible that even when one realizes the factual effects of an action or situation, the *moral significance* of it may still not be recognized. For instance, one may not really empathize with victims due to blaming the

[23] Note that we do not necessarily mean *conscious* awareness here. Intuitive or habitual decision-making and behaviour may be automatic and stay subconscious. However, this kind of decision-making presupposes some form of (subconscious) recognition of what a certain social situation means. This is here regarded as a form of awareness, albeit subconscious.

victim for the harm suffered. Or, due to processes of dehumanization, one may not recognize (or (sub-)consciously downplay) the moral worth of victims (e.g. looking at people from an out-group as being 'inferior') and thereby not become fully aware that a generally endorsed moral value applies to them (Haslam & Loughnan, 2014; Kelman, 1976). Haslam (2006) emphasizes that this is not restricted to contexts of extreme violence but is actually an everyday social phenomenon. Another important aspect of realizing the moral significance of a situation is that it is morally significant to *you* - i.e., that one feels morally responsible to do something. A person may simply not have reflected thoroughly on one's own role in a situation, or, as Bandura et al. (1996) explains, diffuse the moral responsibility to others to evade interference. A generally endorsed moral value will in such cases not influence further decision-making and action.

Psychological research shows that there are many situational and individual factors that can influence these different rounds of interpretation (Craft, 2013; O'Fallon & Butterfield, 2005; M. S. Schwartz, 2016). This then has the result that a generally endorsed moral value, even when it could be relevant within a specific situation, sometimes may not start to play a role in the decision-making process of the individual, causing a first indeterminacy in the empirical general moral value-concrete moral behaviour relationship.

3.4.2 Moral judgment

Once an individual has become aware of the possible lines of actions within a decision situation, of their moral significance, and of one's responsibility, he or she needs to *morally judge* which line of action is the right one. An indeterminacy in the prediction of behaviour from general moral values can arise in this stage due to the fact that in real life a moral decision situation often consists in weighing competing moral values against each other, whether through a deliberative, intuitive, or mixed process. This competition between moral values is not taken into account in a standard moral value questionnaire like MFQ, i.e., one can indicate to find all values (just as) important (Frimer & Walker, 2008). Furthermore, given that some situations are more morally salient than others (Jones, 1991), it is conceivable that in one situation a moral value is more salient than in another. The idea that the relative importance of moral values changes across contexts is also in line with social psychological theories such as Social Identity Theory (Tajfel & Turner, 1979) and Self-Categorization Theory (Turner et al., 1987, 2006), which claim that people can identify themselves differently in different contexts, influencing the importance of values that are partially constitutive of such identities. This makes weighing different lines of actions, informed by underlying moral values, a highly contextual endeavor. To make this more tangible, think of a civil servant who generally prefers the moral value of fairness over loyalty. It may be expected that this person declines a favor asked by a friend to issue a permit without due procedure. However, this does not mean that, when this friend and a stranger are drowning and he can only save one, he will toss a coin, or use any other fair procedure, to determine who to save.

A second indeterminacy in predicting behaviour from general moral values arises within this stage, as the same general moral value may indicate more than one line of action as the right one. In the application of one's general moral value there is simply not one specific action that it determines. This can lead different persons to different

behaviours in the same kind of decision-making context, while claiming to endorse the same general moral value. It can even lead the same person to different behaviours in (only slightly) different contexts. Regarding the first kind of indeterminacy, think of the moral value of care that inclines one person to speed one's mom to the hospital when she is in an emergency, while it may incline another person in the same kind of situation to stick to the speed limit, in order to not risk injuring others. Regarding the second kind, consider that in a subsequent similar emergency situation, the first person may now stick to the speed limit, remembering that he or she almost caused an accident the last time (this reconsideration being induced by the same general moral value of caring for others). The more general problem seems to be here that when measuring a general moral value, it is not clear what specific meanings or instantiations of this general value is considered by respondents (Hanel et al., 2017). These diverting specific meanings can lead to contradictory behaviours between individuals. Also, the meaning given to the value may be linked to a specific context and, therefore, the measured 'general endorsement' by the individual may not be transferable across contexts. This leads to indeterminacy when predicting behaviour from general moral value measures.

3.4.3 Moral intention

When an individual judges which action is the morally right one in light of an endorsed moral value, this does not automatically mean that one will also form the *moral intention* to follow through on one's judgment. Besides moral values, also self-serving values play a role in decision-making, like advancing personal goals and desires. Indeed, moral values are often considered as controlling factors that keep people from only pursuing their short-term selfish desires (Hofmann et al., 2018).

To what extent a moral judgment, based on a moral value, controls for more selfish tendencies and thus to what extent an individual actually sticks to one's moral judgment in a specific decision situation is influenced by many situational and individual factors. Research shows that the social context, like the presence of peers (Warr, 2002), authority figures (Milgram, 1974), or simply being in a hurry (Darley & Batson, 1973) can have a strong influence on whether someone sticks to generally endorsed moral values and corresponding judgments. Also, following Fishbein and Ajzen's (1975) *theory of reasoned action*, people are inclined to follow the subjectively perceived prescriptive norms within a given context. These so-called 'subjective norms', defined as a person's perception of others' expectations and approval of one's behaviour, can be in line with, but may also go against one's considered general moral values and corresponding judgments. This imposes the normative pressure to deviate from them in one's behaviour. Individually, people differ in terms of strength of will (May & Holton, 2012), moral courage (Lachman, 2007) and the extent to which moral values and principles are part of their self-identity (Aquino & Reed ll, 2002; Schlenker et al., 2009), thereby differing in their capacity to 'cling to those values even when faced with pressures to act otherwise.' (M. S. Schwartz, 2016, p761). What becomes clear, is that the extent to which the individual will stick to a generally endorsed moral value in the context of a certain decision situation is quite hard to predict when only relying on a measure of the general endorsement itself. This further contributes to indeterminacy in predicting behaviour from general moral values.

3.4.4 Moral behaviour

Then finally, even in the case a moral intention is established to act upon one's moral judgment, there is still one step to take in order for someone's generally endorsed moral value to affect moral behaviour: actually performing the behaviour itself. Bringing an intention into action involves "working around impediments and unexpected difficulties, overcoming fatigue and frustration, resisting allurements and keeping sight of the eventual goal" (Rest, 1986, p. 15). From this description it becomes clear that many individual- as well as situational factors can play a considerable role. Examples of the first are perseverance and focus to stick to a decision once it is made, especially in long term projects. Situationally, the circumstances can turn out in many degrees of difficulty and complexity, up to the point that they make it just impossible to engage in the intended moral behaviour. In this last step then, from moral intention to moral behaviour, there is still ample possibility for the effect of a generally endorsed moral value on behaviour to perish just before the finish line.

3.4.5 Moral decision-making as a dynamic process

It should be noted, of course, that what appears here in a sequential order and in a demarcated and deliberate fashion can in reality be an intuitive, non-sequential and rather dynamic process of mutual adjustment. Moral awareness and moral judgment may often virtually arise at the same time. Research on basic values and behaviour indicates that the value-behaviour causal relationship can go both ways (Maio, 2016; Maio et al., 2006). Also, theories like cognitive dissonance (Festinger, 1957) make clear that someone's moral judgment and moral value can be adjusted as an effect of an already established intention or performed moral behaviour, instead of the other way around. A vast amount of research on rationalization and neutralization techniques (Sykes & Matza, 1957) shows that people can easily (morally) justify self-serving intentions and behaviours of which they somehow know that these are not (totally) right (e.g. Harris & Daunt, 2011; Johnston & Kilty, 2016; Mckercher et al., 2008). Even initial moral awareness may be rationalized away retroactively by, in second instance, blaming the victim, inspired by self-serving considerations that shape someone's intention. Moral awareness, moral judgment and intention may in this way come up simultaneously, mutually adjusting each other. The potential effect of a general moral value on behaviour can thereby be retroactively annulled or never start off, and the direction of the effect may even be reversed within a specific context of decision-making, possibly further contributing to the indeterminacy of predicting behaviour from generally endorsed moral values. Furthermore, much of the described process may occur intuitively, subconsciously, and within the blink of an eye (although some moral choices can involve long conscious deliberations).[24]

[24] Interesting in this regard is habitual behaviour. It may be argued that habits, harbouring a considered general moral value, evade the different phases of the moral decision-making process and thus its indeterminate effects. Here, it seems important to distinguish habitual behaviour on, at least, two levels. When we consider a developed moral habit on a somewhat abstract level, such as the habit "to help others whenever I can" -comparable to the moral habits which (neo-) Aristotelians would call moral virtues- the different phases of the moral decision-making model still seem relevant. For instance, one still needs to become aware of situations in which help is needed and, possibly, somehow decide between such tendencies and other habits, like "running away when in danger". When habitual behaviour is considered on a more concrete level, so specific automatic behaviour within a particular context, think of "always buying flowers for your neighbour when passing the flower shop on Saturday", or "always ordering meat

3.4.6 Conclusion of the conceptual analysis

Now what can we conclude from this conceptual exposition of the general moral value–concrete moral behaviour relationship? Despite the complexities and the substantial gaps in our knowledge of the moral decision-making process, it has proven insightful to conceptually break down the process of moral decision-making into the described elements and apply it to the general moral value-concrete moral behaviour relationship. This minimal characterization of the relationship shows that someone's generally endorsed moral value needs to go through all four described phases when one enters a certain moral decision situation, before it can have an effect on actual moral behaviour. The outcome of each of these four phases is influenced by many different contextual factors, which render the effect of a general moral value on moral behaviour indeterminate. From this analysis we may expect that general moral values are not strong predictors of concrete moral behaviour. In fact, it may be expected that no or only weak effects are found. Furthermore, the above suggests that when moral values are more specified to the context of the behaviour that they are to predict, this may lead to stronger effects. It can, namely, be expected that these more concrete and specific measures hold more accurate information for a given context on what is considered morally important, i.e., whether one will become aware of this more specific value when it is at stake, or how it holds up against other moral and egoistic values.

In the following, we investigate the empirical relationship between general moral values and concrete moral behaviour in everyday life in three empirical studies. In study 1, we study the relationship between the moral foundations of MFT and participation in voluntary work and informal care. In Study 2, we relate four general moral values provided by MAC to compliance with the nationally proclaimed Covid-19 measures during the 'intelligent' lockdown in The Netherlands in the period of March-May 2020. Finally, in study 3, we look into the effects of general moral foundations on the consumption of meat. Additionally, we compare these to the effects of more specific values -in terms of animal welfare- on meat consumption.

3.5 Methods

3.5.1 Description of the used data sets

For the three conducted empirical studies, we made use of three different datasets. For study 1 and 3, we used existing data by combining measurements from various surveys that have previously been administered among members of the Longitudinal Internet Studies for the Social sciences (LISS) panel -a panel conducted by CentERdata (Tilburg University, The Netherlands). For study 2, we collected our own data. The level of

when dining out", this does seem to surpass the indeterminacy of the moral decision-making model. Indeed, developing concrete habitual behaviour that harbours an endorsed moral value may then be a way to enhance the congruency between one's values and behaviour. At the same time, it seems that such concrete habitual behaviour can also be a source of disconnection between one's values and behaviour. These concrete habits may namely develop in interaction with the social norms within a practical context that do not necessarily reflect one's own values, e.g., think of the habit of scolding on the referee on a football pitch, while this is not in line with your considered general moral values or behaviour in other contexts.

generalization that we aim to make on the basis of this data is that of the Dutch population.

To start with the LISS panel data, this is a representative sample of Dutch individuals who participate in monthly Internet surveys for academic research purposes. The panel is based on a true probability sample of households drawn from the population register. Households that cannot otherwise participate are provided with a computer and Internet connection. Given the data collection procedure, the LISS panel (as a whole) is representative of the Dutch population. Table 1 presents an overview of the surveys that were combined for study 1 and study 3. The surveys were conducted during the end of 2012 and beginning of 2013. The combinations of the surveys yielded different sample sizes for both studies (presented in the final row).[25]

A possible drawback of selecting measurements from multiple surveys is that self-selection biases which may already be present for any individual survey become propagated across multiple surveys. A comparison of the distributions of three socio-demographic characteristics (gender, age and education level) of the different samples with the respective distributions of all LISS panel participants as well as with the Dutch population in 2012, shows that this risk may have indeed manifested itself: especially women and older people are overrepresented, compared to the Dutch population (Table 2). The distribution in terms of level of education is, however, well aligned with the population distribution.

Table 1. Overview of surveys and participants across studies 1, 2 and 3

				Study		
LISS panel survey name	**Data collection period**	**Response (response rate)**	**Measurements**	**1**	**2**	**3**
Social Integration and Leisure, wave 6	February 2013	5,676 (86.0%)	Voluntary behaviour (3 items), providing informal care (1 item)	•		
Consumption decisions and perceptions of animal welfare - Part 2	November 2012	2,648 (87.2%)	MFQ (6 items), animal-specific MFQ (6 items)	•		•
Consumption decisions and perceptions of animal welfare - Part 1	October 2012	3,038 (79.2%)	Consumption of meat (1 item) and meat replacement products (1 item)			•
Own data collection - Compliance to Corona measures 2020	May 2020	N.A.	MAC (12 items), Compliance to Corona measures (10 items)		•	
Sample size (N)				2,320	1,396	2,379

While the bias in terms of gender and age may affect the mean values of the dependent variables in the considered analysis (if these variables significantly influence the dependent variables) there is no reason to expect that the bias will have (large) effects on the estimated relationships between the considered moral values and behaviours. In our study we are explicitly interested in the latter, not the former.

Study 2 is based on a convenience sample, collected by students in the context of a bachelor course running in May 2020 (N=1,396).[26] This gave us the opportunity to

[25] Given that we did not gather the data for study 1 and 3 ourselves, but used existing data, we did not perform power analysis to determine the sample size. The samples that we used for the two studies were constituted by respondents who filled in the two surveys that were combined for each study (see Table 1).

[26] Because we gathered the data following a convenience sampling approach in the context of a bachelor course, no power analysis was performed. To gather the data, students invited their social network to

investigate a very topical moral behaviour that virtually everyone had to deal with in the prior months and that is therefore fresh on people's minds: people's behaviour in relation to nationally proclaimed measures concerning hygiene and social distancing during the 'intelligent' lockdown period (March-May, 2020) of the Corona crisis in The Netherlands. Although it enabled us to collect data relatively fast and thus study a topical phenomenon in a timely fashion, using a convenience sample also runs the high risk of not being representative of the population. As Table 2 shows, comparing our sample to the distributions of the Dutch population in 2019 makes clear that it is indeed biased towards higher educated people and persons in the age group of 15-24. For the same reason as above, however, we believe that the consequences of this bias are limited.

For all three studies, the ethical standards with regards to data collection were met. Informed consent was obtained from all participants. Furthermore, the protocol of Study 2 was accepted by The Human Research Ethics Committee of the university to which we are affiliated. The data collection procedure, conducted by CentERdata, collecting the data for study 1 and 3, abided by the European General Data Protection Regulation (GDPR).

Table 2. The sample distributions of social-demographic characteristics in comparison to LISS panel and population distributions

Variable	Categories	Study 1	Study 2	Study 3	LISS panel	Dutch pop. (2012)	Dutch pop. (2019)
Gender (%)	Male	29	49.6	29	49	49	50
	Female	71	49.9	71	51	51	50
	Other	-	0.5	-	-	-	-
Age (%)	15 - 24 years	2	66	2	16	15	15
	25 - 34 years	8	10	9	13	15	15
	35 - 44 years	16	2	16	16	17	14
	45 - 54 years	20	10	20	18	18	17
	55 - 64 years	25	11	25	18	16	16
	65 years and older	28	2	28	20	19	23
Level of education (%)	Lower education	60	7	59	64	63	59
	Higher education	41	93	41	36	38	41
Sample size (N)		2,320	1,396	2,379			

3.5.2 Dependent variables

When relating moral measures to moral behaviour, a point of discussion is how to determine what counts as 'moral behaviour'. Often the researcher chooses certain forms of behaviour that one regards as clearly morally relevant. However, the researcher may be wrong, in the sense that his or her interpretation of this behaviour as 'moral' does not align with that of the respondents (Meindl & Graham, 2014). Not everyone agrees on what behaviour is actually morally relevant. For instance, some may think that choosing one's mode of transport is a morally relevant choice as it influences one's impact on the environment, while others do not entertain any moral considerations

fill in the survey. The survey was held open for two weeks. The resulting sample was used for the analyses. There were no missing values in the data for all three studies.

when deciding whether to take the bike, bus, train or car. One option that is suggested to tackle this problem is to define moral behaviour from a first-person perspective, that is, to let the respondents themselves decide and report when their behaviour is morally relevant. This option, however, also has a considerable downside. It is vulnerable to socially desirable answers as well as to widespread psychological mechanisms that downplay the moral significance that a subject attributes to their behaviour, such as cognitive dissonance and moral disengagement (Frimer & Walker, 2008). This can lead respondents to not report behaviour as morally relevant, while it clearly is, also according to the general standards of the respondent. As the awareness stage of moral decision-making is an important part of our conceptual model, we find it necessary to include this phase and its possible moral evading mechanisms in our measurements. This information would be missed when we only include behaviours in our models of which the respondents report to have become morally aware. Therefore, we selected as dependent variables a broad set of different daily life behaviours that can all be considered as morally relevant, i.e., where moral considerations can and arguably should play a role. It is for this latter feature that they were deemed suitable for and included in the studies.

Furthermore, as we have noted throughout the paper, we focus in this study on *concrete* moral behaviours. This is to clearly measure distinct phenomena, namely people's general (and specific) moral beliefs and their performed moral behaviour, and not run the risk, described above in reference to Boyd et al. (2015), of plausibly tapping into overlapping sources. By concrete moral behaviours we mean that they have a certain level of specificity, i.e., their description takes the form of concrete acts that are part of certain episodes of behaviour that can be remembered and recounted as such, rather than the form of a general tendency or trait. Compare in this regard 'donating to charity' (Nilsson et al., 2016) to 'I try to help others' (Caprara et al., 2012), where the first kind of description falls within our scope and the latter doesn't. At the same time, we are not aiming at measuring a single behavioural decision by a person. Though concrete, all constructs do reflect behaviour over a certain period of time and thus a behavioural pattern. Even participation in a voluntary organization can usually be assumed to be more than just a one-time action.

In study 1, we focused on participation in voluntary work and providing informal care. As this involves providing help to others without (a large) personal gain, this behaviour has a clear moral component. Participation in voluntary work was operationalized as a dichotomous variable, indicating whether the respondent reported to have performed voluntary work for organizations within one of the following fields: human rights, environmental, and religious (yes/ no) (Table 3). Providing informal care was operationalized as the number of hours during which the respondent reported to have provided informal care per week on average, in the last 12 months (coded in 6 categories, see Table 3).

In study 2, the dependent variable consisted of compliance to the measures proclaimed by The Dutch government during the 'intelligent' lockdown of the Corona crisis March-May 2020. As the measures were meant and presented as an important way to reduce hospitalization and to save lives, we regard it as morally relevant behaviour. Compliance to the Corona measures was operationalized by using items from a national survey, measuring compliance rates with the proclaimed measures conducted by the National Institute of Public Health (RIVM) (www.rivm.nl). We measured on a 5-point

Likert scale (never-always) to what extent people self-reported to comply with certain rules, like washing hands and keeping 1.5m distance from others. In order to bring down the number of models to be estimated, a Principal Component Analysis (PCA) was conducted to summarize the data.[27] This revealed that 9 out of 10 items converged on three distinct components that were logically interpretable: compliance with measures involving personal hygiene, not visiting the most vulnerable groups in society, and general social distancing (Table 4). The one item that did not sufficiently load on any of the three components, asking about coughing and sneezing in the elbow, was left out of the analyses. Regarding reliability of the found components, the social distancing component is with a Cronbach's Alpha of 0.61 just below the 0.7 reliability threshold. As we are here not assuming a latently existing construct, but rather aim to summarize the data in the best way possible (i.e., reduce the data with the least loss of variation), we decided to follow the result given by the PCA and accept the slightly lower reliability for this component. The other components are above the 0.7 reliability threshold (Table 4). For each component the sumscore was computed and included in the analyses.

Table 3. Descriptive statistics of the predicted morally relevant behaviours

Concept	Question and categories	
Voluntary behaviour (study 1)	Whether one has participated in voluntary work for one of the following kinds of organizations: organization for humanitarian aid, human rights, minorities, migrants, environmental protection, peace, animal rights, religious, and/or church	
	0= no (%)	92
	1= yes (%)	8
Providing informal care (study 1)	How many hours of informal care did you provide per week, on average, in the last 12 months?	
	0= 0 (%)	74
	1= 1-8 (%)	18
	2= 9-16 (%)	4
	3= 17-24 (%)	1
	4= 25-32 (%)	1
	5= 33 or more (%)	1
Adherence to Corona measures (study 2)	Can you please indicate to what extent you comply with the following measures of the RIVM? (1-5) mean (SD)	
	Personal hygiene	3.05 (0.91)
	Not visit the vulnerable	3.8 (1.29)
	Social distancing	3.69 (0.83)
Consumption of meat (study 3)	Over the last 4 weeks (28 days), on how many days did you eat chicken meat? mean (SD)	5.5 (4.3)
Consumption of meat replacement products (study 3)	Do you ever eat meat replacement products? Meat replacements products include vegetarian balls or burgers, tofu, soy, tempé, or quorn.	
	1= Never (%)	51
	2= Tasted it once (%)	14
	3= Less than 1 time per month (%)	12
	4= 1 time per month or more often, but less than 1 time per week (%)	12
	5= 1 - 2 times per week (%)	8
	6= 3 - 4 times per week (%)	2
	7= 5 times per week or more often (%)	1

In study 3, we used the consumption of meat and meat replacement products as dependent variables. Studies show that the consumption of meat is regarded as morally

[27] The reason we relied on PCA here and on factor analysis (principal axis factoring) for the MAC-Q-items (and also for the MFQ-items in study 1 and 3), is that for the latter it is assumed that they are reflective of underlying psychological factors. By applying factor analysis the shared variance among the items is extracted, thereby capturing these psychological factors. The Covid-related behaviours are not thought to be caused by underlying factors. Instead, the PCA is meant to provide summary measures that capture most of the variance in the (behavioural) items (not the shared variance).

relevant behaviour by part of the consumer population (Mäkiniemi et al., 2011). The choice to eat less or no meat has also been extensively linked to moral considerations like animal welfare and its ecological impact (Ruby, 2012). (Not) eating meat and eating meat replacement products are therefore considered as behaviours that have a moral component. The consumption of meat was measured in terms of the number of days the respondent reported to have consumed chicken meat during the last four weeks; consumption of meat replacement products was measured on a 7-point Likert scale ranging from never to five times per week or more (Table 3).

Table 4. Rotated component matrix of complying to Corona measures items (study 2)

Questions and items	Component loadings on dimensions		
Can you please indicate to what extent you comply with the following measures of the RIVM? (1=never to 5=always)	**Personal hygiene (α=0.703)**	**Not visit the vulnerable (α=0.850)**	**Social distancing (α=0.609)**
Wash hands often enough	**0.841**	0,039	0,164
Frequently wash hands (more than ten times a day)	**0.835**	0,024	0,020
Wash hands thoroughly (at least 20 seconds)	**0.715**	0,100	0,085
Use paper towels	**0.512**	-0,064	0,250
Do not visit persons older than 70 years old	0,068	**0.923**	0,094
Do not visit persons with ill health	-0,001	**0.913**	0,169
Not shake hands	0,012	0,140	**0.708**
Keep at a sufficient distance from other people (at least 1.5m)	0,203	-0,032	**0.754**
Do not have more than three people visiting	0,218	0,234	**0.717**

Note: Varimax rotation was used to get a simple structure. Number of components extracted was determined based on the component's eigenvalues, where the eigenvalue of 1 was used as the cut-off value. Loadings in bold signify the item's selection for the particular component.

3.5.3 Independent variables

Since we aim at a general investigation of the effects of general moral values on concrete moral behaviour, we made use of two different pre-established scales to operationalize people's general moral values. The Moral Foundations Questionnaire (MFQ) (MoralFoundations.org, 2013; Graham et al., 2011) was used in study 1 and 3. The Morality as Cooperation Questionnaire (MAC-Q) (Curry et al., 2019) was used in study 2.

MFQ has been widely deployed to measure people's general moral values and has been validated across different samples. However, evidence on the structure of the scale does not seem to be fully conclusive. Some studies find evidence for the proposed five-factor structure, though the fit is not always optimal (e.g., Bobbio et al., 2011; Davies et al., 2014; Graham et al., 2011; Nilsson & Erlandsson, 2015; Yilmaz et al., 2016). It has also been suggested that the five-factor structure is further reducible to a higher order two-factor structure, where the care/ harm and fairness/ cheating foundations comprise the higher order individualizing foundation and the loyalty/ betrayal, authority/ subversion, and sanctity/ degradation foundations comprise the higher order binding foundation (Graham et al., 2011; Nilsson & Erlandsson, 2015). This higher order distinction has also been theoretically brought forward by the developers of MFT (Graham et al., 2009). Other studies were not able to reproduce the original five factor structure of the scale. This seems especially the case for samples from non-WEIRD cultures (Atari et al., 2020; Davis et al., 2016; Iurino & Saucier, 2020; but see Doğruyol et al. (2019) for a contrary result). The scale may be regarded as general in nature, i.e., it does not relate to specific moral issues or behaviours.

For study 1 and 3, we used six MFQ-items from the relevance part of the questionnaire, measured on a 6-point Likert scale. Three items are related to the care/harm foundation and three to the fairness/cheating foundation, which together

make up the higher order individualizing foundation. Factor analysis reveals that the data reproduce this higher order structure as the six items converge on one single factor. The scale was found to be sufficiently reliable (Table 5). Sumscores were computed and the constructed variable representing the individualizing foundation was included as an independent variable in the analyses.

Table 5. Factor matrix of MFQ-items (studies 1 and 3)

	Questions and items	Factor loadings
Moral foundations	**When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking? (1='not at all relevant' to 6='extremely relevant')**	**Individualizing moral foundation (α=0.879)**
Care/harm	Whether or not someone suffered emotionally.	0.667
	Whether or not someone cared for someone weak or vulnerable.	0.755
	Whether or not someone was cruel.	0.762
Fairness/cheating	Whether or not some people were treated differently from others.	0.737
	Whether or not someone acted unfairly.	0.763
	Whether or not someone was denied his or her rights.	0.764

Note: The analysis was based on the largest sample (N=2,379, study 3, see Table 2). Respective analysis on the smaller subsample of study 1 yielded a similar result. Number of factors extracted was determined based on the factor's eigenvalues, where the eigenvalue of 1 was used as the cut-off value.

Table 6. Rotated factor matrix of MAC-Q-items

Questions and items	Factor loadings			
To what extent do you agree with the following statements? (1= strongly disagree to 5=strongly agree)	**Group loyalty (α=0.61)**	**Reciprocity (α=0.58)**	**Deference (α=0.49)**	**Fairness (α=0.60)**
People have an obligation to help members of their community.	**0.306**	0.444	0,121	0,158
It's important for individuals to play an active role in their communities.	**0.638**	0,143	0,110	0,096
You should try to be a useful member of society.	**0.603**	0,212	0,110	0,131
You have an obligation to help those who have helped you	0,131	**0.684**	0,137	0,023
You should always make amends for the things you have done wrong.	0,140	**0.362**	0,184	0,198
You should always return a favour if you can.	0,087	**0.501**	0,174	0,077
People should always defer to their superiors.	-0,022	0,163	**0.562**	-0,038
Society would be better if people were more obedient to authority.	0,134	0,084	**0.510**	-0,028
You should respect people who are older than you.	0,139	0,197	**0.339**	0,075
Everyone should be treated the same	0,018	0,137	0,058	**0.732**
Everyone's rights are equally important.	0,078	0,130	-0,025	**0.696**
The current levels of inequality in society are unfair.	0,114	0,016	-0,021	**0.366**

Note: Varimax rotation was used to get a simple structure. Number of factors extracted was determined based on the factor's eigenvalues, where the eigenvalue of 1 was used as the cut-off value. Loadings in bold signify the item's selection for the particular factor.

For study 2, we used 12 items from the MAC-questionnaire. The MAC-questionnaire is a more recently developed measure of morality. Therefore, it still lacks the track record developed by MFQ in terms of longevity and widespread usage. However, it has been developed as a conceptually as well as psychometrically improved tool with respect to MFQ to measure people's moral values (Curry et al., 2019). Like MFQ, also MAC-Q can be regarded as general in nature.

The 12 MAC-Q-items that were used in study 2 belong to four subscales of the judgment part of the questionnaire, with three items for each one. These represent the general moral values of group loyalty, reciprocity, deference, and fairness. The items were measured on a 5-point Likert scale. Factor analysis revealed that our data by and large reproduce the original structure given by MAC. One item ('people have an obligation to help members of their community') did load higher on a different factor (reciprocity) than it originally belongs to (group loyalty). Still, for considerations with regards to content, we decided to keep it as an item of the group loyalty subscale and preserve the original structure. Reliability analyses showed that the Cronbach Alpha's of the four subscales consisting of three items each are below the commonly used 0.7 threshold (Table 6). Sumscores were computed of the four subscales and entered as independent variables in the analyses. To check whether the low reliability of the subscales possibly influenced the results, we ran additional regression models in which the 12 MAC-Q-items were included as individual predictors. These models did not lead to substantially different effects than the ones provided in the Results section below. See the Discussion for a further elaboration on this point.

In study 3 we considered, besides general moral values, also *specific* moral values as independent variables. Here, we used the previously developed items by de Jonge and van Trijp (2014). These researchers formulated three items to reflect the care/ harm and three items to reflect the fairness/ cheating foundation, specific to the context of animal welfare. The items are measured on a 7-point Likert scale (totally agree-totally disagree). Factor analysis reveals that all six items converge on one factor, representing the second-order animal specific individualizing moral foundation (Table 7). The scale was found to be sufficiently reliable and the constructed variable was included in the analyses based on sumscores. For the exact description of the items used in the different studies see Tables 3-7.

Table 7. Factor matrix of the animal-specific moral value items (study 3)

	Questions and items	**Factor loadings**
Animal-specific moral foundations	**Please indicate to what extent you agree with the following statements (1='totally disagree to 7='totally agree')**	**Animal-specific individualizing moral foundation (α=0.779)**
Care/harm	I don't care for animal welfare issues. (reverse coded)	0.585
	I feel a strong emotional bond with animals.	0.610
	People exaggerate the feelings and sensitivity of animals. (reverse coded)	0.609
Fairness/cheating	Animals should be protected for their own sake, rather than simply serving the needs of humans.	0.643
	I believe that society has a moral obligation to promote animal welfare.	0.711
	In principle, we as humans have the right to use animals however we want to. (reverse coded)	0.520

Note: Number of factors extracted was determined based on the factor's eigenvalues, where the eigenvalue of 1 was used as the cut-off value.

3.5.4 Social-demographic characteristics

Finally, in each model, gender, age, and level of education were included to control for possible spurious effects caused by these social-demographic characteristics. Age was entered as a continuous variable. For study 1 and 3, gender was entered as a dichotomous dummy variable. Study 2 included a third category for gender, 'other', and therefore two dummy variables were created with male as the reference category to include them in the analysis. For level of education, we created a dichotomous dummy variable and converted all scores to this scale (0=low/ 1=high). Lower education consists of the following levels: primary school; intermediate secondary education; intermediate

vocational education. Higher education consists of the levels: higher secondary education; higher vocational education; university.

3.5.5 Analysis strategy

To investigate the relationships between the independent and dependent variables we made use of a binary logistic regression analysis and of multiple linear regression analyses. We used the binary logistic regression analysis for the prediction of participation in voluntary behaviour (study 1), as it is operationalized as a dichotomous variable. For the prediction of the other dependent variables, we made use of linear regression models even though most of the dependent variables were measured on Likert scales. While these outcomes are best modelled using ordinal regression models, we relied on the more straightforward linear models, because they are more easily interpreted and because they can provide standardized estimates, indicating the relative importance of the explanatory variables. For our investigation of the relationship between general moral values and behaviour we are most interested in these relative effect sizes of our moral predictors. To make sure that our choice for using linear regression models instead of ordinal regression models has not affected our results and conclusions, we ran additional ordinal regression models for each dependent variable. We found that the estimates of the linear regression models (proportionally) match those of the ordinal regression models[28]. We can therefore rely on the presented linear models.

As it is an assumption of factor analysis and PCA, we assume that our variables are normally distributed. A check whether this is the case, indeed showed that the vast majority of the variables approach the normal distribution, with kurtosis and skewness estimates between -1 and 1. As we make use of a large sample (n>1000), the fact that not all variables are normally distributed does not influence our estimates for factor analysis and PCA (Muthén & Kaplan, 1985).

Regarding the linear regression models, for each dependent variable we estimated two consecutive models. In the first, we entered only the social-demographic characteristics, in the second we added the moral value variables to detect the additional effect[29].

3.6 Results

In study 1, we investigated the effects of general moral values on participation in voluntary behaviour and providing informal care. Table 8 provides the estimates of the binary logistic regression model, estimating the effects of the social demographic characteristics and the individualizing moral foundation (focusing on care and fairness) on participation in voluntary behaviour. The coefficient for the individualizing moral

[28]For reasons of brevity the ordinal regression models are not presented here. The ordinal regression models can be found in the supplementary material of the published article (van den Berg et al., 2022) on which this chapter is based.
Link: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.817860/full#supplementary-material.

[29] All codes of the conducted analyses can be found in the supplementary material of the published article on which this chapter is based (van den Berg et al., 2022). Link:
https://www.frontiersin.org/articles/10.3389/fpsyg.2022.817860/full#supplementary-material

foundation (0.024) is non-significant, providing no evidence that scores on this moral dimension are associated with participation in voluntary behaviour. Table 9 presents the standardized coefficients of the linear regression model explaining informal care. Also here (Model 2), the found coefficient for the individualizing moral foundation (0.027) is non-significant. Hence, no evidence is found that people's scores on this moral dimension influences the extent to which people provide informal care.

Table 8. Coefficients of the binary logistic regression model predicting voluntary behaviour (study 1)

Dependent variable: participation in voluntary behaviour		
Independent variables	**Estimates**	**p-value (two-sided)**
Gender (female)	0,243	0,166
Age	0,015	0,007
Level of education (high)	0,483	0,002
Individualizing moral foundation	0,024	0,142
Constant	-4,309	0,000

Table 9. Coefficients of regression models predicting informal care (study 1)

Dependent variable: providing informal care	**Model 1**				**Model 2**			
Independent variables	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**
Gender (female)	0.057	0.006	0.111	0.040	0.055	0.008	0.107	0.040
Age	0.113	0.000	0.007	0.001	0.109	0.000	0.007	0.001
Level of education (high)	-0.013	0.543	-0.023	0.038	-0.017	0.434	-0.030	0.038
Individualizing moral foundation					0.027	0.196	0.005	0.004
R-square (sign. change)	0.016	(0.000)			0.017	(0.196)		

Note: no multicollinearity was found among the independent variables, all VIF values are between 1 and 2

Similarly, the results of study 2 show weak associations between general moral values and moral behaviours (adherence to the Corona measures) (models 4, 5, and 6 of Table 10). Though the effects are somewhat higher than found in Study 1, and several are statistically significant at the 5%-level. The largest two effects, both positive, are the endorsement of the moral value of fairness on adherence to the personal hygiene measures (beta of 0.10) and on adherence to the social distancing measures (beta of 0.15). This means that as one finds fairness more important, one tends to adhere slightly more strictly to the two kinds of corona measures. The moral value of group loyalty has a significant but very weak effect on the adherence to all three forms of corona measures (beta's of 0.076, 0.061, and 0.067). This suggests that caring about one's community may play a (small) role in the decision to conform to the imposed measures. The above findings seem intuitive, as both fairness and group loyalty can intelligibly motivate conforming to measures, imposed for the benefit of us all. The very weak, but statistically significant, negative effect for reciprocity on 'not visiting the vulnerable' seems less intuitive.

With 0.1 and 0.15 being the largest effects and the other effects being well under 0.1, the effects should overall be considered as weak. The R-square change of the models 4, 5, and 6 confirm that, although statistically significant, adding the general moral values as predictors to the social-demographics explains only a small additional amount of variance (1.8%, 1.3%, and 2.7% respectively).

Table 10. Coefficients of regression models predicting compliance to national Corona measures (study 2)

Dependent variable: personal hygiene		**Model 1**				**Model 4**		
Independent variables	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**
Gender (female)	0.219	0.000	0.399	0.047	0.198	0.000	0.363	0.047
Gender (other)	-0.012	0.652	-0.150	0.333	-0.010	0.694	-0.130	0.331
Age	0.172	0.000	0.010	0.002	0.172	0.000	0.010	0.002
Level of education (high)	-0.057	0.031	-0.201	0.093	-0.060	0.024	-0.212	0.094
Group loyalty					0.076	0.009	0.118	0.045
Reciprocity					-0.013	0.644	-0.019	0.041
Deference					0.011	0.689	0.015	0.038
Fairness					0.100	0.000	0.134	0.036
R-square (sign. change)	0.089	(0.000)			0.107	(0.000)		
Dependent variable: not visit the vulnerable		**Model 2**				**Model 5**		
Independent variables	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**
Gender (female)	-0.013	0.613	-0.034	0.068	-0.028	0.289	-0.073	0.069
Gender (other)	-0.024	0.362	-0.438	0.481	-0.025	0.338	-0.458	0.479
Age	-0.168	0.000	-0.014	0.002	-0.177	0.000	-0.015	0.002
Level of education (high)	0.091	0.001	0.452	0.135	0.081	0.003	0.403	0.136
Group loyalty					0.061	0.042	0.133	0.065
Reciprocity					-0.001	0.968	-0.002	0.060
Deference					-0.053	0.062	-0.103	0.055
Fairness					0.082	0.003	0.155	0.052
R-square (sign. change)	0.044	(0.000)			0.057	(0.001)		
Dependent variable: social distancing		**Model 3**				**Model 6**		
Independent variables	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**
Gender (female)	0.130	0.000	0.215	0.043	0.102	0.000	0.168	0.043
Gender (other)	0.057	0.028	0.667	0.303	0.056	0.029	0.654	0.299
Age	0.259	0.000	0.014	0.001	0.256	0.000	0.014	0.001
Level of education (high)	0.092	0.001	0.295	0.085	0.089	0.001	0.285	0.085
Group loyalty					0.067	0.023	0.093	0.041
Reciprocity					-0.065	0.026	-0.083	0.037
Deference					0.013	0.644	0.016	0.035
Fairness					0.147	0.000	0.179	0.033
R-square (sign. change)	0.082	(0.000)			0.109	(0.000)		

Note: no multicollinearity was found among the independent variables, all VIF values are between 1 and 2

In study 3, we investigated both the effects of general moral values as well as specific moral values -tailored to the context of animal welfare- on the consumption of meat and meat replacement products. Table 11 (models 3 and 4) shows that the effect of the individualizing moral foundation is non-significant for the frequency of eating chicken meat, and very weak (with a beta of 0.07) for the frequency of eating meat replacement products. Models 5 and 6 show that the context specific individualizing

moral foundation (related to animal welfare) is a stronger predictor than its generic counterpart. Its effect on the number of days eating chicken meat is statistically significant at the 5%-level, but is still very weak (beta of -0.07); its effect on eating meat replacement products is substantially stronger than the general moral value (beta of 0.23, significant at 1%-level). Including the specific moral value in model 6 attenuates the initial statistically significant effect of the general moral value on eating meat replacement products (model 4) downwards, rendering it statistically insignificant. This indicates that the initially detected effect of the general moral value is actually explained by its specific counterpart.

In sum, the found effects of general moral values on the considered moral behaviours must be regarded as weak to very weak. Most effects are well under 0.10, while many do not reach statistical significance at the conventional 5%-level. For the models considered, the general moral values are only able to explain 2.7% of variance, at best. Study 3 indicates that context specific values are somewhat stronger predictors than their general counterparts.

Table 11. Coefficients of regression models predicting consumption of meat and meat replacement products (study 3)

Dependent variable: consumption of meat	**Model 1**				**Model 3**				**Model 5**			
Independent variables	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**
Gender (female)	0.017	0.399	0.164	0.194	0.017	0.416	0.158	0.195	0.024	0.246	0.227	0.196
Age	-0.129	0.000	-0.038	0.006	-0.130	0.000	-0.038	0.006	-0.131	0.000	-0.039	0.006
Level of education (high)	-0.007	0.719	-0.066	0.182	-0.009	0.678	-0.076	0.184	-0.009	0.670	-0.078	0.183
Individualizing moral foundation					0.009	0.665	0.008	0.017	0.025	0.247	0.021	0.018
Animal-specific individualizing moral foundation									-0.066	0.002	-0.047	0.015
R-square (sign. change)	0.017	0.000			0.017	0.665			0.021	0.002		
Dependent variable: consumption of meat replacement products	**Model 2**				**Model 4**				**Model 6**			
Independent variables	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**	**Beta**	**p-value (two-sided)**	**B**	**Std. Error**
Gender (female)	0.071	0.000	0.240	0.068	0.067	0.001	0.224	0.068	0.042	0.035	0.140	0.066
Age	-0.037	0.071	-0.004	0.002	-0.046	0.025	-0.005	0.002	-0.042	0.035	-0.004	0.002
Level of education (high)	0.202	0.000	0.628	0.064	0.192	0.000	0.596	0.064	0.192	0.000	0.598	0.062
Individualizing moral foundation					0.074	0.000	0.022	0.006	0.020	0.322	0.006	0.006
Animal-specific individualizing moral foundation									0.225	0.000	0.057	0.005
R-square (sign. change)	0.047	0.000			0.052	0.000			0.099	0.000		

Note: no multicollinearity was found among the independent variables, all VIF values are between 1 and 2

3.7 Discussion

Our results show that in all three conducted empirical studies, only weak to very weak effects between general moral values and concrete moral behaviours were found. Overall, the largest effect found within the linear regression models was 0.15, while the large majority of the effects were well under 0.1. Many did not reach statistical significance at the conventional 5%-level. Adding the general moral values as predictors to the models consisting of only the social-demographic characteristics resulted in explaining only very small additional amounts of variance of the behaviour. These

findings suggest that general moral values are poor predictors of people's concrete moral behaviour.

The findings in study 3 lend support for the idea that a reason for these weak associations may be sought in the context specificity of moral decision-making. Here, we found larger effects for moral values that were tailored to the context of animal welfare than we found for their generic counterparts. In addition, inclusion of the more specific moral value to the model rendered the (initially found) effect of the general moral value insignificant. This can be explained by the notion that more specific moral values may harbour more accurate information about what is found morally important in a certain kind of context of decision-making, i.e., about how it holds up against factors that may hamper awareness and against other moral and egoistic values that can play a role.

These empirical results are in line with the derived expectations from our conceptual analysis. There, we stated that for a general moral value to act as a predictor of a concrete form of behaviour, its presence within an individual would have to (regularly) lead to this behaviour. Our conceptual analysis -one we argued to be acceptable to 'rationalists' as well as 'intuitionists'- suggests that this is unlikely. Before someone's general moral value can determine behaviour in a given situation it needs to go through a process of moral decision-making which consists of four different phases. These all need to become manifest in such a way that they harbor the potential influence of one's general moral value. When looking into this process, it becomes clear that each phase is influenced by contextual factors, which are not taken into account in a general measurement of moral values and which can potentially annul their influence. We therefore expected to find rather weak effects when predicting concrete moral behaviour from general moral values and higher effects for more specific moral values, corresponding to what was found in the empirical studies.

Now the question remains what these results can tell us about predicting concrete moral behaviour from general moral values and about the influence of morality on behaviour more generally. The results of our empirical study confirm the expectations derived from the conceptual model. They also line up well with previously reported empirical findings on general moral values in relation to moral behaviour and with findings from studies on values and behaviour more generally. Together, this points towards the notion that general moral values are, in fact, poor predictors of concrete moral behaviours. Our results furthermore suggest that a possible reason for this, is that morality's influence on behaviour may be more context specific than a general questionnaire can grasp. However, we do need to be cautious about drawing too strong conclusions just yet. For our empirical study, we made use of data and data collection methods that were readily available to us. As such, our findings are subject to limitations (but also strengths) that need to be taken into account.

First of all, the samples used in the different studies are not representative of the population. For study 1 and 3, this may be due to selection effects as the samples were composed of respondents who completed both surveys that were combined for each study. In study 2, the bias is likely due to the use of convenience sampling. As mentioned in the Method section, we think this has a limited influence on our results regarding the relationship between general moral values and behaviour. The bias most directly affects the estimation of means of the dependent variables, which is not the focus of this study. Though it is possible that the relationship between moral values and behaviour for

younger or more highly educated people is different than for the Dutch population in general, we have no reason to expect this, also given that no theory of morality argues that morality only influences behaviour for select demographic subpopulations.

A strong asset of our study, in terms of preventing bias, is that the measurements for the independent and dependent variables used in study 1 and 3 (retrieved from the LISS panel data) were collected in different instances with a substantial time in between. This avoids the risk of inducing associations due to the measurements being part of the same survey. For example, if a person just completed a set of items related to general moral values and is then asked whether he or she is engaged in specific forms of moral behaviour, the person may be inclined to provide answers that are consistent with his or her stated moral values, thereby inflating the correlations. For study 2, this is a possible limitation as the moral value items and Covid-19 behaviour items were part of the same survey. Note, however, that if this has indeed manifested itself, then the actual relationship between moral values and conforming to the considered Covid-19 measures is even weaker than found in our study, which is in support of our main conclusions.

Another limitation of our study is that not all the subscales of the original scales of MFQ and MAC were available for our analyses, due to using existing data (study 1 and 3) and space limitations in the conducted survey (study 2). Specifically, concerning MFQ (study 1 and 3), we miss measurements for the binding foundations of loyalty, authority, and sanctity; concerning MAC (study 2) we miss measurements for the values of family, property and heroism. Although we believe that the moral values that were included seem relevant to the behaviours they were to predict (e.g., care with regards to voluntary work and eating meat; deference with regards to compliance behaviour with rules), this does limit our empirical findings and corresponding conclusions to the combinations of moral values and behaviours that were studied. Space limitations were also the reason to not include MFQ-items in the survey of study 2, but limit it to MAC-Q-items.

In addition, for the same reasons as stated above, not all the items per used subscale were available for MFQ and MAC. The latter may be the reason for finding low reliabilities for the MAC's subscales in our data, which could also be (partly) the reason for finding small effects between these scales and behaviour in study 2. To check whether the outcomes of study 2 were sensitive to using lesser reliable scales, we ran additional regression analyses including all the 12 MAC-items separately. The results of these analyses provided a similar picture as the results of the analyses based on the MAC-constructs, which were presented in the Results section. We found only weak associations and few significant effects between the separate items and the compliance level of the different types of Covid19-measures.[30] This result suggests that the found weak effects are not primarily due to the low reliability of the used scales.

Another issue due to using existing data for study 1 and 3, is that we were somewhat restricted in our choice of the dependent variables. Particularly, though eating chicken meat and eating meat replacement products are instances of (not) eating meat and, therefore, morally relevant behaviours in itself, it would have been more ideal to measure to what extent people (do not) eat meat overall. Especially the eating chicken meat-item is vulnerable to critique in this regard, as eating less chicken meat

[30]These results are not reported here, for brevity, but are published as supplementary material to the published article (van den Berg et al., 2022) on which this chapter is based. Link: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.817860/full#supplementary-material

could also mean that people eat more pork or beef. Also, eating chicken meat may be regarded by some respondents as morally better behaviour than eating beef or pork. The weak effect found between the moral value and eating chicken meat may be due to this possible moral ambiguousness of the behavioural item. Whether this is the case can be easily investigated in future research, by using a more encompassing item for measuring individual meat consumption.

A final point concerns the choice of moral behaviour more generally. As explained in the Method section, there may of course be discussion as to what extent these are actually morally relevant behaviours, pertaining to the discussion on imposing morally relevant behaviour upon respondents versus having the respondent indicate what he or she deems as moral behaviour. There, we substantiated our choice for the first option. However, this does mean that behaviour which we have indicated as morally relevant may not be viewed as such by respondents. For instance, '(not) eating meat' or '(not) washing hands during the Covid19 pandemic' may not be regarded by everybody as morally relevant. Partly, this can be due to not becoming aware that such behaviour is actually relevant to one's moral values. In this case, as explained, these findings are relevant to our conclusions about the relationship between general moral values and behaviour. However, if people generally view such behaviours as, for example, health issues, submitted to egoistic considerations rather than moral ones, then our findings say less about the influence of moral values on moral behaviour, i.e., we then did not test moral values against genuine moral behaviour. As it is possible to question the moral relevance of virtually any behaviour to some degree, this seems to be a deeper and almost inevitable issue for the study of moral behaviour in general, in particular for studies that choose -for possibly good reasons- to impose what behaviour is morally relevant. This dilemma does oblige researchers to sufficiently substantiate this choice, and, in case of imposing the moral behaviours, to substantiate their choices of behaviour as well. We hope to have done this to a sufficient degree. Another way to approach this problem is to select a variety of morally relevant behaviours that reflects the rich palette of moral behaviours that exist. We have made an attempt at this in our choices for this study.

In sum, especially the somewhat ad-hoc character of the selection of moral value- and behaviour-items and their accompanying limitations suggests that further research is needed before we can be fully conclusive about whether general moral values are indeed poor predictors of concrete moral behaviour in everyday life. Preferably, such research would have to include all and complete subscales and a broad range of selected behaviours. Furthermore, it would be interesting to see whether our findings hold up in more representative samples, as we expect. Another interesting research direction is to use observational data for behaviour that is eligible to observe in 'real life', as complementary to self-reports (see e.g., Nilsson et al., (2016) and O'Grady et al. (2019) for such efforts).

That said, in light of the fact that our results are in line with studies which have related general moral values to specific forms of moral behaviour, as well as with findings from studies on the broader concept of basic values and behaviour, and considering that our conceptual framework makes these results intelligible, we do believe that evidence is building up towards the conclusion that general moral values are poor predictors of concrete moral behaviours in daily life. Now let's assume, for a moment, that we are on the right track with the idea that knowing the moral values that

people endorse in general cannot tell us much about their more concrete behaviours. Does this imply that morality or moral values barely influences individual decision-making? And, in its wake, does this mean a blow to the evolutionary foundations and the usefulness of empirical moral value theories? The answer to these questions is left for further research efforts, but we here offer one possible direction.

Our analysis suggests that the contextual aspect of moral decision-making does not square well with the prevailing method used to measure moral values (which is done in a context-free manner through a generic questionnaire, like MFQ and MAC-Q). More specifically, the idea -inevitably assumed by this method - of general moral values being relatively stable personal dispositions that cause a similar kind of behaviour over different contexts, does not seem to align with the more dynamic role played by moral values arising from our conception of moral decision-making. In reality, in different specific decision situations a person can become aware of different moral values being at stake, if any; can make different appraisals between moral values as well as between moral values and self-serving considerations; and can make a different assessment of what actions are feasible. This points to a moral influence that is context dependent and possibly much more dynamic than general measurements are able to detect.

Of course, adding more general measures to the model, like people's moral identity (Aquino & Reed ll, 2002; Smith et al., 2014), religiosity (Huber & Huber, 2012), or people's more egoistic values (Seuntjens et al., 2015), may lead to better predictions. This is another aspect that needs to be addressed in further research. However, there are at least two reasons to worry about whether this may solve the indeterminate effect of morality with regards to behaviour. First of all, these added variates would again consist of general measures, while things like following through on one's moral considerations may prove to be rather contextual (Ross & Nisbett, 2011). Secondly, if our elaboration on the process of moral decision-making is on the right track, the indeterminacy seems to be more fundamental. Specifically, in the moral judgment phase, moral values or moral values-inspired considerations are pitted against each other. This seems to constitute an indeterminacy in the heart of the moral endeavour itself. This is no question of whether one is inclined to become aware of the moral relevance of a situation or whether one wants to act morally at all, but how to morally relate to the situation at hand. To get a better understanding of how moral agents weigh considerations within a specific decision situation seems to ask for other measurement techniques in which such weighing has a place.

In line with this, to get a better grasp of the influence of morality on decision-making and action, it is important to know what specific moral values and moral considerations become important within a certain context. This asks for measuring people's local moral considerations within a certain delineated context or scenario (for an interesting example, see Navarick & Moreno (2022) focusing on people's moral choices within the delineated setting of Covid-19 triage dilemmas in the hospital). A further interesting direction in this regard would be to study people's specific moral considerations across different contexts and see whether and to what extent these, as well as their underlying values, alter per situation.

In conclusion, the fact that very small effects were found for general moral values and somewhat larger effects for more specific moral values, suggests that people's moral beliefs can influence decision-making, but that context matters and that this needs to be reflected in the measuring method. This relates to a fundamental requirement for a valid

measurement instrument: it should trigger the same kind of behavioural mechanisms in the measuring process compared to what happens in the real world. To move forward in our understanding of the relationship between morality and behaviour this seems crucial. In other words, to know to what extent and how morality influences concrete forms of behaviour, such as conforming to Covid-19 measures or eating meat, simply measuring people's general moral values does not seem the best way to go. Our study suggests that improvement lies in using and developing methods that can better incorporate the contextual aspect of moral decision-making when measuring people's morality and studying its influence on behaviour.

References

Aquino, K., & Reed ll, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Atari, M., Graham, J., & Dehghani, M. (2020). Foundations of morality in Iran. *Evolution and Human Behavior*, *41*(5), 367–384.

Bandura, A., Barbaranelli, C., & Caprara, G. V. (1996). Mechanisms of Moral Disengagement in the Exercise of Moral Agency. *Journal of Personality and Social Psychology*, *71*(2), 364–374.

Bardi, A., & Schwartz, S. H. (2003). Values and Behavior: Strength and Structure of Relations. *Personality and Social Psychology Bulletin*, *29*(10), 1207–1220. https://doi.org/10.1177/0146167203254602

Bobbio, A., Nencini, A., & Sarrica, M. (2011). Il Moral Foundation Questionnaire: Analisi della struttura fattoriale della versione italiana. *Giornale Di Psicologia*, *5*(1), 7–18.

Boyd, R. L., Wilson, S. R., Pennebaker, J. W., Kosinski, M., Stillwell, D. J., & Mihalcea, R. (2015). Values in words: Using language to evaluate and understand personal values. *Proceedings of the 9th International Conference on Web and Social Media, ICWSM 2015*, 31–40.

Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning. *Judgment and Decision Making*, *3*(2), 121–139.

Caprara, G. V., Alessandri, G., & Eisenberg, N. (2012). Prosociality: The contribution of traits, values, and self-efficacy beliefs. *Journal of Personality and Social Psychology*, *102*(6), 1289–1303. https://doi.org/10.1037/a0025626

Cieciuch, J. (2017). Exploring the complicated relationship between values and behaviour. In S. Rochas & L. Lilach (Eds.), *Values and Behavior: Taking a Cross-Cultural Perspective* (pp. 237–247). Heidelberg: Springer.

Clark, C. B., Swails, J. A., Pontinen, H. M., Bowerman, S. E., Kriz, K. A., & Hendricks, P. S. (2017). A behavioral economic assessment of individualizing versus binding moral foundations. *Personality and Individual Differences*, *112*, 49–54. https://doi.org/10.1016/j.paid.2017.02.052

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology*, *107*(5), 943–963. https://doi.org/10.1037/a0037245

Craft, J. L. (2013). A Review of the Empirical Ethical Decision-Making Literature: 2004-2011. *Journal of Business Ethics*, *117*(2), 221–259. https://doi.org/10.1007/s10551-012-1518-9

Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008

Darley, J. M., & Batson, C. D. (1973). " From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*(1), 100–108.

Davies, C. L., Sibley, C. G., & Liu, J. H. (2014). Confirmatory factor analysis of the Moral Foundations Questionnaire. *Social Psychology*, *45*(6), 431–436.

Davis, D. E., Rice, K., Van Tongeren, D. R., Hook, J. N., DeBlaere, C., Worthington, E. L., & Choe, E. (2016). The moral foundations hypothesis does not replicate well in black samples. *Journal of Personality and Social Psychology*, *110*(4), e23–e30. https://doi.org/10.1037/pspp0000056

de Jonge, J., & van Trijp, H. (2014). Heterogeneity in consumer perceptions of the animal friendliness

of broiler production systems. *Food Policy*, *49*(P1), 174–185. https://doi.org/10.1016/j.foodpol.2014.07.008

Díaz, R., & Cova, F. (2022). Reactance, morality, and disgust: the relationship between affective dispositions and compliance with official health recommendations during the COVID-19 pandemic. *Cognition and Emotion*, *36*(1), 120–136. https://doi.org/10.1080/02699931.2021.1941783

Dickinson, J. L., McLeod, P., Bloomfield, R., & Allred, S. (2016). Which Moral Foundations Predict Willingness to Make Lifestyle Changes to Avert Climate Change in the USA? *PLOS ONE*, *11*(10), 1–11. https://doi.org/10.1371/journal.pone.0163852

Doğruyol, B., Alper, S., & Yilmaz, O. (2019). The five-factor model of the moral foundations theory is stable across WEIRD and non-WEIRD cultures. *Personality and Individual Differences*, *151*(July), 109547. https://doi.org/10.1016/j.paid.2019.109547

Duc Huynh, T. L. (2020). Replication: Cheating, loss aversion, and moral attitudes in Vietnam. *Journal of Economic Psychology*, *78*(March), 102277. https://doi.org/10.1016/j.joep.2020.102277

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017. *Personality and Social Psychology Review*, *23*(4), 332–366. https://doi.org/10.1177/1088868318811759

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behaviour: an introduction to theory and research*. Reading, MA: Addison-Wesley.

Frimer, J. A., & Walker, L. J. (2008). Towards a new paradigm of moral personhood. *Journal of Moral Education*, *37*(3), 333–356. https://doi.org/10.1080/03057240802227494

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental Social Psychology, 47*, 55-130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and Conservatives Rely on Different Sets of Moral Foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Meindl, P., & Beall, E. (2012). Integrating the Streams of Morality Research: The Case of Political Ideology. *Current Directions in Psychological Science*, *21*(6), 373–377. https://doi.org/10.1177/0963721412456842

Graham, J., Nosek, B., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385.

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., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. *The Innate Mind*, *3*, 367–391.

Hanel, P. H. P., Vione, K. C., Hahn, U., & Maio, G. R. (2017). Value instantiations: the missing link between values and behavior? In S. Roccas & S. Lilach (Eds.), *Values and behavior: Taking a Cross-Cultural Perspective* (pp. 175–190). Heidelberg: Springer.

Harris, L. C., & Daunt, K. L. (2011). Deviant customer behaviour: A study of techniques of

neutralisation. *Journal of Marketing Management*, *27*(7–8), 834–853.

Haslam, N. (2006). Dehumanization: An Integrative Review. *Personality and Social Psychology Review*, *10*(3), 252–264. papers2://publication/uuid/014DC7A5-77D8-47BE-9341-D730C2237A67

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65*, 399–423. https://doi.org/10.1146/annurev-psych-010213-115045

Hofmann, W., Meindl, P., Mooijman, M., & Graham, J. (2018). Morality and Self-Control: How They Are Intertwined and Where They Differ. *Current Directions in Psychological Science*, *27*(4), 286–291. https://doi.org/10.1177/0963721418759317

Hoover, J., Dehghani, M., Johnson, K., Iliev, R., & Graham, J. (2019). Into The Wild: Big Data Analytics in Moral Psychology. In K. Gray & J. Graham (Eds.), *Atlas of Moral Psychology* (pp. 525–541). Guilfords Publications.

Huber, S., & Huber, O. W. (2012). The Centrality of Religiosity Scale (CRS). *Religions*, *3*(3), 710–724. https://doi.org/10.3390/rel3030710

Iurino, K., & Saucier, G. (2020). Testing Measurement Invariance of the Moral Foundations Questionnaire Across 27 Countries. *Assessment*, *27*(2), 365–372. https://doi.org/10.1177/1073191118817916

Johnston, M. S., & Kilty, J. M. (2016). "It's for their own good": Techniques of neutralization and security guard violence against psychiatric patients. *Punishment & Society*, *18*(2), 177–197.

Jones, T. M. (1991). Ethical Decision Making by Individuals in Organizations : An issue-contingent model. *Academy of Management Review*, *16*(2), 366–395.

Kasachkoff, T., & Saltzstein, H. D. (2008). Reasoning and moral decision-making: A critique of the social intuitionist model. *International Journal of Developmental Sciences*, *2*(3), 287–302. https://doi.org/10.3233/DEV-2008-2307

Kelman, H. C. (1976). Violence without restraint: Reflections on the dehumanization of victims and victimizers. In G. M. Kren & L. H. Rappoport (Eds.), *Varieties of psychohistory* (pp. 282–314). New York: Springer.

Kennett, J., & Fine, C. (2009). Will the real moral judgment please stand up? : the implications of social intuitionist models of cognition for meta-ethics and moral psychology. *Ethical Theory and Moral Practice*, *12*(1), 77–96. https://doi.org/10.1007/s10677-008-9136-4

Kohlberg, L. (1984). *The Psychology of Moral Development*. Cambridge, MA: Harper and Row.

Kroesen, M., & Chorus, C. (2018). The role of general and specific attitudes in predicting travel behavior – A fatal dilemma? *Travel Behaviour and Society*, *10*(June), 33–41. https://doi.org/10.1016/j.tbs.2017.09.004

Kugler, M., Jost, J. T., & Noorbaloochi, S. (2014). Another Look at Moral Foundations Theory: Do Authoritarianism and Social Dominance Orientation Explain Liberal-Conservative Differences in "Moral" Intuitions? *Social Justice Research*, *27*(4), 413–431. https://doi.org/10.1007/s11211-014-0223-5

Lachman, V. D. (2007). Moral courage: a virtue in need of development? *Medsurg Nursing : Official Journal of the Academy of Medical-Surgical Nurses*, *16*(2), 131–133.

Latane, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* New York: AppletonCentury-Crofts.

Lee, J. A., Bardi, A., Gerrans, P., Sneddon, J., Van Herk, H., Evers, U., & Schwartz, S. (2021). Are value--behavior relations stronger than previously thought? It depends on value importance.

European Journal of Personality, *36*(2), 133–148. https://doi.org/https://doi.org/10.1177/08902070211002965

Levitt, S. D., & List, J. A. (2007). About the Real World ? *Journal of Economic Perspectives*, *21*(2), 153–174.

Maio, G. R. (2016). *The psychology of human values*. London: Psychology Press.

Maio, G. R., Olson, J. M., Bernard, M. M., & Luke, M. A. (2006). Ideologies, values, attitudes, and behavior. In J. Delamater (Ed.), *Handbook of social psychology* (pp. 283–308). Springer.

Mäkiniemi, J. P., Pirttilä-Backman, A. M., & Pieri, M. (2011). Ethical and unethical food. Social representations among Finnish, Danish and Italian students. *Appetite*, *56*(2), 495–502. https://doi.org/10.1016/j.appet.2011.01.023

May, J., & Holton, R. (2012). What in the world is weakness of will? *Philosophical Studies*, *157*(3), 341–360. https://doi.org/10.1007/s11098-010-9651-8

Mckercher, B., Weber, K., & Cros, H. (2008). Rationalising Inappropriate Behaviour at Contested Sites Contested Sites. *Journal of Sustainable Tourism*, *16*(4), 369–385. https://doi.org/10.2167/jost778.0

Meindl, P., & Graham, J. (2014). Know thy participant: The trouble with nomothetic assumptions in moral psychology. In H. Sarkissian & J. C. Wright (Eds.), *Advances in experimental moral psychology* (pp. 233–252). A&C Black.

Miles, A. (2015). The (Re)genesis of Values: Examining the Importance of Values for Action. *American Sociological Review*, *80*(4), 680–704. https://doi.org/10.1177/0003122415591800

Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.

MoralFoundations.org. (2013). Retrieved October 13, 2021, from https://moralfoundations.org

Musschenga, A. W. (2009). Moral intuitions, moral expertise and moral reasoning. *Journal of Philosophy of Education*, *43*(4), 597–613. https://doi.org/10.1111/j.1467-9752.2009.00707.x

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*(2), 171–189.

Navarick, D. J., & Moreno, K. M. (2022). Moral Dilemmas in Hospitals: Which Shooting Victim Should be Saved? Running head: Moral Dilemmas in Hospital Care. *Frontiers in Psychology*, 13:770020. https://doi.org/doi: 10.3389/fpsyg.2022.770020

Nilsson, A., & Erlandsson, A. (2015). The Moral Foundations taxonomy: Structural validity and relation to political ideology in Sweden. *Personality and Individual Differences*, *76*, 28–32. https://doi.org/10.1016/j.paid.2014.11.049

Nilsson, A., Erlandsson, A., & Västfjäll, D. (2016). The congruency between moral foundations and intentions to donate, self-reported donations, and actual donations to charity. *Journal of Research in Personality*, *65*, 22–29. https://doi.org/10.1016/j.jrp.2016.07.001

O'Fallon, M. J., & Butterfield, K. D. (2005). A review of the empirical ethical decision-making literature: 1996-2003. *Journal of Business Ethics*, *59*(4), 375–413. https://doi.org/10.1007/s10551-005-2929-7

O'Grady, T., Vandegrift, D., Wolek, M., & Burr, G. (2019). On the determinants of other-regarding behavior: Field tests of the moral foundations questionnaire☆,☆☆,★,★★. *Journal of Research in Personality*, *81*, 224–237.

Presti, S. Lo, Mattavelli, G., Canessa, N., & Gianelli, C. (2021). Psychological precursors of individual

differences in COVID-19 lockdown adherence: Moderated-moderation by personality and moral cognition measures. *Personality and Individual Differences*, *182*, 111090.

Qian, K., & Yahara, T. (2020). Mentality and behavior in COVID-19 emergency status in Japan: Influence of personality, morality and ideology. *PloS One*, *15*(7), e0235883.

Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.

Rest, J. R. (1994). Background: Theory and Research. In J. R. Rest & D. Narvaez (Eds.), *Moral development in the professions: Psychology and applied ethics* (pp. 1–26). Psychology Press.

Reynolds, S. J., & Ceranic, T. L. (2007). The Effects of Moral Judgment and Moral Identity on Moral Behavior: An Empirical Examination of the Moral Individual. *Journal of Applied Psychology*, *92*(6), 1610–1624. https://doi.org/10.1037/0021-9010.92.6.1610

Rohan, M. J. (2000). A rose by any name? The values construct. *Personality and Social Psychology Review*, *4*(3), 255–277. https://doi.org/10.1207/S15327957PSPR0403_4

Ross, L., & Nisbett, R. E. (2011). *The person and the situation: Perspectives of social psychology*. London: Pinter \& Martin Publishers.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication Bias in Meta-Analysis. In H. R. Rothtein, A. J. Sutton, & M. Borenstein (Eds.), *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments* (pp. 1–7). https://doi.org/10.1002/0470870168.ch1

Rozin, P. (2006). Domain denigration and process preference in academic psychology. *Perspectives on Psychological Science*, *1*(4), 365–376.

Ruby, M. B. (2012). Vegetarianism. A blossoming field of study. *Appetite*, *58*(1), 141–150. https://doi.org/10.1016/j.appet.2011.09.019

Sagiv, L., & Roccas, S. (2021). How do values affect behavior? Let me count the ways. *Personality and Social Psychology Review*, *25*(4), 295–316.

Sagiv, L., Roccas, S., Cieciuch, J., & Schwartz, S. H. (2017). Personal values in human life. *Nature Human Behaviour*, *1*(9), 630–639.

Schier, U. K., Ockenfels, A., & Hofmann, W. (2016). Moral values and increasing stakes in a dictator game. *Journal of Economic Psychology*, *56*, 107–115. https://doi.org/10.1016/j.joep.2016.06.004

Schlenker, B. R., Miller, M. L., & Johnson, R. M. (2009). Moral identity, integrity, and personal responsibility. In D. Narvaez & D. K. Lapsley (Eds.), *Personality, identity, and character: Explorations in moral psychology*. Cambridge: Cambridge University Press.

Schwartz, M. S. (2016). Ethical Decision-Making Theory: An Integrated Approach. *Journal of Business Ethics*, *139*(4), 755–776. https://doi.org/10.1007/s10551-015-2886-8

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 25, pp. 1–65). Academic Press. https://doi.org/10.1016/S0065-2601(08)60281-6

Schwartz, S. H., & Butenko, T. (2014). Values and behavior: Validating the refined value theory in Russia. *European Journal of Social Psychology*, *44*(7), 799–813. https://doi.org/10.1002/ejsp.2053

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J. E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, *103*(4), 663–688. https://doi.org/10.1037/a0029393

Schwartz, S. H., Cieciuch, J., Vecchione, M., Torres, C., Dirilen-Gumus, O., & Butenko, T. (2017).

Value tradeoffs propel and inhibit behavior: Validating the 19 refined values in four countries. *European Journal of Social Psychology*, *47*(3), 241–258.

Seuntjens, T. G., Zeelenberg, M., de Ven, N., & Breugelmans, S. M. (2015). Dispositional greed. *Journal of Personality and Social Psychology*, *108*(6), 917.

Smith, I. H., Aquino, K., Koleva, S., & Graham, J. (2014). The Moral Ties That Bind. Even to Out-Groups: The Interactive Effect of Moral Identity and the Binding Moral Foundations. *Psychological Science*, *25*(8), 1554–1562. https://doi.org/10.1177/0956797614534450

Sykes, G. M., & Matza, D. (1957). Techniques of neutralization: A theory of delinquency. *American Sociological Review*, *22*(6), 664–670.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of inter-group conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of inter-group relations* (pp. 33–47). Monterey, CA: Brooks/ Cole.

Tarry, H., & Emler, N. (2007). Attitudes, values and moral reasoning as predictors of delinquency. *British Journal of Developmental Psychology*, *25*(2), 169–183. https://doi.org/10.1348/026151006X113671

Thornberg, R., Landgren, L., Wiman, E., & Thornberg, R. (2018). *It Depends ': A qualitative study on how adolescent students explain bystander intervention and non-intervention in bullying situations*. *39*, 400–415.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory.* Oxford: Basil Blackwell.

Turner, J. C., Reynolds, K. J., Haslam, S. A., & Veenstra, K. E. (2006). Reconceptualizing personality: Producing individuality by defining the personal self. In T. Postmes & J. Jetten (Eds.), *Individuality and the Group: Advances in social identity* (pp. 11–36). London: Sage.

Vainio, A., & Mäkiniemi, J. P. (2016). How Are Moral Foundations Associated with Climate-Friendly Consumption? *Journal of Agricultural and Environmental Ethics*, *29*(2), 265–283. https://doi.org/10.1007/s10806-016-9601-3

van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2020). Does morality predict aggressive driving? A conceptual analysis and exploratory empirical investigation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *74*, 259–271. https://doi.org/10.1016/j.trf.2020.08.017

van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2022). Why are general moral values poor predictors of concrete moral behavior in everyday life? A conceptual analysis and empirical study. *Frontiers in Psychology*, 3834. https://doi.org/10.3389/fpsyg.2022.817860

Warr, M. (2002). *Companions in crime: The social aspects of criminal conduct*. Cambridge: Cambridge University Press.

Yilmaz, O., Harma, M., Bahçekapili, H. G., & Cesur, S. (2016). Validation of the Moral Foundations Questionnaire in Turkey and its relation to cultural schemas of individualism and collectivism. *Personality and Individual Differences*, *99*, 149–154. https://doi.org/10.1016/j.paid.2016.04.090

4. Moral Foundations Theory and The Narrative Self

Towards an Improved Concept of Moral Selfhood for the Empirical Study of Morality

This chapter has been accepted for publication as a journal article in *Phenomenology and the Cognitive Sciences* and is co-authored by Dr. mr. L. D. A. Corrias (second author).

Abstract

Within the empirical study of moral decision making, people's morality is often identified by measuring general moral values through a questionnaire, such as the Moral Foundations Questionnaire provided by Moral Foundations Theory (MFT). However, the success of these moral values in predicting people's behaviour has been disappointing. The general and context-free manner in which such approaches measure moral values and people's moral identity seems crucial in this respect. Yet, little research has been done into the underlying notion of self. This article aims to fill this gap. Taking a phenomenological approach and focusing on MFT, we examine the concept of moral self that MFT assumes and present an improved concept of moral self for the empirical study of morality. First, we show that MFT adopts an essentialist concept of moral self, consisting of stable moral traits. Then, we argue that such a notion is unable to grasp the dynamical and context sensitive aspects of the moral self. We submit that Ricoeur's narrative notion of identity, a self that reinterprets itself in every decision situation through self-narrative, is a viable alternative since it is able to incorporate context sensitivity and change, while maintaining a persisting moral identity. Finally, we argue that this narrative concept of moral self implies measuring people's morality in a more exploratory fashion within a delineated context.

Keywords: Moral values; Moral identity; Moral self; Moral Foundations Theory; Narrative self; Ricoeur

4.1 Introduction

Over the last few decades, the empirical study of moral decision-making has established itself as an important sub-field of psychology, known as moral psychology. Within this field, measuring people's basic and general moral values, through a general moral questionnaire, has become a common academic practice to map out people's morality (Curry et al, 2019; Graham et al, 2013). Furthermore, these measures have been used to investigate the influence of people's morality on other attitudes and behaviour (e.g., Clark et al., 2017; Cohen, 2014; Dickinson et al., 2016; Graham et al., 2009; Hoover et al., 2021; Miles, 2015; Nilsson et al., 2016; O'Grady et al., 2019; Qian & Yahara, 2020; Vainio & Mäkiniemi, 2016).

Theories and accompanying questionnaires that are used to measure people's moral values are, for example, Schwartz Theory of Basic Values (Schwartz, 1992; Schwartz et al., 2012)[31] and Moral Foundations Theory (MFT) (Graham et al., 2013; Haidt & Joseph, 2004). A more recently developed one is Morality As Cooperation theory (MAC) (Curry et al., 2019). These, what we will call, 'empirical moral value theories' define people's morality through a set of distinct basic moral values. These moral values are regarded as psychological mechanisms that were formed during the evolution of the human being and which are further individually developed during someone's life. The extent to which an individual has developed a general moral value as part of his or her morality is empirically measured through a generic questionnaire.

Though these values are often presented as determinants of behaviour, their success in predicting people's actual moral behaviour has been disappointing (Boyd et al., 2015; Graham et al., 2012). Studies that have specifically investigated the link between general moral values and specific moral behaviours report weak associations at best (e.g., O'Grady, 2019; Schier et al., 2016; Van den Berg et al., 2020; Van den Berg et al., 2022)[32]. It is far from clear whether people who score higher on a general moral value when filling in a questionnaire also show more behaviour in accordance with that value (Graham et al., 2012). This is problematic as predicting and explaining individual differences in actual moral behaviour seems to be a central goal when measuring people's morality (Ellemers et al., 2019), and as such a central goal of moral psychology.

When considering this issue, one element that catches the eye is the general character of the measured moral values that are attributed to the individual, and the contextless manner in which these are measured. In light of psychological studies that have emphasized the situational (e.g., Doris & Doris, 2002; Ross & Nisbett, 2011), the social contextual (e.g., Potter & Wetherell, 1987; Southerton et al., 2004) and issue-contingent nature (Jones, 1991) of (moral) values and behaviour, it is questionable whether regarding such general individual measures as direct determinants of behaviour is in accordance with important presuppositions of how people morally function. This brings us to question some of the more fundamental assumptions underlying general moral value theories. In particular, it raises questions about a

[31] Note that Schwartz Theory of Basic values entails a broader set of general values than just morally relevant ones.

[32] A similar problem has been identified with regard to the association between people's moral judgments and their behaviour, known as 'the judgment-action gap' (see e.g., Lapsley & Narvaez (2004). In this paper we will not go further into this discussion as we confine ourselves explicitly to empirical moral value theories and the relation between moral values and behaviour. However, we do not rule out that findings and insights in our analysis may prove relevant to that discussion.

concept that seems to be central to our moral functioning and our relationship with moral values, namely that of the moral self. With this, we mean a concept of moral identity that corresponds to the agent of moral decision-making and action. Empirical moral value theories assume a certain conception of moral self, underlying their theory and measurement strategies, as they, in fact, identify people in moral terms. However, what their concept of the moral self entails is not clear. The question then is what this implicit moral self, underlying empirical moral value theories, looks like; how it should be evaluated and, possibly, improved; and what this means for the method of measuring morality's influence on behaviour.

In this paper, we investigate and critically assess the concept of moral self that underlies empirical moral value theories and argue for an improved concept of moral selfhood for the empirical study of morality. For this purpose, we confine ourselves specifically to Moral Foundations Theory (MFT) and its Moral Foundations Questionnaire (MFQ). We will take the following approach: first, we explicate the implicit notion of the moral self that MFT assumes through a description and analysis of the theory and questionnaire. This results on the one hand in a concept of moral self that consists of stable moral traits, while on the other hand the theory seems to adopt aspects of virtue theory that indicate a more flexible and dynamic moral self, without further developing such a concept. Then, to make sense of this somewhat ambivalent result, we turn to a field that has extensively studied the self: phenomenology. More particularly, we describe and discuss Ricoeur's concept of narrative moral self (Ricoeur, 1992). This elaboration is used to interpret and evaluate MFT's concept of moral self and, at the same time, present the narrative moral self as a more viable alternative. In the discussion, we make a few suggestions as to how insights from Ricoeur's narrative self could possibly enhance the empirical measurement of the influence of moral considerations on behaviour.

4.2 Moral Foundations Theory and the implied moral self

In our analysis, we focus explicitly on MFT because it is arguably one of the more prominent empirical moral value theories within moral psychology today. Furthermore, it is referred to and used in other fields than moral psychology to study the moral influence on decision-making, such as consumer and environmental studies (e.g., Chowdhury 2019; de Jonge & van Trijp, 2014; Vainio & Mäkiniemi, 2016) and behaviour during the COVID-19 pandemic (Presti et al., 2021; Diaz & Cova, 2022). Another reason is that the theory has been extensively described over the last two decades in several empirical and more theoretical papers, which can be used to derive the concept of moral self that it assumes. In the following, we will explicate this implicit conception of the moral self through a description and analysis of the theory and its questionnaire.

4.2.1 Moral Foundations Theory

Moral Foundations Theory (MFT) positions itself opposite to rational and monist models of morality, such as the Kohlbergian model of moral development (Kohlberg, 1969, 1984). Instead, MFT claims that our morality consists of a plurality of fundamental moral values[33], the so-called moral foundations, which intuitively influence our

[33] The developers state that their theory is primarily about moral virtues instead of moral values, given the cultural developmental element that is part of it. Furthermore, it is stated that the moral foundations

decision-making. Next to moral pluralism and intuitionism, it takes two other elements as central to its theory: nativism and cultural learning. Furthermore, MFT presents itself as a descriptive moral theory, it aims to describe which moral values people actually have, instead of making normative claims about which they should have. People's moral foundations are measured through the accompanying Moral Foundations Questionnaire (MFQ) (Graham et al., 2013). We will now briefly go into the four central elements of the theory and, subsequently, into the structure of MFQ.

The first element MFT regards as central to the theory is *pluralism*. It consists in the fact that the theory distinguishes more than one fundamental moral value, i.e., the moral foundations. At this point, at least five moral foundations are distinguished: care/ harm, fairness/ cheating, loyalty/ betrayal, authority/ subversion, and sanctity/ degradation. [34]

Secondly, the moral foundations are described as *innate* psychological learning modules or mechanisms that developed as adaptive solutions to distinct social problems of (group) survival, during human's evolution. As innate learning modules, they are part of every individual's moral mind in advance of experience. This means they bear a universal *a priori* sensitivity to certain moral beliefs, values, virtues, concepts, principles, judgments etc. congruent to the specific moral domains that the different foundations designate (Graham et al., 2013).

Third, it is claimed that this *a priori* universal moral mind only constitutes the 'first draft' of every individual's morality. During an individual's life, the first draft is 'edited' through *cultural learning*. This means that the five domain-specific learning modules develop and create more specific modules within their moral domain as people grow up in a specific social and cultural environment. These specific modules constitute people's more specific moral values, virtues and intuitions. For example, the innate fairness learning module develops several specific fairness modules during a person's life that are culture-specific (e.g., a module for 'not cutting in line' in a culture in which waiting for your turn is an accepted means of allocating goods). To what extent an individual develops the different moral foundations, and the according sensitivities to domain-specific moral concepts and beliefs, into their actual adult morality, depends on the social and cultural environment one is brought up in and the individual's personal experiences, especially during childhood. What specific moralities different people develop, in the sense of specific values and virtues, is therefore in itself not innately given. According to MFT, this process of cultural learning that is universally structured by the possibilities and limitations of an initial organization of the moral

themselves are not one's moral virtues, they are the fundamental learning modules that give rise to people's broader set of culturally embedded virtues (see below) (Graham et al., 2009). However, in the broader literature on MFT, the moral foundations themselves are often referred to as people's (basic or general) moral values, including in studies by its developers. Furthermore, the questionnaire itself does not contain anything that designs it specifically for tapping into moral virtues instead of moral values. The items in the questionnaire seem to be suitable for both. We therefore refer to what is measured by MFQ as one's moral foundations, general moral values, or general virtues interchangeably.

[34] The first two are considered the 'individualizing' foundations, as these protect the interests of the individual. The last three are considered the 'binding' foundations as they protect the interests of the group (Graham et al., 2009). Note that the authors state that they are confident about the existence of these five moral foundations but that more may be discovered. See Graham et al. (2013) for a checklist noting the criteria candidate foundations should satisfy before they can be recognized as a genuine one. Promising candidates that have been proposed are the values of 'liberty/ oppression', equity/ undeservingness, and 'honesty/ lying' (Graham et al., 2018).

mind, can explain the stark differences as well as the (more fundamental) similarities that are found between moralities across persons, groups and cultures (Graham et al., 2018; Graham et al., 2013; Haidt & Joseph, 2004, 2007).

Besides pluralist, nativist, and cultural learning elements, the fourth important component of the theory is its *moral intuitionism*. MFT builds on the Social-Intuitionist model of moral judgment developed by Haidt (2001), which claims that our moral judgments are directly *caused* by an intuitive process of moral evaluation. That is, an effortless and affective process that automatically results in a moral evaluation of like or dislike: 'moral intuitions [are].. bits of mental structure that connect the perception of specific patterns in the social world to evaluations and emotions that are not fully controllable or revisable by the person who experiences them.' (Haidt & Joseph, 2007, p. 381). People's moralities are regarded as consisting of intuitive 'input-output programming' (p. 379), i.e., the more specific moral modules (e.g., "not-cutting-in-line"), which largely encapsulate moral judgment.

Deliberate reasoning is considered as *post-hoc* rationalization that does not have a direct influence on the already intuitively established moral belief or judgment. It only serves a social function in explaining the intuitively derived moral judgment to others, justifying what is already established and will not be altered. Haidt (2001) states that it is the kind of reasoning that is usually associated with a lawyer instead of a scientist, namely, fitting the reasoning towards an already accepted conclusion instead of impartial reasoning leading towards a yet unknown answer.

4.2.2 Moral Foundations Questionnaire

To what extent an individual has developed the different moral foundations as part of his or her morality can be measured with the Moral Foundations Questionnaire (MFQ) (Graham et al., 2011; MoralFoundations.org, 2013). The questionnaire consists of two parts. In the first, the so-called 'relevance part', respondents are asked to rate to what extent different general considerations are relevant to their thinking when judging between right and wrong (not at all relevant- extremely relevant). Each item taps into one of the moral foundations. For example, the item 'Whether or not someone suffered emotionally' is related to the care/harm foundation; the item 'Whether or not someone acted unfairly' to the fairness foundation; and the item 'Whether or not someone conformed to the traditions of society' connects to the authority/subversion foundation.

The second part of the survey, the so-called judgment part, asks about the respondents' agreeableness with moral statements (strongly disagree-strongly agree). An example of a judgment-item related to the moral foundation of loyalty is 'It is more important to be a team player than to express oneself' and one related to purity/degradation is 'Chastity is an important and valuable virtue'. Though the developers claim that the judgment part was designed to ask about context-specific moral judgments to complement the general moral considerations of part one, the judgment items rather seem to consist of general and abstract moral principles, often applicable to a wide variety of contexts (Clifford et al., 2015; Gray & Keeney, 2015).

The scores on the six items belonging to the same foundation are summed up, which results in an individual's sum score per foundation. This expresses the degree to which a moral foundation underlies one's moral beliefs and concerns (Graham et al., 2011). The sum score for each foundation can consequently be used for testing association with other variables, such as attitudes and behaviours.

4.2.3 MFT's implicit concept of moral self

With the above description of the theory and the design of the questionnaire, it is possible to discern what concept of the moral self is implied by MFT. People's moral identity is defined here by their endorsement of the moral foundations, expressed by their individual scores on MFQ. The moral self that seems to be implicitly supposed can then be characterized as a fairly stable moral personality or moral character. People's intuitive moral regularities are developed by, and can be brought back to -or better: aggregated to- at least five general psychological moral dispositions that latently exist within the individual. As the development of these general psychological moral dispositions is measured outside of any specific context, it is assumed that these exist independently of any specific context and have a relatively stable hierarchy and efficacy across contexts. People's general moral values are in this sense very much presented as moral personality traits. The moral self that underlies MFT seems to be an example of what Frimer and Walker (2008, p. 344) call an "essential" self-concept: a moral self that is '*unified, internally consistent and has an essence that exhibits agency across contexts*'. The essential nature of one's moral self here consists of the endorsement or development of the general moral values, which is expressed by an individual's MFQ-score. This does not mean that one's moral values and moral self cannot further develop or change over time. Yet, it does suggest that this moral self amounts to a relatively stable moral character that causally affects decision making and behaviour in a similar way across different contexts and over a longer time frame. In its core, the essential aspect of this concept of moral self lies in the proposed intuitive moral regularities that produce a certain output when receiving a certain input and which can allegedly be summarized by abstract moral value scores without referring to any context.

4.2.4 Leeway for a more dynamic concept of moral self within MFT

The described implicit conception of the moral self as a stable moral character is then derived from two interlinked sources. On the one hand, it follows from the main theoretical description of MFT, where domain-specific learning modules develop intuitive moral regularities that causally determine people's moral judgments. On the other, it is derived from the nature of the questionnaire that asks about general and contextless moral principles. This presumes that these moralities can be effectively summarized into *general* moral values or virtues, functioning as individual moral traits and affecting the same kind of behaviour across contexts. This then sketches a quite rigid and determined picture of the moral self and the influence of people's moral values on behaviour.

The view on the human being as having general (moral) traits determining behaviour across contexts, has been criticized by situationist theories that emphasize the influence of the situation and social context on (moral) decision-making and behaviour (e.g., Doris & Doris, 2002; Ross & Nisbett, 2011). Interestingly, Haidt and Joseph (2007) defend MFT against this possible critique and, thereby, create leeway for a moral self (without further developing it) that seems to be in tension with the concept of moral self that is suggested by the main theory and questionnaire. Here, MFT is placed in the tradition of virtue ethics and, while first describing people's developed moralities as intuitions in the sense of input-output regularities, these moral modules are now also linked to virtues and to characteristics of virtues that suggest a more dynamic and contextually sensitive interpretation of people's moral nature.

Haidt and Joseph (2007) state that virtues are characteristics of a person or traits, but not in the sense of broad behavioural dispositions or 'global tendencies to act in a particular way (e.g., honest, brave) across widely varying circumstances' (p. 386). Rather, virtues are described here as 'dynamic patternings', 'capacities', or situation-specific 'social skills' (p. 386): '[t]o possess a virtue is to have extended and refined one's abilities to perceive morally-relevant information so that one is fully responsive to the local sociomoral context. To be kind, for example, is to have a perceptual sensitivity to certain features of situations, including those having to do with the well-being of others, and for one's motivations to be appropriately shaped and affected.' (Haidt & Joseph, 2007, p. 386). Developing a virtue is a 'comprehensive attunement to the world' (p. 387) and, furthermore, it is explicated: 'what it means for a personality characteristic to be a *virtue* and not simply a behavioural regularity, is largely that it consists in functioning well in a specific "sphere of existence."' (p. 387). Here, people's moral character, consisting of certain virtues or moral values, receives a more dynamic and contextual nature in the sense that it attunes to and is embedded in the social context. This seems to presuppose a different relationship between the moral agent and his or her values, and, ultimately, a different concept of the moral self than the one that we derived from the main description of the theory and questionnaire.

Another aspect that the authors bring forward as relevant to MFT and that has been linked to virtue ethics is the role of narrativity in moral thinking and moral development. It is argued that through our moral intuitions, produced by our moral foundations, that narratives can become compelling moral stories. At the same time, it is through moral narrative that the intuitions of our moral foundations are socialized and developed into coherent moralities while growing up (Haidt & Joseph, 2007). The notion of narrativity is however neither further developed nor connected to a concept of the moral self. The connection with narrativity does seem to point towards a self that interprets and tries to make sense of oneself and its social environment, presuming a certain dynamic and context sensitivity as part of moral thinking.

The link that is made by MFT to virtue ethics and narrativity then opens the door to a more dynamic moral self that is sensitive to the specific situational context in which it finds itself. However, such a concept is not further developed, and, also, seems to be at odds with the implicit essential moral self that underlies MFT's main theoretical elements and structure of the questionnaire. In fact, a more dynamic and contextual moral self with a corresponding relationship to one's moral values, seems problematic for MFT's idea that morality mainly consists of intuitive input-output regularities as well as for predicting judgments and behaviours on the basis of generally measured moral values. One of the main aims of the analysis in the following sections is to explicate and problematize this ambivalence in MFT and to propose a more viable concept of moral self for the empirical study of morality.

4.2.5 Turning to phenomenology

To investigate and explicate the found ambivalence in MFT and evaluate its concept of moral self, we will now continue our analysis on the fundamental level of the moral self. As such, it seems prudent to turn to a field that has a long tradition in thinking about the (moral) self, namely *phenomenology*. Phenomenology is the philosophical field which systematically studies the first-person perspective of the experiencing and meaning giving subject. More particularly, in the next section, we turn to Ricoeur's

narrative concept of the self. This theory presents a thoroughly developed (moral) self-concept, where three central notions that have arisen from our above investigation form fundamental elements, namely: *moral character*, *a self that attunes to and is embedded in the social context*, and *narrativity*.

By explicating Ricoeur's phenomenological concept of self we are, first of all, able to articulate a substantiated moral self that has fully developed the three above notions. Secondly, we are able to relate Ricoeur's concept of moral self to the one that MFT implicitly assumes through its main theory and questionnaire. We can then evaluate the latter in terms of the former. This will also make clear what it would mean for MFT and its measurement tool if it in fact adopts a full-fledged dynamic concept of the moral self, like Ricoeur's narrative moral self, and leaves its ambivalent position. Third, this brings us to suggestions as to how it may be possible to incorporate the found phenomenological insights into the empirical study of people's morality.

4.3 Ricoeur's concept of the moral self

4.3.1 Ricoeur's phenomenological and hermeneutical approach

In this section and the next, we draw on the work of Paul Ricoeur, to present a thoroughly developed concept of the moral self which explicates ideas that MFT points to, but does not elaborate on. We will argue that this conception is ultimately at odds with aspects of MFT's main theory and questionnaire.

Ricoeur's overall aim is to develop a notion of moral selfhood that sails between the Scylla of the Cartesian essential Ego and the Charybdis of the Nietzschean splintered subject, offering an alternative to both extremes (Ricoeur, 1992, p. 1-16). His phenomenological approach consists of a hermeneutics of the self[35] that seeks a position between these two alternatives. *Hermeneutics* can be understood as the philosophical approach that argues that the specific methodology of the humanities consists of interpreting (*Verstehen*) its objects of study, usually texts. Ricoeur applies this method of interpretation to the notion of self, as it is experienced by the first-person subject, to grasp and explicate this fundamental experience in a systematic way. Accordingly, we will call this hermeneutics of the self a phenomenological approach, since it stays true and further develops the first-person account.[36] This hermeneutics proceeds by detours in order to tackle the question of identity or self by devising a theory of human action.

[35] We are aware of the fact that Ricoeur's notion of a narrative self has not been uncontested. In particular, we are familiar with the critique of Zahavi (2007, 2008 and 2014), who argues that a phenomenological, embodied notion of selfhood precedes and underlies the hermeneutical, narrative self. A full treatment of this issue goes beyond the limits of this article. Nevertheless, we do want to stress that we are of the opinion that within Ricoeur (1992), and in particular when read against the background of other parts of his oeuvre, phenomenological and hermeneutical considerations are combined, leading to a notion of selfhood that is both embodied and discursive. Accordingly, we make no strict distinction between phenomenology and hermeneutics in this article. For an in-depth defence of this interpretation of Ricoeur's notion of selfhood, see Halsema (2019).

[36] Ricoeur clarifies his position in Ricoeur (1975). He rejects one specific version of phenomenology; the idealistic one of the early Husserl. Apart from that, he sees a relationship of 'mutual belonging' between phenomenology and hermeneutics. On the one hand, hermeneutics builds on phenomenology as its necessary presupposition. On the other hand, the goal of phenomenology – the return to the *Lebenswelt* – cannot be attained without the hermeneutical presupposition of interpretation.

Here it seeks to connect the questions 'what' and 'why' of action (what is action and how can we explain it?), which are the focus of analytic philosophy, to the question 'who' (who is acting?), which is easily concealed but constantly presupposed by the first two. According to Ricoeur, it takes a hermeneutical approach that builds upon phenomenology, to bring forward this aspect of the acting person, which is selfhood. The only certainty this approach may claim is that of attestation. Attestation reaches an epistemic level that stands in opposition to the 'ultimate and self-founding knowledge' of the Cartesian Ego. However, it is not mere belief in the sense of *doxa*, which is inferior to knowledge. Rather, attestation links to the epistemic value notion of credence or trust and can ultimately be understood as 'the assurance of being oneself acting and suffering' (Ricoeur, 1992, p. 22).

4.3.2 *Idem*-identity and *ipse*-identity

Ricoeur's concept of moral selfhood is a narrative notion of personal identity that should be understood as existing in time and as the ongoing dynamic interplay between two poles of identity: selfhood (*ipse*) and sameness (*idem*). In the following we will first describe these two distinct poles of personal identity and their specific way of existing in time. Subsequently, we will go into narrativity as mediating between them and constituting the moral self.

Idem-identity is identity in the sense of "sameness". Overall, sameness is an answer to the question of identity in terms of "what?", making re-identification possible. It has three different components: numerical identity, qualitative identity, and uninterrupted continuity. These three components may be contrasted with plurality, difference, and change, respectively. First, numerical identity means oneness in the sense that two occurrences of a thing are one and the same (I saw the plant in the room and now that I re-enter the room, I see the very same plant). Second, qualitative identity denotes the situation of extreme resemblance to the point of interchangeability. That is, between two things there is no qualitative difference (you are wearing the exact same dress as I!). Third, uninterrupted continuity harbours sameness as permanence in time, in the sense that one and the same individual goes through different stages of development (think of scrolling through someone's photo-album or Facebook timeline covering several decades, and identifying the changing appearance as the same individual) (Ricoeur, 1992, p. 116-118).

Selfhood, or *ipse*-identity, is not sameness. It is another form of permanence in time, another way of answering the question of identity that is particularly relevant to the question of personal identity. While also things have *idem*-identity, *ipse*-identity belongs to persons only. Selfhood denotes reflexivity in the sense of a relation to self (Ricoeur, 1992, p. 1-2). It is a form of permanence in time that is an answer to the question "who?", specifically when we are looking for the agent of an action: "who did that?" (Ricoeur, 1991). Persons are the privileged bearers of this notion of agency, in the sense that the actions belong to the self, the self owns them. So, the reflexivity resides in an agent being able to recognize herself as the subject of a certain action[37]. This self-ascription presupposes an identity –selfhood, self-designated by "I, myself"- that cannot be expressed in terms of sameness but that resides in one's experience as a subject. The contrasting notion of selfhood is then not difference, but otherness.

[37] In the same sense this reflexivity of the self is contained in recognizing one's body as one's own body, i.e., in Husserl's idiom *Leib*; or in one's thoughts as one's own thoughts, i.e., as belonging to oneself.

When it comes to the permanence in time of persons, sameness (or *idem*) manifests itself as character. In other words, character points to one understanding of permanence in time, one way of answering the identity question: "who am I?", which takes the form of the question "what am I?" as it is answered in terms of sameness or "what". Character is seen as 'the set of distinctive marks which permit the re-identification of a human individual as being the same' (Ricoeur, 1992, p. 119). These lasting dispositions are related to habit, understood as a notion of sedimentation, and to acquired identifications with what is other than self, such as norms, values, or role models. Interpreted in this way, character offers stability, in terms of permitting the re-identification of persons, through the three senses of sameness: numerical identity, qualitative identity, and uninterrupted continuity. Though *idem* is emphasized at this pole of personhood, Ricoeur stresses that character is actually where *ipse* and *idem* overlap. Or better, 'nearly overlap', as their difference is not annulled. It is namely *my* character that belongs to *me*: 'precisely as second nature, my character is me, myself, *ipse*' (Ricoeur, 1992, p. 121). Or, as Ricoeur notes, character can be understood as the expression of selfhood in terms of sameness, i.e. 'the "what" of the "who"' (1992, p. 122).

Selfhood, on its own, harbours another form of permanence in time belonging to persons, namely self-constancy, in the sense of 'that manner of conducting himself or herself so that others can *count on* that person. Because someone is counting on me, I am *accountable for* my actions before another. The term "responsibility" unites both meanings: "counting on" and "being accountable for"' (Ricoeur, 1992, p. 165). The notion of keeping one's word can be taken as emblematic for this self-constancy of selfhood. Indeed, the keeping of one's promise appears 'as a challenge to time, a denial of change: even if my desire were to change, even if I were to change my opinion or my inclination, "I will hold firm"' (Ricoeur, 1992, p. 124). In this sense, for Ricoeur, the self has an inherently moral aspect from the outset. Keeping one's word forms the opposite pole in Ricoeur's model of permanence in time with regard to character, namely where sameness and selfhood are separated by an extreme gap (Ricoeur, 1992, p. 124). To make this pole of personal identity and its permanence in time more tangible, one can think of the practice of people who give each other their wedding vows in which they promise to take care of and be there for each other *no matter what* (whether it be a change of circumstances or of character). It is in this practice that we rely on the self and its self-constancy, independent of character.

To relate this back to the idea of *my* character, the pole of identity where *idem* and *ipse* nearly overlap, this notion of self-constancy or maintaining oneself is what characterizes my relation with my (moral) dispositions, such as the values I consider as my own. Ricoeur makes clear that this is a reflexive relationship, one of loyalty or fidelity towards these values and one of recognizing oneself in these values. This makes these values my own, while at the same time it permits a certain movement, flexibility, and adjustment in my relationship with values. It is exactly narrative that mediates this dialectic between myself and my values which is described in the next section.

4.3.3 The moral self as a narrative notion of personal identity

As stated, Ricoeur's concept of the moral self consists of the dynamic interplay of the two described poles of personal identity –*idem* and *ipse,* or character and self-constancy. This interplay is mediated by narrativity (Ricoeur, 1992, p. 140-151). Instead of what Haidt and Joseph (2007) emphasize, the role of narrativity is neither solely nor primarily pedagogical. Narrativity is, rather, constitutive for a viable account of the moral self.

Indeed, because the (moral) self comes into existence by being narrated, the self should be understood as constructed by narrative (Halsema, 2019).

Before going into the technical details, it might be good to first get a basic idea of why Ricoeur gives narrativity a central place in his theory of identity. Think of the very first question you often get at a job interview: "Please tell us something about yourself." In answering this question, you usually do not start enumerating your (best) character traits. Instead, you tell a story: the story of who you are. Naturally, through this story your character traits transpire. However, they are only one element in your story that encompasses your actions and behaviours and relations to others. If, at a later stage in the interview, you are asked which character traits make you perfect for the job, you might enumerate them. Yet, notice how this alone is seldom sufficient since you are usually asked to illustrate these with a concrete example. Here, again, your answer takes the form of a narrative: you tell the story of how flexibly you reacted when confronted with a sudden change in your schedule. The bottom line is that a full account of one's identity takes the form of a narrative that mediates selfhood and sameness.

In his narrative account of the self, Ricoeur connects narrativity to the plot. The plot has an integrative function (Ricoeur, 1984, 1992). Understood as the movement of 'discordant concordance', the plot generates a 'synthesis of the heterogeneous' (Ricoeur, 1992, p. 141). More particularly, through the plot individual events and the story as a whole are configured, and brought into one coherent whole: 'the narrative event is defined by its relation to the very operation of configuration; it participates in the unstable structure of discordant concordance characteristic of the plot itself. It is a source of discordance inasmuch as it springs up, and a source of concordance inasmuch as it allows the story to advance' (Ricoeur, 1992, p. 142). The nature of the plot is therefore one of permanence and change.

As the plot makes a coherent whole of the heterogeneous elements in a story, it also provides the characters within the story with their identity. This is a narrative identity, correlating to the events of the story. A narrative, namely, does not describe events in an impersonal way. In a narrative, characters are linked to events as the ones who perform the actions or who are affected by the events, i.e., the narrative describes the character in its acting and suffering, defining the character. In other words, by telling the story through the structure of the plot, the questions of 'who?', 'what?', and 'why?' are answered by connecting these answers through time. This gives the story a temporal configuration that makes it possible to follow it and, at the same time, renders the character a comprehensible identity that has duration in time (Ricoeur, 1984, 1992). As the character of a story achieves her narrative identity via the movement of the plot - mediating between change and permanence – this narrative identity itself also has the structure of the plot. As Ricoeur (1992, p. 143) notes: 'characters, we will say, are themselves plots'. This means that the discordant concordance characteristic of the plot also applies to the character itself. This structure of change and permanence is the very structure of the dialectic between sameness and selfhood.

Now in the same vein, the identity of a person is constructed by telling one's life story. In a hermeneutics of the self, it is the self that interprets her life through its acting and suffering and, simultaneously, her character through the mediation of narrativity with the structure of the plot. This interpretation takes the form of an appropriation, in the specific sense of making one's own. Ricoeur (1992, p. 160-163) argues in this regard that narratives and life itself remain two distinct things. Think of how literary narratives

cannot be simply applied to life but need to be appropriated: in this interaction between reader and text, the reader becomes the co-author of the meaning of a story. In the same vein, the self-narrative is an interpretation of one's experiences of acting and suffering through an appropriation that organizes one's life, integrating one's past, present and future.

The mediation of narrativity, more precisely, relates the two poles of permanence in time of persons that Ricoeur distinguishes: self-constancy – the moral dimension of selfhood – and character. Through the narrative interpretation of the self, the self connects the question 'who?' to that of 'what?'. It provides the self with some 'flesh on the bones' through the story that is told and the dispositions of character that figure within it. At the same time, it gives the character the possibility to innovate when sedimentation has rendered traits rigid. It returns character to the movement that was lost in the acquired set of properties.

The constitution of one's identity or the self through narrative, mediating *ipse* and *idem*, then provides character with a dynamic nature. This is, first of all, seen in the fact that, just as other stories, self-narratives can be told more than once and in different ways. Also, it is conceivable that different stories harbouring different kind of character traits apply to different social contexts (e.g., at work you are an authoritarian boss, but at home a timid husband). Furthermore, Ricoeur (1992) emphasizes that a person is only the co-author of her own life story as also other people tell and add to a person's life story; entailing that one's narrative identity remains open to changes and revisions until people stop talking about the person (Halsema, 2019). Finally, the dynamic nature of identity is shown in the interpretive act by which one configures the discordant events as part of a concordant life story. The appropriation of new events entails a constant reinterpretation of one's life story and thereby of oneself. When having new experiences and figuring out what to do, these are interpreted in light of who you are, while attuning to the context of the specific situation. The appropriation of the experience as part of one's life story then takes a mutual fitting by relating to and questioning oneself, as well as the specific social context. This implies a dynamic and context sensitive concept of the self, where acquired dispositions are brought back to the process of acquiring them, while self-constancy is given recognizable features. In this way the self is constituted as a permanence in time that is ever changing and developing.

4.3.4 Narrative identity and moral identity

This narrative self remains faithful to the general claim of hermeneutics: it offers an interpretation of the self by a reconfiguration of cultural signs into symbolic circuits. This means that action and, accordingly, the self as interpreted in its acting and suffering, is always embedded within a certain culture and symbolically mediated. This entails that it takes place in a practical field that is articulated by rules, norms, values and signs (Ricoeur, 1984, p. 57-59). These give the practical field a meaning that is inherently public, i.e. available to and readable for actors within the field, while it also makes the interpretation of oneself and what should be done context-specific. Furthermore, this symbolic circuit or texture of action opens up to the idea of the prescriptive or normative, making it possible to evaluate action. Hence, neither action nor the self can be morally neutral.

In fact, narrative has had moral implications from its very start: right from the oral tradition of storytelling, which was about exchanging experiences and examples of exercising practical wisdom. In a similar way, fiction provides us with imaginative

explorations of judging characters and actions. As Ricoeur (1992) makes clear, a narrative is never morally neutral, but peppered with evaluations. The plot, for instance, does not only logically structure a narrative, but also provides it with a certain end goal or good (implicitly) put on the horizon. Furthermore, narrative theory anticipates and supports moral theory by the concept of action that it offers (Ricoeur, 1992). This is a specific, layered notion of action, of which practices (basic actions, nesting relations, constitutive rules) and life plans (the narrative unity of a life as brought about by actual experience and fabulation) are important elements, forming an integrative whole expressing certain values and goals that are related to 'the good life'. Narration, therefore, marks the transition between ascribing action to an agent and prescribing obligations to act to an agent. It is the narrative self that is the agent of moral action, for such notions as 'the narrative unity of life', 'life plans', 'the good life' assume both how life is rooted in biology and the way in which an agent regards this life as her own (Ricoeur, 1992, p. 178). This agent is a self-interpreting animal in the sense of Charles Taylor: interpreting and trying to bring into agreement her notion of the good life with her actions (Taylor, 1985).

At the same time, narrative identity, Ricoeur argues, is also characterized by the dialectic of the self and the *other* (Ricoeur, 1992). In this respect, narrative identity does not solely consider my own life and life plans, but it is also concerned with duties towards others and how to treat them well. The interconnectedness of narrative and moral identity entails, on the one hand, that narrating implies morally evaluating one's actions, on the other hand, moral identity assumes that one is able to give a narrative account of one's actions, reflecting upon them and giving reasons for them (Halsema, 2019).

Our interpretation of the self in its acting and suffering is thus always normative. This is already given by the necessarily interpretive approach that we take to our life and our experiences (Van Tongeren, 2020). These experiences are rooted in a way of being in the world that is always already normative, i.e. characterized by meaning and values. In this regard Sayer (2011) claims that we are 'beings for whom things matter', i.e. our relationship to the world is primarily evaluative, marked by import, significance, or, indeed, meaning. It is in these interpretations, or narratives, that our moral dispositions like virtues and values, defining our moral character, have their place and through which they are expressed. As explained, this entails a dynamic and context-sensitive relationship with moral values, which involves a constant relating and questioning of our values within a specific context to which one attunes. This leads to an idea of moral decision-making that amounts to moral evaluation in situation.[38] This

[38] Ricoeur discusses the moral evaluation in situation under the heading of the Aristotelean concept of *phronesis*, see especially study 7 and 9 of Oneself as Another (1992). While an extensive discussion of *phronesis* in the work of Ricoeur falls outside of the scope of this paper, it is noteworthy that he explicates the relationship between Aristotelian ethics and Kantian morality in Ricoeur (2007). The Kantian framework is crucial in delimitating 'the hard core of the issue', that is designating morality as the realm of the obligatory (Ricoeur 2007, p. 45). Aristotelian ethics situates moral decision-making in concrete situations against the horizon of a person's ideas about 'the good life'. It is here that virtues, understood as 'excellences of action', take central stage: 'Each one of these excellences picks out its intended good against the background of an open-ended intention magnificently designated by the expression of a "good life" or, better, "living well." This open horizon is inhabited by our life projects, our anticipations of happiness, our utopias, in short by all the changing figures of what we take as signs of a fulfilled life.' (Ricoeur 2007, p. 50). The notion of *phronesis* enters the picture because the exercise of virtue in concrete

also means that, through the mediating role of the narrative moral self, the relation between our moral values and our behaviour is characterized by interpretation, making the effects of moral values on behaviour dynamic, in the sense of varying in kind and strength across contexts.

4.4 Confronting MFT's concept of moral self with Ricoeur's narrative moral self

Let us retrace our steps. In the second section we explicated the moral self that MFT implicitly assumes through its main theory and questionnaire, while we also referred to three notions related to virtue ethics which the developers of the theory have linked to MFT: *moral character*, *a self that attunes to and is embedded in the social context*, and *narrativity*. Together, these three notions seem to point to an alternative, possibly more dynamic and context-sensitive moral self. However, such a moral self is not developed within MFT-scholarship. In section 3, we turned to phenomenology to investigate what such a dynamic moral self would look like. Here we expounded on the narrative conception of the moral self, as developed by Ricoeur, to bring forward a thoroughly developed conception of the moral self that elaborates these three notions. We can now relate back to MFT to see what these insights can tell us about the moral self that it implicitly presumes through its main theory and questionnaire, i.e. how we can qualify MFT's moral self in relation to Ricoeur's narrative moral self, and, in its wake, what this suggests about empirically measuring morality.

4.4.1 MFT defines the moral self solely in terms of *idem*-identity

We ascertained in section 2 that MFT's implicit conception of the moral self consists of an internalization of general moral values or development of general moral virtues that function as character traits and together form a fairly stable moral personality. This was derived from the notion that MFQ measures the extent to which distinct general learning modules (the moral foundations) are reflected in people's developed moralities (consisting of intuitive input-output regularities). The questionnaire consists of general items such as: 'When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking? –'Whether or not someone suffered emotionally'/ 'Whether or not someone acted unfairly' etc., each tapping into one of the five defined moral foundations. The aggregate measure of each foundation can be regarded as the endorsement of a *general* moral value or virtue. These are subsequently used to explain different kinds of attitudes and behaviours across contexts. As the general moral values are measured outside of any specific context, it is assumed that their endorsement exists independently of any specific context and that they have a relatively stable hierarchy and efficacy across contexts. Together with the idea that these general moral values are the aggregate reflection of input-output moral regularities, this amounts to an essentialist concept of moral self, where people's general moral values or virtues can be regarded as dispositional moral traits that, through moral regularities, intuitively and causally determine certain attitudes and behaviour.

situations cannot be separated 'from the personal quality of the wise human being – the *phronimos* – the sensible person.' (Ricoeur 2007, p. 54).

Following Ricoeur's terminology, the above description of MFT amounts to a conception of the self that -at most- can be considered in terms of *idem*-identity. People's score on the MFQ is an answer to the question 'who am I (morally speaking)?' in terms of *what*: the moral values that someone supposedly endorses in general. It is this general moral value endorsement that defines the person morally and that bears the characteristic of sameness. It constitutes a moral character that offers stability through the three senses of sameness: numerical identity, qualitative identity and uninterrupted continuity.

Importantly, this is not to say that MFT's implicit moral self can be equated to Ricoeur's concept of *idem*-identity or to what he brings forward as the *idem*-aspect of character. Where the moral regularities may come close to what Ricoeur calls "habits", the innate origin of MFT's moral character is something Ricoeur would reject (e.g., Changeux & Ricoeur, 2000). By interpreting MFT's implicit concept of moral self in terms of *idem*-identity we merely emphasize its permeation with sameness.[39] It is an example of what Ricoeur calls '...the inscription of character in Sameness' (Ricoeur, 1992, p. 119 nt. 4). That is, an idea of moral self or moral character that is simplified and ossified as it earns stability solely in terms of "what" without making reference to a "who", nor to the reflexivity of self-constancy. In other words, what is hard to grasp for MFT is the idea that *someone is relating to one's own moral values*. For Ricoeur, in contrast, it is clear that even at the pole of moral character the self never vanishes entirely, as 'one cannot think the *idem* of the person through without considering the *ipse*, even when one entirely covers over the other' (Ricoeur, 1992, p. 121)

Without doubt, the three senses of sameness are vital to empirical analyses: first, scores on the different moral values belong to one and the same (numerical) individual, defining and individualizing him or her in moral terms, giving the opportunity of re-identification. Second, MFQ-scores also offer stability in the sense of qualitative identity. Two individuals that have the same score on a certain item or on the aggregated foundation score are regarded as the same, in the sense of being similar. It gives the ability to compare individuals in moral terms. Thirdly, the MFQ-score defining an individual's moral character offers stability in the sense of uninterrupted continuity. It offers continuity of an individual in moral terms over time and in different situations.

All three senses of sameness, harboured by an individual's moral character in terms of MFQ-scores, are prerequisites to doing meaningful empirical analyses. It is because scores can be attributed to one and the same numerical individual that certain combinations of scores can lead to associations between variables (such as values, attitudes, and behaviours) on the population level (e.g., regularly finding the combination of relatively high scores on variable X and Y within different individuals, while also regularly finding the combination of relatively low scores on X and Y, leads to a positive association between the two variables). To establish such associations we need, of course, a measure of qualitative sameness between individuals. These associations can only be meaningful when the scores express a certain continuity in an individual's morality.

As should be clear, though essential to empirical psychological investigation, this approach easily loses sight of the other crucial aspect of moral personhood that has been brought forward by Ricoeur: selfhood. This is, of course, not a particularly surprising

[39] It is in this specific sense that we will refer to MFT's concept of the moral self in terms of *idem* in the rest of our analysis.

conclusion, nor a fault of psychological methods. The psychological sciences necessarily take an observational and thus third-person perspective, aiming at scientific objectivity. However, this does not make the first-person perspective, which phenomenology can bring forward, less relevant for understanding morality and moral behaviour and thus for the central aims of (moral) psychology. Phenomenology is able to articulate the person as subject and its structures of experience that are central to the moral life, reaching a verity level of attestation. Empirical psychology is able to objectify these experiences, losing a certain richness in experiential information, but enabling methods of generalization on the population level and of prediction that can claim scientific objectivity. The approaches complement each other.

At the same time, following our analysis, it can be concluded that the three notions of virtue theory that have been related to MFT actually presuppose *ipse*. When general moral value theories would accept such a conception of the moral self, as we think they should, this is not without implication for their theory and measuring methods. In the discussion we will further go into what it could mean for the empirical study of morality to incorporate *ipse*-identity to a certain extent.

4.4.2 Reflexivity (*ipse*) in a theme park

We will now continue by making clear what a moral self in terms of *idem* without *ipse* and the mediation of narrativity amounts to and how it contrasts to a moral self that does incorporate them. This is best described using an example. Let's take the situation where you are waiting in line for a ride in a theme park. Two boys of about 10 years old sneak in and cut in line just in front of you. Now say that you in general highly endorse the moral value of fairness in the sense of MFT. You have developed a certain sensitivity for social situations in which this value is jeopardized, as in the situation you find yourself in right now. In fact, part of the development of the fairness foundation into your morality is the development of the more specific not cutting in line rule. The not cutting in line rule is what you, in general, find morally important. In case you would fill in the MFQ you would score highest on the fairness foundation, your score on the care foundation would be somewhat lower and the binding foundations again a bit lower. If the idea of a moral self solely in term of *idem* is correct then this situation would always lead you to judge the behaviour by the boys as wrong (following your intuitive dislike evaluation) and, accordingly, to consider telling them off (or some other proportional action) as the right thing to do.

Of course, such a course of events is conceivable. But another scenario may just be as conceivable. Namely, that you, though perhaps initially put off by the cutting in line of the boys, remember yourself at that age. How you used to play around with your brother, being cheeky sometimes but not intending any harm. Feelings of affection and care come up and, at the same time, you can feel again the emotional stress when some older person told you off in these cases, as well as the disappointment you imagine the boys would feel when you tell them to get out of the line. You look around you, there are mainly adults in the line. Besides, it's a really quiet day at the theme park and the waiting times for the different attractions have been short. You decide to let them be and have them enjoy their ride. Now the point here is not that this scenario is more probable than the former, or that this is necessarily the right decision. Rather, the point is that it is at least imaginable that someone would reason and feel like this, even when fairness is regarded as his most important value in general, and that it seems to be a possible morally acceptable way of dealing with the situation. In fact, it is conceivable

that, to the decision maker, this decision is what it means to act 'fair' in these circumstances.

Would such a scenario be possible if the moral self solely consists of *idem*-identity? It seems not. What the second story implies is *reflexivity*. Recall how reflexivity refers to the relationship towards oneself, i.e. an agent is able to recognize herself as the subject of a certain action. The question becomes what your *own* values tell you to decide in this situation. In other words, what is emphasized in terms of selfhood is not that fairness, care and binding values are the ones that are always called upon by you when you take moral decisions. Rather, what is important is that in recognizing values like fairness, care and loyalty as your own values when making a moral decision in a situation, you recognize yourself as a moral agent.[40] That is, these values are part and parcel of your moral identity, or selfhood. The permanence in time at play here is self-constancy: you are responsible for your actions in the double sense of others being able to count on you and you being accountable for your actions. You are holding firm, not in the sense that you stubbornly hold on to fairness ('whoever comes in line first, is first to take the ride') but that you recognize yourself in your decisions and actions. Letting the boys get in front of you in this situation is consistent with who you are.

Note how the decision here is highly influenced by the concrete social context in which you find yourself. The moral self is embedded in a concrete practical field of actions that comes with its own specific rules and values to which it attunes. You are in a theme park, i.e. a place imagined and designed with primarily children and young adults in mind. The park, from the attractions to the food options and from the availability of baby change rooms to the walking routes, is catered to their desires and needs. Adults are 'less important' in such a place. It's even the implicit rule of theme parks to be in a good mood, have fun and let children do things that they normally can't do, skipping the line may be one of these things. You pick up the signs that help you guide your moral decision in this specific situation. Before letting the boys go first, you checked whether there were no other children waiting in line. You estimated that the other adults in line will not mind that you let the boys go first. Furthermore, lecturing the boys about the importance of queuing seems especially 'out of place' and even against the unwritten rules of a theme park. Letting the boys go first is a decision that attests to the moral self as embedded in and attuned to a concrete situation.

Given the above analysis, we argue that whenever you need to take a moral decision the narrative plays a mediating role between yourself and your values. Before taking the decision, you were reminded of you and your brother at the same age, a concrete episode in your life story. How the fairness, care and binding foundations play out as moral values that may be called upon to guide moral decisions in concrete situations was influenced by this. As we have discussed above, Ricoeur argues that narrative mediates the character traits of *idem* and the self-constancy of *ipse*. Fairness, care and binding values are taken up in a concrete situation by narratively connecting them to the moral agent. In appropriating these values, you make them your own, you acknowledge them as part of your self-narrative, or life story. This process of appropriation is done through a plot; integrating the heterogeneous through a 'discordant concordance', providing a narrative unity to different events and situations over time and figuring the values that are important to you. The narrative answers the

[40] This decision process does not necessarily have to take a conscious deliberative form, but could also be intuitive.

question *who* is the moral agent? In answering this question, you tell your life story, i.e. you relate to those aspects of your life story, how you acted and were affected, that are important to this situation and that are imbued with evaluations.

Here, in retelling your life and connecting it to you as a kid, having fun with your sibling, not being preoccupied with rules and conventions but without intending any harm, you answer the question of the *who*. At the same time, your values are given life by being confronted with the narrative of the moral agent in a concrete situation. In finding yourself in line in a theme park with the boys trying to get in front of you, you are confronted with the question what these values mean for you in this specific situation. In answering that question, you make use of narrative to reinterpret fairness, care and binding values, ultimately relating them to who you are. In a theme park confronted with boys who want to get in front of you in line, you tell of yourself as a person who values fairness, care and binding values and has a brother with whom he played and had fun. Subsequently, the decision you take is informed by and shapes the story you tell about yourself. The self-narrative organizes your life, integrating your past, present and future, and you bear this life story with you and put it at play in every moral decision you take. Letting the boys go first is a decision that attests to the narrative nature of the moral self.

4.4.3 Problems for predicting behaviour from general moral values

Now, what does the above analysis mean for predicting behaviour from general moral value measures? First of all, it can be said that general moral questionnaires, like MFQ, do tap into people's self-narrative. Questions that ask you to reflect on certain moral considerations and moral statements ask for self-interpretation and actually presuppose *ipse*-identity. Questionnaires measuring moral values or moral personalities, then, do not go beyond, but rather make use of people's self-narratives. The result is a certain reflection of (parts of) this self-narrative. The main problem arises with how this reflection is subsequently interpreted and treated. Reduced to a stable set of general moral value traits, the self-narrative loses its defining features, namely that it is an interpretation that is in need of constant reinterpretation; an idea of self that is in constant flux and fits and adjusts itself to the different contexts it encounters. With this reduction, one's moral identity loses its defining *ipse*-aspect, exposing at least two problems for the prediction of behaviour from general measures of moral values.

The first is that a general questionnaire fails to grasp the flexibility of decision-making and action that our values permit, due to the reflexivity given in *ipse*-identity. Though MFQ-scores may be regarded as a possible expression of one's interpretation of the moral self in terms of general moral values, it fails to grasp that these values are related to and reinterpreted in every new situation through the mediation of narrative, in order to decide what to do. By fitting the episode of the decision situation into the other episodes that make up our life story, we again question and appropriate our general moral values in a way that is specific to the decision's context and that fits them into who we are. Furthermore, within a specific situation, you may call upon specific parts or episodes of your self-narrative. In other words, a particular version of you that does not feature as prominent in the general interpretation of yourself (e.g., think of recognizing yourself in two little kids in a theme park). The meaning that general moral values acquire within a specific decision situation, their mutual relationship, and what kind of behaviour they determine, is thereby highly contextual and hard to predict. This flexibility is not simply incorporated by measuring values (or modules) that are

somewhat more specific or contextualized (like the cutting in line rule), as a person still needs to relate to such (more specific) principles within a particular situation, implying the *ipse*-aspect of the moral self. Though measuring more specific values can be expected to render somewhat better results, as it may hold better information of the role of certain specific moral values in that context.

A second problem that the above analysis suggests is that individuals may have dispersing understandings of the moral concepts figuring in the questionnaire. If we, indeed, understand our moral self and moral values through narrative, this suggests that when we fill in a questionnaire, which triggers reflection on moral concepts and values, we also use certain episodes and situations to see what we think. These episodes can be ones that we have appropriated as part of our life story. This means that broad moral concepts like "unfair treatment" or "emotional suffering" are understood through and are given meaning within our particular life stories. The meaning we give is thereby never really general, but always has some specificity. This can have the result that, for instance, unfair treatment for one individual is predominantly understood in terms of the unequal treatment of different groups of people by institutions, due to one's life story, while another person may predominantly relate to other kind of episodes, such as about persons that give themselves a preferential treatment (i.e., when cutting in line or when cheating in a game). People may also think of widely differing contexts, ranging from unfair treatment in the work-place to the unfair treatment of animals. The problem is that these different understandings are not reflected by people's scores on generally formulated questions. Behind two relatively high scores on the importance of fairness for one's moral judgments can exist two quite different meanings, stemming from different narratives and life stories. As it is assumed that these are the meanings behind the scores that trigger behaviour, it seems logical that this dispersion affects the predictive value of these items. Making items more specific to a certain context may be a way to reduce this problem. This is of course at cost of the idea of a general moral value that has predictive value over many different contexts. But, as we have argued, this is not how we believe that the moral self in relation with its moral values functions.

4.5 Discussion

Like any full-fledged moral theory, either normative or descriptive, MFT assumes a conception of moral self. Using insights from Ricoeur's notion of personal identity, we have argued that MFT assumes an underdeveloped concept of moral self, which is reflected in a naïve way of measuring people's morality. Following Ricoeur, we have presented an improved concept of moral self for the empirical study of morality.

As we have aimed to make clear, Ricoeur's narrative concept of the moral self connects to aspects of virtue theory that the developers of MFT themselves consider as part of their theory. The notions of *moral character*, *a self that attunes to and is embedded in the social context*, and *narrativity*, are elements that are claimed by Haidt and Joseph (2007) to underlie MFT, but which are neither further developed nor connected to a more elaborated concept of moral self. With Ricoeur's notion of the moral self, we were able to further develop these elements and show what it would mean for MFT and for measuring morality when these would be seriously incorporated. Our analysis thereby exposes an ambivalence within MFT on a fundamental level. Accepting

the proposed more substantial moral self leads to a clash with those elements of the core theory that may only claim a moral self in terms of *idem* and to rejecting its according method of measurement (MFQ).

In particular, considering MFT's theory, it follows from our analysis that accepting a certain attunement to the social context and a place for narrativity in moral judgment and decision making is not intelligible without accepting the reflexivity of the self (*ipse*). That is, if MFT is serious about incorporating these elements, it must acknowledge a moral self who relates to one's moral values and moral regularities, if only, by interpreting them and deciding between conflicting or multiple possible ones in a situation. Yet, this, in turn, strikes at the roots of MFT's core theoretical idea of morality simply consisting of input-output moral regularities that can be aggregated in terms of general moral dispositions and of the implicit essential concept of moral self that follows from it. MFT would have to leave its underlying idea of decision-making where a defined set of general moral values as moral dispositions causally determine behaviour across contexts. Instead, general moral values should rather be regarded as touchstones that are called upon in decision-making to figure out whether a specific action is in line with one's moral understanding of oneself. Ricoeur's concept of the narrative moral self makes such an idea of decision-making and its according interpretive and dynamic relationship with one's values intelligible. It is the narrative aspect of the moral self that leads the way here, where action depends on one's ideas about "the good life", turning decision-making and action into an interpretation of the self in situation, offering flexibility as well as stability.

In its wake, the improved conception of the moral self, incorporating both *ipse*- and *idem*-identity and the mediation of narrativity, exposes considerable difficulties for measuring people's morality in general terms and using these to predict behaviour across a variety of contexts. Our analysis implies that the specific meanings that our moral values receive, their importance vis-à-vis each other, and the decisions and actions they determine are situation-specific and, therefore, cannot simply be measured in a general way, out of context. Another aspect complicating prediction from general measures of moral values, is that the concepts used in the questionnaire do not have a univocal interpretation. Following our analysis, people give meaning to moral concepts, such as 'fairness' or 'emotional suffering', through their particular life stories. Similar scores on a moral foundation may therefore actually harbour quite diverting meanings.

In sum, due to the mediation of the narrative moral self, the relation with our general moral values and –thereby the relationship between general moral values and behaviour- is interpretive in nature, instead of being characterized by causal determination. Therefore, the influence of moral values on behaviour varies in kind (i.e. depending on its specific meaning) and strength across persons and contexts. This dynamic relationship makes it fruitless to predict behaviour from moral values when *not* taking these specificities into account, as empirical moral value theories try to do.

As discussed, the fact that MFT's measurements solely reflect *idem*-identity, seems to be part and parcel of the observational, third-person stance of the empirical sciences. By definition, this view focuses on the 'what' of the 'who'. For empirical sciences, the phenomenological structure of the self is hard to grasp. On the other hand, empirical psychology is able to make generalizations about populations and predictions and can give insight into broader tendencies, which are not part of the

phenomenological toolbox. It is therefore not a matter of choosing one or the other, but rather of finding ways where the two realms can complement each other.

Though *ipse*-identity may be hard to grasp directly by empirical investigation, it can, to some extent, be taken into account in the used measuring method. In the following, we suggest possible ways for the empirical investigation of morality to incorporate the explicated phenomenological insights on the moral self. With regard to empirical moral value theories, this entails a substantial revision of their measuring method. First of all, in light of the dynamic and context-sensitive relation with our moral values, studies focusing on the influence of morality on moral behaviour should limit their study to a delineated context (i.e., a certain professional environment, a school, a retirement home, car drivers, or the playground). As people's moral values get their specific meaning and importance within a decision context, abstracting from this will inevitably lead to a loss of information. Though a certain level of generality is of course necessary to make general empirical claims, the more abstract these get, the more meaningless they become. For a better understanding and prediction of moral behaviour it is, therefore, necessary to understand the particular meanings of moral values and the importance given to them by individuals within that context.

This brings us to a second implication of our analysis for the study of morality in relation to behaviour. Given the context specificity of our moral considerations, as well as the richness of our self-narratives in terms of moral meaning, studies should take an exploratory rather than a confirmatory approach to measuring morally relevant phenomena, like moral values. Instead of imposing and limiting the choices of what can be possibly considered morally important or relevant for people from the top down, studies should start from a bottom-up approach to map out what morally matters to people when deciding within a certain context. This could be done by having quantitative studies be preceded by more qualitative investigations in which people's moral considerations are brought forward in a narrative form.

Here, the empirical study of morality could learn from qualitative approaches in the behavioural sciences. For instance, discourse analysis and social practice theory map out people's considerations, interpretations and social practices, playing a role in decision-making in a specific behavioural context (see Dickinson et al (2010) for an interesting study on people's considerations with regard to climate change and their choice on holiday travel mode). After mapping out such context-specific interpretations and considerations, these can subsequently be quantified into context-specific measures. Such a procedure is a double-edged sword, as it reveals the specific moral considerations that play a role within a certain context and, at the same time, measuring more precise concerns harbours less risk to arrive at diverting meanings. Also, the measurement of importance of such context-specific measures vis à vis each other can be expected to be more robust. Another procedure in which a qualitative phase informs quantitative analysis has been conducted by Boyd et al., (2015). In this study, on the influence of general moral values on everyday behaviour, participants were asked to describe their most important values in relation to who they are in their own words. From these narratives, people's values were deducted by counting certain theme words. These measurements proved to be more successful in predicting behaviour than pre-established moral scales. This indicates that starting from such narratives is a better way to grasp people's moral identities. Following our analysis, applying such a method to a more delineated context of behaviour may prove to enhance predictions further.

A third point of enhancement, following from our analysis, is designing the method of measurement in such a way that respondents actually need to rank moral concerns against each other. The dynamical aspect of the moral self, by relating to and interpreting one's moral values within a decision situation, involves weighing conflicting concerns against each other. This is not reflected in a general questionnaire, such as MFQ, where all items can receive the same score.

An example of a methodology that has been thoroughly developed in the last decades and which incorporates a large part of the above suggestions is Q-methodology. Combining qualitative and quantitative methods and focusing on one specific context, Q-methodology seems a promising approach for the empirical study of morality (Brown, 1980). Here, different perceptions about a subject, existing within a population, are measured. The first step is to exploratively collect statements, opinions, preferences etc. about a certain subject. For instance, this is done through conducting interviews with focus groups. Then, this broad collection is brought back to a representative set of statements. Subsequently, participants rank these statements vis à vis each other in terms of agreement. These individual views are correlated, resulting in several different perceptions on a subject that are, to a certain extent, generalizable to the population (Brown, 1993). Such subjectivities seem more insightful to understanding behaviour and the role of morality within a practical context and may also prove to be better predictors of behaviour within that context than general moral measures, which turn narratives into a general moral disposition. Of course, all above suggestions for enhancing the measurement of morality and its prediction of behaviour is subject to further empirical study.

Indeed, the described quantitative methods can still only grasp persons in terms of *idem*. However, they do this while taking into account the dynamic aspects and context sensitivity of *ipse*-identity, to some extent. With regard to empirical moral value theories, and MFT in particular, the question is to what extent these are willing and able to incorporate this aspect of personhood in their theory and method. In this article, we have aimed to show that if theories, like MFT, are serious about incorporating a more developed concept of the moral self, like the one we have explicated in this paper, and we believe they should be, their core theory as well as their method of measuring people's morality needs substantial revision.

References

Boyd, R., Wilson, S., Pennebaker, J., Kosinski, M., Stillwell, D., & Mihalcea, R. (2015). Values in words: Using language to evaluate and understand personal values. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 9, No. 1, pp. 31-40).

Brown, S. R. (1980). *Political subjectivity*. New Haven, CT: Yale University Press.

Brown, S. R. (1993). A primer on Q methodology. *Operant subjectivity*, *16*(3/4), 91-138.

Changeux, J. P., & Ricoeur, P. (2000). *What makes us think? A neuroscientist and a philosopher argue about ethics, human nature, and the brain*. Princeton University Press.

Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: A standardized stimulus database of scenarios based on moral foundations theory. *Behavior research methods*, *47*(4), 1178-1198.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology*, *107*(5), 943–963. https://doi.org/10.1037/a0037245

Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008

Chowdhury, R. M. (2019). The moral foundations of consumer ethics. *Journal of Business Ethics*, *158*(3), 585-601.

Clark, C. B., Swails, J. A., Pontinen, H. M., Bowerman, S. E., Kriz, K. A., & Hendricks, P. S. (2017). A behavioral economic assessment of individualizing versus binding moral foundations. *Personality and Individual Differences*, *112*, 49–54. https://doi.org/10.1016/j.paid.2017.02.052

de Jonge, J., & van Trijp, H. (2014). Heterogeneity in consumer perceptions of the animal friendliness of broiler production systems. *Food Policy*, *49*(P1), 174–185. https://doi.org/10.1016/j.foodpol.2014.07.008

Dickinson, J. E., Robbins, D., & Lumsdon, L. (2010). Holiday travel discourses and climate change. *Journal of Transport Geography*, 18(3), 482-489.

Dickinson, J. L., McLeod, P., Bloomfield, R., & Allred, S. (2016). Which moral foundations predict willingness to make lifestyle changes to avert climate change in the USA?. *PloS one*, *11*(10), e0163852.

Díaz, R., & Cova, F. (2022). Reactance, morality, and disgust: the relationship between affective dispositions and compliance with official health recommendations during the COVID-19 pandemic. *Cognition and emotion*, *36*(1), 120-136.

Doris, J. M., & Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. Cambridge University Press.

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017. *Personality and Social Psychology Review*, *23*(4), 332–366. https://doi.org/10.1177/1088868318811759

Frimer, J. A., & Walker, L. J. (2008). Towards a new paradigm of moral personhood. *Journal of Moral Education*, *37*(3), 333–356. https://doi.org/10.1080/03057240802227494

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental*

Social Psychology, 47, 55-130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and Conservatives Rely on Different Sets of Moral Foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitch, C., & Mooijman, M. (2018). Moral foundations theory. *Atlas of moral psychology*, 211-222.

Graham, J., Meindl, P., & Beall, E. (2012). Integrating the Streams of Morality Research: The Case of Political Ideology. *Current Directions in Psychological Science*, *21*(6), 373–377. https://doi.org/10.1177/0963721412456842

Graham, J., Nosek, B., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385.

Gray, K., & Keeney, J. E. (2015). Disconfirming moral foundations theory on its own terms: Reply to Graham (2015). *Social Psychological and Personality Science*, *6*(8), 874-877.

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., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55-66.

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. *The innate mind*, *3*, 367-391.

Halsema, A. (2019). The Self is Embodied and Discursive: Tracing the Phenomenological Background of Ricoeur's Narrative Identity. In S. Davidson (ed.), *A Companion to Ricoeur's Fallible Man* (pp. 125-144). Lexington Books.

Hoover, J., Atari, M., Mostafazadeh Davani, A., Kennedy, B., Portillo-Wightman, G., Yeh, L., & Dehghani, M. (2021). Investigating the role of group-based morality in extreme behavioral expressions of prejudice. *Nature Communications*, *12*(1), 1-13.

Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of management review*, *16*(2), 366-395.

Miles, A. (2015). The (Re)genesis of Values: Examining the Importance of Values for Action. *American Sociological Review*, *80*(4), 680–704. https://doi.org/10.1177/0003122415591800

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of Socialization Theory and Research* (pp. 347-480). Chicago, IL: Rand McNally.

Kohlberg, L. (1984). *The Psychology of Moral Development*. Cambridge, MA: Harper and Row.

Lapsley, D. K. & Narvaez, D. (Eds.). (2004). *Moral development, self, and identity*. Psychology Press.

MoralFoundations.org. (2013). Retrieved November 4, 2021, from https://moralfoundations.org

Nilsson, A., Erlandsson, A., & Västfjäll, D. (2016). The congruency between moral foundations and intentions to donate, self-reported donations, and actual donations to charity. *Journal of Research in Personality*, *65*, 22–29. https://doi.org/10.1016/j.jrp.2016.07.001

O'Grady, T., Vandegrift, D., Wolek, M., & Burr, G. (2019). On the determinants of other-regarding behavior: Field tests of the moral foundations questionnaire☆,☆☆,★,★★. *Journal of Research in Personality*, *81*, 224-237.

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour*. Sage.

Presti, S. L., Mattavelli, G., Canessa, N., & Gianelli, C. (2021). Psychological precursors of individual differences in COVID-19 lockdown adherence: Moderated-moderation by personality and moral cognition measures. *Personality and Individual Differences*, *182*, 111090.

Ricoeur, P. (2007). *Reflections on the Just* (D. Pellauer, Trans). The University of Chicago Press.

Ricoeur, P. (1992). *Oneself as another* (K. Blamey, Trans.). The University of Chicago Press.

Ricoeur, P. (1991). Narrative identity. *Philosophy today*, *35*(1), 73-81.

Ricoeur, P. (1984). *Time and Narrative. Volume 1*. The University of Chicago Press.

Ricoeur, P. (1975). Phenomenology and Hermeneutics. *Noûs*, *9*(1), 85-102.

Ross, L., & Nisbett, R. E. (2011). *The person and the situation: Perspectives of social psychology*. Pinter & Martin Publishers.

Schier, U. K., Ockenfels, A., & Hofmann, W. (2016). Moral values and increasing stakes in a dictator game. *Journal of Economic Psychology*, *56*, 107–115. https://doi.org/10.1016/j.joep.2016.06.004

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 25, pp. 1–65). Academic Press. https://doi.org/10.1016/S0065-2601(08)60281-6

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., ... & Konty, M. (2012). Refining the theory of basic individual values. *Journal of personality and social psychology*, *103*(4), 663.

Southerton, D., Warde, A., & Hand, M. (2004). The limited autonomy of the consumer: implications for sustainable consumption. *Sustainable consumption: The implications of changing infrastructures of provision*, 32-48.

Taylor, C. (1985). *Philosophical papers: Volume 1, Human agency and language* (Vol. 1). Cambridge University Press.

Vainio, A., & Mäkiniemi, J. P. (2016). How Are Moral Foundations Associated with Climate-Friendly Consumption? *Journal of Agricultural and Environmental Ethics*, *29*(2), 265–283. https://doi.org/10.1007/s10806-016-9601-3

van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2020). Does morality predict aggressive driving? A conceptual analysis and exploratory empirical investigation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *74*, 259–271. https://doi.org/10.1016/j.trf.2020.08.017

van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2022). Why are general moral values poor predictors of concrete moral behavior in everyday life? A conceptual analysis and empirical study. *Frontiers in Psychology*, 3834.

van Tongeren, P. (2020). *The Art of Living Well: Moral Experience and Virtue Ethics*. Bloomsbury Publishing.

Qian, K., & Yahara, T. (2020). Mentality and behavior in COVID-19 emergency status in Japan: Influence of personality, morality and ideology. *PloS one*, *15*(7), e0235883.

Zahavi, D. (2014). *Self & Other. Exploring Subjectivity, Empathy, and Shame.* Oxford University Press.

Zahavi, D. (2008). *Subjectivity and Selfhood. Investigating the First-Person Perspective.* MIT Press.

Zahavi, D. (2007). Self and Other: The Limits of Narrative Understanding. In *Narrative and Understanding Persons*. (pp. 179-201). Cambridge University Press.

5. Perpetrators of International Crimes and the Role of the Moral Self

This chapter is co-authored by Dr. mr. L. D. A. Corrias (second author).

Abstract

In this study we present a concept of the moral self that can serve as a fundament to the empirical study of perpetrators and resisters in the context of international crimes. Through providing an overview of the empirical field, we show that situational, dispositional and interactionist approaches presuppose a dynamic and reflexive concept of the moral self, which is not articulated. Through recourse to Paul Ricoeur's phenomenological insights on personal identity, we formulate such a dynamic and reflexive notion of the moral self, which is characterized by narrativity. Our analysis emphasizes both the narrative and reflexive structure of the moral transformation process of perpetrators. This narrative moral self grounds -and makes explicit- a true interactionist approach to explaining behaviour within the context of international crimes. Such an approach transcends the prevailing dichotomy in this field between explaining behaviour either in terms of situational or dispositional factors and goes further than regarding it as the simple product of both. Our analysis leads to several methodological implications for empirically investigating this behaviour.

Keywords: International Crimes; Perpetrators; Resisters; Interactionism; Moral Self; Moral identity; Narrative Identity; Ricoeur

5.1 Introduction

Within extant literature, perpetrators of international crimes[41] are generally typified as 'ordinary men[42] in extraordinary circumstances' (Smeulers 2019; Overy, 2014; Browning, 1992). This phrase brings two important characteristics to the fore, which are broadly considered to be typical for the events of international crimes and its perpetrators. The first pertains to the fact that these crimes are by definition committed in a collective context, and ordinarily are massive in scale insofar as they involve many perpetrators and victims[43]. Perpetrators are often embedded in an (state) institution or system in which a perverted moral climate reigns, and, in turn, commit their crimes either on the express orders of or with the knowledge of the ruling authorities (Smeulers and Grunfeld, 2011; de Swaan, 2015). The second characteristic is that the majority of the perpetrators of international crimes cannot be considered as either mentally ill or as possessing an aggressive or otherwise divergent personality, which is patently impossible given the numbers of perpetrators that are often involved (Waller, 2007), but rather appear to be relatively ordinary (Nollkaemper 2009; Punch, 2009; Drumbl, 2007; Browning, 1992).

Given these two characteristics, the predominant approach to explaining these crimes is "situationistic", that is, it refers to the extraordinary circumstances within a certain collectivity or system in which these "ordinary men" are embedded in order to explain their behaviour. Such situational factors include obedience to authority (Kelman and Hamilton, 1989; Milgram, 1974), conforming to peers (Asch, 1951; Browning, 1992), and atrocity-producing situations (Zimbardo, 2011). Together with important social-psychological mechanisms, such as cognitive dissonance (Festinger, 1957) and neutralization techniques (Sykes & Matza, 1957), these situational factors feature in extant explanations of the gradual transformation of a "normal" person into someone who is able to *go along* with (systematically) committing horrific acts (Staub, 1989; Smeulers, 2004). An important feature of this transformational process is that perpetrators adopt the perverted moral norms of the system, and, in turn, come to see their acts as legitimate and morally right. What was once regarded as wrong – participating in systemic harmful actions towards others- is now regarded *as the right thing to do* (Staub, 1989, 2014; Üngör, 2019; Waller, 2007; Smeulers 2012; 2004; Kelman

[41] International crimes are characterized and legally distinguished from conventional or domestic crimes, by the fact that perpetrators can be held individually criminally responsible under *international* criminal law (Shaw, 2008, pp. 399-400; Smeulers, Hola and Van den Berg, 2013, p. 7). What acts fall and should fall within the category of international crimes is the source of ongoing debate (see Einarsen (2012) for an interesting discussion). However, there is a general agreement that at least four crimes definitely come under this category, namely: genocide, crimes against humanity, war crimes, and the crime of aggression. The criminology of international crimes, which aims at both the description and explanation of these acts of extreme violence, has hitherto primarily focused on the events and perpetrators of the first three. These three crimes also constitute the focus of this paper.

[42] The vast majority of international crimes are indeed committed by men. However, women have also been involved. For an interesting study on the role of women in mass atrocity, see Smeulers (2015). Because the majority of perpetrators are males, we have opted to use masculine personal pronouns when referring to perpetrators.

[43] Though uncommon, following the juridical definition, an international crime can also be committed by the individual action of one person, as long as it falls within a period of structural violence. Structural violence does, of course, indirectly, refer to a relevant larger collective context.

and Hamilton, 1989; Bauman, 1989; Hilberg, 1961). In this respect, they can be said to go through a process of *moral transformation*.

However, not everyone who is embedded within the same systemic context acts in a similar manner. First of all, several types of perpetrators have been distinguished (Smeulers, 2008). Secondly, there are also always those people who have the capacity *to resist* the pressure to go along[44] with the perverted norms of the system (De Swaan, 2015). This suggests that people's personal dispositions are also relevant for understanding the different types of behaviour observed in this context. Extant research on the role of personal dispositions primarily focuses on relatively common personal dispositions, which in the specific context of international crimes could increase someone's propensity to commit crimes, rather than on outright violent personality traits or psychological disorders (Smeulers, et al., 2019). However, research in this area has thus far not generated any 'conclusive empirical evidence…as to what kind of personality traits would be more prominent within perpetrators of mass violence compared to other people' (Smeulers et al., 2019, p. 33). Despite this, the renewed interest in personal dispositions has nonetheless paved the way for exploring relevant interactionist insights into which dispositional and situational factors mutually and dynamically influence each other (Newman, 2002; Haslam and Reicher, 2007).

Interestingly, similarly to the more situational approaches, the explanations which take individual factors into account also typically refer to a moral transformation process by which perpetrators come to see their acts "as the right thing to do" (Haslam and Reicher, 2007; de Swaan, 2015). This means that, on the other side of the same coin[45], those who are able to resist within this context are the ones that make the opposite moral judgment. What becomes clear here, is that a key aspect of the explanation of the behaviour of perpetrators and resisters in the context of international crimes –whether from a situational, or dispositional/interactionist approach- is the moral dimension of the human being and its potential for transformation and resistance.

More specifically, the moral judgment that an act "is the right thing to do" reflects a person's individual morality, or better yet, is an expression of a person's moral identity or moral self[46]. By one's moral identity or moral self we mean someone's understanding of themselves as a moral person, which, in turn, stipulates what is right for him to do and regulates his moral behaviour. The moral transformation which perpetrators go through thus implies a change in moral identity in interaction with their environment, while, simultaneously, the individual continues to identify as one and the same 'good person'. In this respect, resisters can also be expected to go through some form of moral change in that they can no longer simply follow the ruling law, authorities, and social norms of society.

By making this moral dimension of the human being an important part of both the explanation and understanding of international crimes, empirical approaches thus

[44] In this article, when we speak of resistance it is always meant in this minimal sense of 'not going along' with the ruling systemic context.

[45] This framing is not to suggest that the individuals involved in the context of international crimes are designated to one of these two roles. Other relevant groups are, of course, victims and bystanders and it has been shown that these roles are dynamic and fluid during mass atrocities, i.e., individuals can shift from one role to the other (Anderson, 2019; Williams, 2020). Rather, we merely wish to say that the people who resist do make an opposing moral judgment with regard to those that do go along with it.

[46] We use the concept of moral self and moral identity interchangeably.

assume a concept of moral identity that although fundamental to their explanation, is not further articulated. Rather, the change in moral identity is merely presented as a factual outcome of situational and personal psychological processes. Therefore, the structure of moral identity itself, its capacity for change while remaining one and the same, and its pivotal role in constituting behaviour within the context of international crimes is not further investigated. Given that it forms the basis upon which further (empirical) theorizing is built, getting a better grasp of this foundation of moral agency can both greatly enhance our understanding of behaviour within this context and open up expedient avenues for future research.

In this paper we aim at articulating a concept of moral identity that can serve as the fundament to extant empirical approaches explaining behaviour within the context of international crimes. In section 5.2, we begin with a (non-exhaustive) overview of empirical explanations of this behaviour. Here, we show that people's moral identity and its capacity for transformation and change constitutes an important underlying assumption of these theories. Furthermore, we show that the moral transformation process involves a change in the *interpretation of the self* in moral terms, that is, that the person comes to interpret themselves as someone who is morally able to commit such acts. This implies that the concept of moral identity or moral self that is assumed in extant research is *reflexive*, insofar as it entails a relation of the self towards the self. Or, as Philip Pettit (2001, p. 79) describes the reflexivity involved in the self: 'That an agent is a self means that he can think of himself, or she can think of herself, in the first person as the bearer of certain beliefs and desires and attitudes and as the author of the actions, and perhaps other effects, to which they give rise.'

To be able to articulate such a dynamic and reflexive concept of moral identity, we adopt what philosophers call a phenomenological approach. Phenomenology is characterized by adopting the perspective of the first-person, or 'a stance of radical reflexivity', as Charles Taylor (1989, p. 130) referred to it:

> The world as I know it is there for me, is experienced by me, or thought about by me, or has meaning for me. Knowledge, awareness is always that of an agent. (...) In our normal dealings with things, we disregard this dimension of experience and focus on the things experienced. But we can turn and make this our object of attention, try to experience our experiencing, focus on what the world is *for* us (p. 130).

In section 5.3, we draw upon Paul Ricoeur's (1992) phenomenological insights into personal identity and narrative self in order to explicate a concept of moral identity that exists in dynamic interaction with its environment, and which is both reflexive and characterized by self-interpretation. In section 5.4, we argue that this concept of moral identity transcends the overly simple dichotomy between situational factors and personal dispositions, and, as such, can serve as a fundament to more integrative approaches, such as dynamic interactionist accounts of behaviour in the context of international crimes. Next, we briefly delineate some methodological directions for future research that follow from this analysis. In the conclusion, we provide a brief summary of our argument and main findings.

5.2 The moral and reflexive dimension of the transformation process featuring in the explanation of perpetrators

To bring forward the role of morality and -more specifically- moral identity and its reflexivity in explaining international crimes, we now explore some of its most prominent explanatory approaches. First, we focus on obedience to authority as an important situational factor in explaining these crimes, specifically Kelman and Hamilton's (1989) theory of obedience to authority, which attests to a process of moral transformation. Next, we look into social-psychological mechanisms of cognitive dissonance and neutralization techniques that are regarded as essential to the transformation process. In the third sub-paragraph, we examine relevant (moral) dispositional factors and the interactionist approach. Finally, we discuss what this cursory overview tells us about the concept of moral identity that appears to be presupposed by these empirical approaches.

5.2.1 Obedience to authority

The systemic and hierarchical context in which international crimes more often than not take place, has led scholars to claim that the essence of these crimes is the fact that they are crimes of obedience (Kelman, 2009; Smeulers and Grunfeld, 2011). The relevance of "obedience to authority" for explaining the harm carried out in the context of mass atrocity was first brought forward by the seminal experiments on obedience conducted by Stanley Milgram in the 1960s. This link was further theoretically developed by Kelman and Hamilton (1989) in their book *Crimes of Obedience*. In the following section, we first briefly discuss Milgram's obedience experiments, before then proceeding to unpack Kelman and Hamilton's theory of crimes of obedience.

Milgram's obedience experiments

Milgram conducted a host of experiments to test to what extent and under what circumstances subjects are willing to administer what they believe to be painful shocks to another person on the orders of an authority figure. The subject is told that they are taking part in a learning experiment, where the subject plays the role of a 'teacher' testing the memory of a 'learner', who -in the baseline set-up- is seated in an adjacent room. Every time the learner gives a wrong answer, the teacher is ordered by the experimenter to administer gradually increasing electric shocks, starting with 15 Volts and rising up to the maximum of 450 Volts, by pushing a lever. Unbeknownst to the subject, the learner in the other room is actually a confederate of the experiment, who is purposely giving wrong answers at fixed moments and receives no real shocks. Of course, the learner does act as if they are receiving painful shocks. This starts off with a grunt at 75 Volts, before rising to screams and predetermined lines such as 'get me out of here', which is clearly audible to the subject giving the shocks. When they at some point hesitate or refuse to administer the shock, the experimenter orders the subject to continue with the experiment, using standard probes, such as 'the experiment requires that you continue' and 'you have no other choice, you must go on' (Milgram, 1974).

Infamously, the majority of the subjects were willing to administer severe shocks. In the above-described base-line condition, 62.5 per cent of the subjects continued up to the highest level of 450 volts. This was a much higher obedience rate than expected (Milgram, 1974). Most subjects did show signs of discomfort and tension during the

process, while many at some point protested in some way or begged the experimenter to go check on the learner (Milgram 1965). Subsequent variations of this experiment showed that several factors considerably influenced obedience rates. For instance, when the subject was physically closer to the learner, such as when the subject had to physically touch the learner to administer the shocks, obedience rates dropped. The position of the experimenter also influenced obedience rates. For example, in variations where the orders were given from a different room, when the experimenter left and a random person took their place, and when a second experimenter contradicted the given orders, people were less likely to obey (Milgram, 1974).

Many scholars saw the results of this experiment as yielding important insights into the influence of an authority figure on behaviour and explicitly related this to their explanations of the behaviour of perpetrators of international crimes (Waller, 2007; Browning, 1992; Smeulers and Grünfeld, 2011; Harrendorf, 2014; Kelman and Hamilton, 1989). At the same time, the experiments received substantial criticism. Besides the evident ethical issues, pertaining to the stressful position the subjects were put in (Perry, 2013), both the interpretation of the results of the experiment (Haslam and Reicher, 2017; Smeulers, 2020) and what it actually says about behaviour outside of the laboratory setting (Burger, 2014; Fenigstein, 2015) have been the source of ongoing critical debate (Smeulers, 2020). Moreover, the perceived realness of the experiment has also been questioned, namely concerning the perceived realness of the shocks (Perry, 2013), and the ambiguity of the cues on the harmfulness of the shocks (Harrendorf, 2014). One example is that the experimenter would reassure the subject during the experiment that the shocks would not leave any permanent damage, whilst the lever with the highest shocks read: 'Danger, Severe Shock.'

In questionnaires completed right after the experiment, however, the majority of the subjects did indicate that they believed the shocks were real (Smeulers, 2020). Milgram (1965, 69) also took the observed tension people were experiencing as evidence that they perceived the situation to be real. In subsequent years, there have been several replications of the experiment, including ones that removed any ambiguous cues (see Smeulers and Grünfeld, 2011, pp. 222-227 for an overview). These have generally confirmed Milgram's results (Blass, 2009; Beauvois et al., 2012). Smeulers and Grünfeld (2011, p. 232) conclude that, at the very least, the experiments show that 'people without bad intentions can get locked up in a situation in which they feel obliged to obey an authority and thus hurt others' and that 'when confronted with two conflicting demands: namely, to obey an authority and the moral value not to hurt anyone else, many people yield and obey the authority'.

Kelman and Hamilton's crimes of obedience theory

One study that has aimed to provide a thorough theoretical explanation for this phenomenon in the context of international crimes, emphatically linking obedience to authority to the perpetration of international crimes, is Kelman and Hamilton's (1989) book *Crimes of Obedience*. It develops a substantial theoretical explanation in which three processes are presented as of central importance: *authorization*, *routinization* and *dehumanization*.

"Authorization" is described as a process of social influence by which an individual perceives an authority figure and their orders as legitimate and therefore follows it. The concept of "a perceived legitimate authority" indicates a role-relationship

in which one person is perceived as possessing the power and right to order a second person who, taking the role of the subordinate, feels they ought to voluntarily obey out of a sense of duty (Kelman & Hamilton, 1989; Tyler, 1997, 323; Lamb, 2010). The subordinate feels a moral obligation to follow the orders, that is, a duty that entails the acceptance of the order on a moral basis –it is judged to be "right"- rather than for instrumental reasons, such as personal gain: 'The person's focus is not on what he wants to do... but on what he sees required of him.' (Kelman & Hamilton, 1989, p. 90). Whether or not an authority and its orders are perceived as legitimate depends, for instance, on whether the larger system is perceived as legitimate, how an authority gained their position, and the authority's behaviour in this role (Kelman & Hamilton, 1989).

According to Kelman and Hamilton (1989), the reason people feel committed to their role as subordinates in the first place is because it is personally meaningful to them. This is because a role not only defines someone's duties but also someone's rights, and, hence, it can become part of someone's identity. It provides a means through which individuals can maintain a meaningful relationship with their group and present themselves as being a good group member (Kelman and Hamilton, 1989). On the moral level, this entails that the subordinate feels a responsibility to follow the authority in order to fulfil their role well, while the moral meaning of the action itself becomes less important (Smeulers and Grünfeld, 2011, p. 219). An orientational shift then takes place from personal preferences and moral considerations about the act itself to role requirements. Here, moral considerations do not vanish, but rather transfer to how the order is fulfilled. The ultimate effect of authorization in this context is therefore a moral recategorization from the morally unacceptable to the morally acceptable and obligated. That is to say, the harmful act comes to be seen as the right thing to do, which is regarded as a central feature of crimes of obedience in the context of international crimes (Kelman and Hamilton, 1989; Kelman, 2001).

However, even if an order to participate in mass violence is perceived as legitimate, it can still lead to moral inhibition and further moral questioning. In order to quell the potential rise of moral questions, organizations involved in the ordering of international crimes tend to *routinize* the subordinate's criminal conduct (Kelman, 2009). Activities are divided into separate tasks and allocated to different individuals who are only given control and responsibility over their own subcomponent of the operation. As individuals only focus on performing their own task, their conduct increasingly consists of '...routine, mechanical and highly programmed operations' (Kelman and Hamilton, 1989, p. 18). Individuals simply carry out their routine without having to think about either the entire process or the potential immoral consequences. Furthermore, this division of work results in a "diffusion of responsibility", so that nobody feels truly responsible for the actual acts that are committed (Kelman and Hamilton, 1989). The process of routinization thereby brings about a normalization of violence: 'altogether, the actions come to be seen as part of a normal job rather than participation in massacre or genocide' (Kelman, 2009, p. 37).

Despite this, participation in systematic violence may still be confronted with inhibition (Kelman & Hamilton, 1989). The process of *dehumanizing* the targeted group can contribute to the lowering of such inhibitions. Through the process of dehumanization, the perpetrator comes to exclude victims from their moral 'universe of obligation' (Fein, 1984, p. 6). They are no longer deemed to be genuine human beings to which standard moral principles apply (Kelman, 2009; Kelman and Hamilton, 1989).

In most instances, dehumanization forms part of a larger ideology that stereotypes and stigmatizes a certain group along racial, ethnic, or religious lines and deems them to be inferior (Alvarez, 2009). During the 1994 Rwandan genocide, for instance, Hutus systematically labelled Tutsis as "cockroaches" (Mullins, 2009). In bureaucracies, dehumanization can be realized through a technical approach to the killings, such as, for example, presenting victims merely as bodies to be counted (Kelman and Hamilton, 1989, p. 19).

The three processes of authorization, routinization and dehumanization help to explain how a normalization of violence becomes possible. In particular, through a shift from standard moral considerations about the moral meaning of a violent act to an orientation on role obligations and morality qua duty (Kelman and Hamilton, 1989; Bauman, 1989; Hilberg, 1961), the perpetrator undergoes a moral transformation by which the acts come to be seen as the right thing to do. This moral shift is accordingly presented as an important aspect of how perpetrators come to commit these crimes.

5.2.2 Social-psychological mechanisms involved in the transformation of perpetrators

People do not turn into mass murderers overnight. Rather, studies investigating how (relatively) ordinary people come to commit international crimes almost univocally refer to a certain *gradual process* of transformation by which a person adopts a role within a criminal system and embraces its perverted morality (e.g., Waller, 2007; Smeulers, 2004; Staub, 1989). Within Kelman and Hamilton's theory, this gradual transformational process can clearly be discerned in the interaction between the three aforementioned processes. Similarly, explanations that refer to other situational or social pressures, such as conformity to the peer group (Browning, 1992), or those that position the atrocity producing situation as being crucial (Zimbardo, 2011), ordinarily situate this explanatory factor within a larger process of (moral) transformation.

There are two social-psychological mechanisms which often play a central role in these explanations of the transformational process, namely *cognitive dissonance* and *neutralization techniques*. In certain respects, these two mechanisms can be regarded as the conductors of the transformation process, which serve to turn (initial) action in response to social pressures into an almost unstoppable train of violence and an internalization of its corresponding norms.

Cognitive dissonance refers to the uncomfortable feeling that arises when someone experiences an incongruity between their actions and internalized norms and beliefs (Festinger, 1957). Given that people feel the need to be consistent with themselves, this discrepancy needs to be neutralized (Staub, 2011, p. 171). This can either be realized by changing one's behaviour (and admitting wrongdoing for past behaviour) or by changing one's beliefs and understanding of the situation. As people have a strong urge to justify their committed acts, the second option is often the preferred means through which this incongruence is resolved (Aronson, 2004). Within the context of perpetration, this means that instead of resisting the pressure to harm others, perpetrators instead begin to interpret the harm they are doing in a way that makes it acceptable and justifiable to themselves.

This manipulation of beliefs and the meaning of actions is facilitated by justifications and rationalizations of behaviour, known as neutralization techniques, which Sykes and Matza (1957) first described in the context of youth delinquency. They

distinguish between five techniques: denial of responsibility, denial of injury, denial of victim, condemning the condemners, and appealing to higher loyalties. By, for instance, denying responsibility for one's conduct or by denying even that there was a genuine victim –in the eyes of the perpetrator the victim deserved this treatment or were somehow to blame for what happened- the immoral meaning of actions and one's blameworthiness are effectively denied. These rationalizations can function as a justification after the fact to assuage one's conscience, while, simultaneously, by changing one's beliefs, also motivates to engage in further harmful behaviour (Sykes & Matza, 1957).

According to Smeulers (2004), the reaction after a first harmful act is decisive. Once individuals start to rationalize their actions, it becomes increasingly more difficult to reconsider how such actions are viewed. This would require admitting to oneself that one was wrong before, both in terms of one's action and judgment. It leads to processes which Bandura (1999) referred to as "moral disengagement", by which people somehow deny the immoral meaning of and moral responsibility for their conduct and evade self-sanction. At the same time, the rationalizations change people's (moral) beliefs, motivating them to engage in inflicting further harm. The combination of cognitive dissonance and neutralization techniques can therefore lead to a lethal trap. Through forms of self-justification, an individual reinterprets both one's actions and, ultimately, oneself as someone who is allowed or even obligated to perform these acts.

Falling into this trap becomes more likely when one is gradually led into the behaviour, with an initial small and non-significant first involvement and then subsequent incremental increases. As the differences between steps appear rather insignificant, refusal at some point requires an admission that one has done wrong already, thus leading to cognitive dissonance. This "foot in the door technique" (Burger, 1999; Freedman and Fraser, 1966) can be discerned in Milgram's experiments (Burger, 2014), where participants started off with a rather light shock that gradually increased by only 15 volts each time. Also in situations of mass atrocity we often see such a process. As Smeulers (2004) makes clear, there is usually an initiation phase, where prospective perpetrators become gradually involved in aggression and violence. She cites the example of the Greek torture school, as described by Haritos-Fatouros (2003), where torture recruits would first only witness and clean-up after torture sessions, and occasionally be ordered to (physically) mistreat prisoners, before eventually fully participating in the torture sessions themselves.

Staub (1989; 2011) places the above transformation process within the larger macro-political context and makes clear that a 'continuum of destructiveness' is constituted. The first step in the described process is for an individual within a collective to be influenced by a ruling destructive ideology, which scapegoats and devalues members of another group, often consciously imposed by political leaders during periods of social and economic upheaval. As individuals are committed to their group and their own respective role within it, and given that these destructive beliefs become more common and institutionalized, individuals are more likely to engage in harmful actions towards members of the devalued group. Once they become involved in these initial harmful acts, the already developed negative beliefs about the other are accelerated through further rationalizations to alleviate the cognitive dissonance. Staub (2011) describes how individuals, in this way, change their behaviour and their beliefs as they "learn by doing". This continuum of destructiveness can lead to a total '*reversal of*

morality' (Staub, 2011, p.168) (emphasis in original), where 'Killing, eliminating this particular group and its members, comes to be the right, moral thing to do' (Staub, 2011, p. 168). According to Smeulers (2012), it is indeed the human ability to manipulate (moral) reality as well as the attendant belief in the legitimacy of the criminal orders given and the perverted moral norms ruling in the context of collective violence that help us explain 'why so many ordinary and law-abiding citizens become involved in international crimes and become law-abiding criminals' (p. 27). [47]

What emerges from the description of the transformation process is that it presupposes a reflexive attitude on the behalf of the people involved. This transformation takes place as people reinterpret *themselves*, in an effort to remain consistent with *themselves*, thus leading to a new understanding of *themselves*. Furthermore, this understanding has an important moral dimension, namely that the person wants to maintain their self-concept of being a good person and, as such, reinterprets what actions are regarded as morally acceptable and obligated. Hence, this presupposes a notion of moral identity that remains constant ("a good person"), while, simultaneously, changing how someone's values and beliefs are understood within a specific context (law-abiding citizens become law-abiding criminals). This moral transformation is of paramount importance to this explanation. As was also the case in Kelman and Hamilton's crimes of obedience theory, the fact that the perpetrator comes to see his conduct as "the right thing to do" is presented as an important motivational factor for the perpetration of mass violence.

5.2.3 Personal dispositions and interactionism

Individual dispositional factors

Although the situationistic approach has been widely utilized to explain the behaviour of perpetrators of international crimes, there has also been critique concerning the one-sidedness of these explanations. For instance, the different types of perpetrators that can be distinguished within one systemic context already suggests the relevance of more than just situational factors (Smeulers, 2008). This idea becomes even more lucid when considering the fact that there are also people who do not conform and instead resist these situational pressures (de Swaan, 2015). For instance, while the Milgram experiment has generally been interpreted as showing that ordinary people will obey the violent orders of an authority, the results also demonstrate that a considerable amount of people, namely one-third to four-fifths of the subjects, depending on the setup, do *not* obey (Jetten & Mols, 2014; Miller, 2014; Blass, 1991). Similarly, studies investigating actual international crime situations have shown that there are people who resist by not going along with the systemic context and its norms (Whitehead, 2021). This suggests that there must be something about the individual that also plays a role here. Therefore, many contemporary scholars posit that we need both individual and situational factors to explain behaviour within the context of international crimes (e.g., Smeulers e.a, 2019; de Swaan, 2015; Staub, 2014, Halsam and Reicher 2007; Blass, 1993).

In line with this reasoning, in recent decades, several studies have focused on the influence of personal dispositions on behaviour within Milgram-like obedience experiments. These studies primarily focus on common personality traits, that is, traits

[47] This quote from Smeulers (2012) was translated from Dutch by the authors.

that are widespread amongst the population and are not necessarily pathological, but rather which in interaction with the systemic context of international crimes can still lead to a higher risk of involvement in crimes (Smeulers et al., 2019). In other word, these are dispositions which are not necessarily directly linked to violent behaviour itself, but rather make the transformation into a perpetrator more likely in a given context.

Traits that have been related to obedience outcomes in experimental studies are, for example: authoritarianism (*F*-scale) (Elms & Milgram, 1966), general personality traits (Schurz, 1985), empathic concern and desire for control (Burger, 2009), the Big Five personality traits (Bocchiaro & Zimbardo, 2010), and conscientiousness and agreeableness (Bègue et al., 2015). Although in some cases (weak) associations have been found between particular traits and obedience behaviour (e.g., Bègue et al., 2015), most studies have not found convincing evidence for such relationships. Overall, they have produced contradictory and mixed results (Blass, 1991). The current consensus is that there is no clear and univocal empirical evidence concerning which personal characteristics distinguish those who obey from those who do not, within an experimental setting. Therefore, it also remains unclear which characteristics are important candidates for explaining the difference between those who perpetrate and those who resist under the same systemic circumstances in the context of international crimes (Blass, 1991; Bocchiaro & Zimbardo, 2010; de Swaan, 2015; Smeulers et al., 2019).

Besides these experimental studies, scholars have also argued the case for certain relevant individual dispositions on theoretical grounds. De Swaan (2015), for instance, argues that situations of mass atrocities are characterized by a process of compartmentalization, both on a societal and psychological level. This entails that individuals strongly bond with people from the same group, while, simultaneously, strongly de-identifying with (groups of) people who they regard as different to them. The main personal disposition put forward by de Swaan (2015) to explain whether people will yield to the pressures of adopting the role of either perpetrator or resister in such a context, is the (in-)ability to have empathy with those outside of one's direct circle of identification. Whilst this supposition by de Swaan is of course a possibility, it also remains largely hypothetical, insofar as it is not substantiated with empirical evidence on both groups of perpetrators versus those who do not go along. Rather, de Swaan (2015) appears to infer this from the fact that perpetrators use violence against persons outside of their own group and very rarely display sincere remorse afterwards. However, both could also be the result of other situational and individual processes of course, such as a susceptibility to authority or conformity and a tendency to morally disengage and internalize alternative (moral) beliefs.[48]

Staub (2014) stresses that certain individual moral dispositions are crucial for resisting situational pressures. He views both Milgram-like situations and mass atrocities from a systems perspective, where situational influences and a process of learning by doing often 'overwhelm moral concerns' (p. 509). A "prosocial value orientation" may in this regard be highly relevant, insofar as it has to do with concern for others and feeling responsible for other's wellbeing. It is emphasized, however, that what general values become important in a certain situation is highly contextual (Staub,

[48] It even appears somewhat circular to derive the apparent explanatory difference of empathic ability between perpetrators and 'ordinary' people (de Swaan, 2015, p. 230) from what needs to be explained, namely the more violent behaviour and the lack of remorse of perpetrators.

2014; see van den Berg et al. (2022) for a similar point). One important aspect, here, is that such moral concerns are triggered by the situation (think of the vocalized distress of the learner in the Milgram experiment). Another potentially relevant moral individual disposition referred to by Staub (2014) is people's level of moral courage.

In an interesting recent study, Munch-Jurisic (2018) adopts a more sceptical stance towards simply regarding certain moral dispositions as protective against perpetration, and instead emphasizes the complex role that these can play within the specific context of international crimes. In her study on perpetrators' disgust, which is the disgust perpetrators can experience when witnessing or committing atrocities, she shows that the moral meaning of this emotion is often not univocal. In fact, in cases where disgust invokes moral conflict, it is most often interpreted as a weakness that must be overcome, which, in turn, induces further and more extreme harm doing: 'Their ability to manage these emotions becomes a crucial element of their self-understanding and their genocidal mentality.' (Munch-Jurisic, 2018, p. 161). What becomes clear here is that what matters for behaviour is not the moral disposition as such but rather the meaning it receives when the person relates to it in the reflexive relationship towards oneself.

Interactionism

Whilst it may not be clear which personal dispositions are most relevant to explaining behaviour in the context of mass atrocities, an integrative approach that incorporates the interaction between dispositional and situational factors is generally regarded as the best way forward (Harrendorf, 2014; de Swaan, 2015, Smeulers, 2019; Newman, 2002; Blass, 1991). However, Newman (2002), amongst others, stresses that a genuine *interactionist* approach must go one step further than merely combining 'static' individual and situational factors to both explain and predict behaviour within the context of international crimes. Following the interactionist tradition in social psychology, he claims that both factors are entangled to such a degree that to speak of them only as separate, albeit combined, effects is insufficient.

Newman (2020) purports that an interactionist approach defends four key principles. First, interactionists point out that neither dispositions nor situations in and of themselves determine behaviour. Rather, both do, and, crucially, they do so in interaction with each other. The influence of situations differs amongst people with different traits, whilst the effect of people's traits may only become effective in certain situations and not in others. Second, interactionism does not understand situations as static influences to which individuals are passively submitted, but rather regards them as, albeit to a certain extent, being within the control of social actors themselves: people can select situations. Third, situations do not only affect behaviour but also people. More specifically, situations can change people's characteristics. A person's self-concept must thus be understood in a more dynamic way. This relates directly to both the aforementioned transformation process undergone by perpetrators and the influence of cognitive dissonance and rationalizations by which people change their beliefs and, ultimately, their moral identity. Finally, also here, the relationship is reciprocal since people and their specific characteristics also affect situations. Newman (2002, 2020) argues that if personality influences the situations one encounters, and the situation influences our behaviour -or when the situation determines your personal dispositions which subsequently influence behaviour- then it becomes impossible to attribute

influence to one or the other factor, even if they are considered in combination with one another.

Haslam and Reicher (2007) apply a similar interactionist perspective whilst discussing three dynamics at work in the rise of tyrannical violence. To start, they point out that people with certain characteristics (e.g., supporting hierarchy) are drawn to tyrannical groups. They maintain that although individuality obviously matters, it should nevertheless be understood 'as the sum of social relations, both past and present, that determine how people exercise choices about the future that are available to them' (Haslam and Reicher 2007, p. 620). Second, Haslam and Reicher (p.620) argue that tyrannical groups change their members, both in terms of their individual traits and the way in which these characteristics express themselves. More particularly, with reference to Turner et al.'s (1987) self-categorization theory, they state that 'what people learn about their groups as members changes the way they understand themselves'. Finally, when the social context changes, formerly uninfluential individuals can suddenly become leaders, thus giving them the ability to define the moral norms and further transform the social context to gain even greater influence. This, in turn, transforms what is regarded as normal in society and what it means to be a law-abiding citizen.

It should be clear from this dynamic interactionist explanation that the transformation of the individual is an essential constitutive element, whereby individuality takes a more central role than in situational approaches, but still always occurs in a dynamic interaction with the social context, and is reflexive. Haslam and Reicher stress that this transformation process requires creativity and hard work, and that for many perpetrators this resulted in a situation that 'they *really believed* that what they were doing was right' (Haslam and Reicher (2007, p. 619) (emphasis in original). The critical question at this juncture, Haslam and Reicher (2007) point out, is no longer why people follow the leader but rather why some others do not.

5.2.4 The relevance of moral identity and its dynamic and reflexive aspects for explaining international crimes

What emerges from this –albeit non-exhaustive- overview is that first and foremost, both explanations that primarily rely on situational factors and those which include personal dispositions refer to the moral dimension of the human being. This dimension is expressed through the personal conviction on the behalf of the perpetrator that committing international crimes is "the right thing to do". This points towards the relevance of a person's individual morality, or better yet still, moral identity, for understanding perpetrators of international crimes. Secondly, the empirical approaches refer to the moral transformation of the individual, in order to explain how people (do not) go along with the systemic context of international crimes. This moral transformation can be said to exist by virtue of the fact that what the perpetrator once regarded as morally wrong (killing and using other forms of extreme violence against helpless victims) is now deemed to be "right". Hence, the transformation takes place at the level of one's moral identity, that is, the interpretation of one's own values, upon the basis of which one then acts. Third, the moral identity that is presupposed is reflexive, insofar as the transformation process entails a reinterpretation of oneself and results in a different understanding of oneself.

Although it is evident that empirical research on international crimes presupposes such a concept of moral identity or moral self, it is not further explicated.

Rather, in these explanations the moral transformation and adoption of the perverted moral code is typically presented as the mere outcome of the different empirical factors and psychological mechanisms discussed above. The transformation of moral identity itself is not subjected to any further scrutiny, as if all that can be said about it is exhausted by reference to the empirical factors and mechanisms. This is wholly unjustified, as we will make clear by delving deeper into the three aforementioned points.

First of all, in the empirical approaches, the reference to people's moral identity functions as an important element for both explaining and understanding behaviour in the context of international crimes. Specifically, it makes the behaviour intelligible, insofar as it explicates an important motivation for committing these horrific acts, namely "doing the right thing". As such, this moral dimension warrants separate attention when seeking to understand the perpetrators of international crimes.

The question of people's moral identity becomes even more inescapable when considering the act of resistance. Because of the "fact of resistance", scholars have rightly pointed to the influence of dispositional factors for explaining perpetration and resistance. Yet, it appears that it is simply not possible to understand resistance without due reference to the morality of the resisting individual. For instance, while a high level of authoritarianism or agreeableness could potentially help explain perpetration in the sense of going along, and –on the other side of the same coin- a lack of such dispositions could help explain not going along, ultimately, the latter does not work without also referring to the idea that not going along is also the right thing to do. From a more positive perspective, dispositions like empathy, in the sense of feeling or understanding the feelings of others, also appear insufficient for explaining resistance if it does not also entail the further determination that the act is "not right". This is supported by the fact that feelings of empathy can still be *rationalized* away, thus making way for the judgment that the killing is somehow justified. Munch-Jurisic's study (2018) of perpetrators' disgust provides clear support for this train of thought.

The fact that an individual's morality is important for both the explanation and understanding of resistance (and by implication for that of perpetration) is also explicitly acknowledged by Bauman (1989). When considering that there must be more to morality than the standard sociological conception, inspired by Durkheim, according to which an individual's morality is determined by the ruling moral norms, Bauman (1989, p. 168) states: 'some ordinary people, normally law-abiding, unassuming, non-rebellious and unadventurous, stood up to those in power and, oblivious to the consequences, gave priority to their own conscience…'. Both the perpetrator and resister perspective therefore give ample reason to further investigate the moral dimension of the human being in relation to behaviour within the context of international crimes, and particularly the concept of moral identity that underlies it.

This brings us to the second point. The concept of moral identity that arises from the discussed empirical research is dynamic and capable of change, while an individual remains one and the same person. This emerges when considering the moral transformation of perpetrators. Furthermore, people who resist can also be said to go through a process of moral development, in that they no longer simply follow social norms, the law, and state authorities, but rather relate to them from the standpoint of their "own conscience". This raises the question of how such a dynamic conception of moral identity is possible and what it looks like, i.e., questions about its structure.

The third point already brings an essential aspect of this structure to the fore, namely: the dynamic moral identity that is assumed by the empirical approaches appears to be reflexive. The transformation process is described by these theories as an essentially reflexive process of self upon self. The individual transforms as it reinterprets itself under the influence of psychological mechanisms, such as cognitive dissonance and neutralization techniques, when it is subjected to certain situational influences and is susceptible to these processes and influences as a result of certain personal characteristics. The process of moral transformation as it occurs in the rationalisation of behaviour, then, entails that a subject reinterprets its moral identity, e.g., its values, to whom its values apply, and its moral behaviour, while, simultaneously, continuing to understand themselves as a "good person". This moral self-understanding can be summed up by the conviction that "I can commit these acts".

This dynamic and reflexive concept of moral identity remains largely implicit within the discussed empirical approaches. In other words, although this concept of moral identity or moral self is assumed by these approaches, both the question of how such a moral identity is possible and an investigation of its reflexive structure is lacking. Given that this question inquires after a priori assumptions about the subject, it is to be expected that empirics may not have much to say about it. However, as it pertains to a fundamental view on the moral subject, it can direct further empirical theorizing and investigations. Specifically, gaining a better understanding of this dynamic and reflexive moral identity could turn out to be integral for understanding how the moral transformation of "ordinary men" into violent perpetrators becomes possible and, moreover, may yield valuable insight into the possibilities for resistance.

It is for this reason that we turn to Ricoeur's theory of the narrative self. Ricoeur (1992) describes a moral self that is reflexive, in the sense that it understands itself through the (morally relevant) stories it tells about itself and that can incorporate change while remaining one and the same person. The stance of simple interactionism of strictly distinguishing situational factors from personal dispositional factors becomes less useful in this idea of the subject, insofar as both are entangled in the subject's dynamic interaction with its environment. Given this, we argue that this concept of self can function as a basis for the described more dynamic interactionist explanations of perpetrators and resisters of international crimes.

5.3 Ricoeur's concept of the narrative moral self

5.3.1 *Ipse*- and *idem*-identity

In his work *Oneself as Another*, Ricoeur (1992) tackles the problem of personal identity as the fundament for action and ethics. He addresses the ancient problem of how it is possible that we understand persons (and things) as remaining identical and persisting across time, while everything is constantly changing, by distinguishing between two possible uses of the word identity: identity as *sameness*, designated with the Latin *idem*; and identity as *selfhood*, designated with *ipse*.

Idem-identity or sameness, essentially a concept of relation, is the concept that we use to deal with and understand the permanence in time of the objective or empirical world, i.e., the world of things. *Idem*-identity encompasses the components by which we are able to identify and recognize something or someone as the same. The first of these components is numerical identity, which refers to the fact that we understand two

occurrences of a thing or individual as being the same one (e.g., someone walks out a room and comes back in again). Second, qualitative identity refers to sameness in the sense of 'extreme resemblance' (e.g., "your cup looks exactly the same as mine") (Ricoeur, 1992, p. 166). As Ricoeur (1992) makes clear, qualitative identity often serves as a criterion for numerical identity: by relating two occurrences in terms of qualitative sameness, we can become convinced that it is one and the same thing. Moreover, because one individual (thing) changes across time, we also make use of a third component of identity, namely 'uninterrupted continuity,' which can be described in terms of a seed developing into a flower or a girl becoming a woman. Demonstrating the uninterrupted continuity of the same individual or thing 'rests upon the ordered series of small changes which, taken one by one, threaten resemblance without destroying it' (Ricoeur, 1992, p. 117) (e.g., think of flipping through one's photo album or Facebook timeline). Still, given that time functions as a factor of change and difference, the concept of identity must assume a "principle of permanence in time" as its strongest criterion, in order to be able to conceive of change while a thing remains the same. This permanence in time thereby functions as the condition of possibility for numerical identity. This is, for instance, the unchanging structure or organization of a thing, that which remains the same while its components change. With regards to the biological individual, one could think of someone's genetic code as a principle of permanence in a flux of change (Ricoeur, 1992).

On the other hand, there is identity as selfhood or *ipse*. Selfhood, as Ricoeur (1992) emphatically states, is not sameness. At its very core, selfhood denotes a permanence in time which is the reflexive relationship of the self towards the self. In this respect, *ipse* stands for the fundamental experience that we have of ourselves as subjects. This experience is well expressed in the primary experience of *our* actions, *our* body, and *our* thoughts as belonging to *ourselves*. This is in contradistinction to an impersonal description from a third point of view, where, for instance, a body can be described as one amongst others. For Ricoeur (1992), the non-reducibility of this reflexive experience in which actions, bodies, and thoughts, are ascribed to ourselves, attests to the non-reducibility and presupposition of the self as a fundamental aspect of the person. This is an aspect that always comes into view when we ask the indelible question "who?," especially when concerning action, in the sense of "who is acting?." While both persons and objects carry *idem*-identity, only persons are characterized by *ipse*-identity or selfhood, as well. By bringing in *ipse* as a fundamental part of identity the analysis thus becomes phenomenological.

Apropos *personal identity*, selfhood and sameness are engaged in a dialectical relationship. Exactly what this relationship entails becomes clear when Ricoeur turns to the question of a form of permanence in time that could constitute our personal identity, i.e., an answer to the question "Who am I?". Ricoeur (1992) states that there are two forms of permanence in time that could serve as a reply here: "character" and "faithfulness to oneself in keeping one's word," which expresses self-constancy and self-maintenance as the effort of self-constancy in the act of promising. Both replies stand for a persistence in time that belongs to the individual person and represent opposite poles of personal identity. At the pole of character, *ipse* and *idem* overlap, i.e., character is the expression of *ipse* in terms of *idem*; the "who" or self is explained in terms of distinctive and recognizable features. At the pole of self-constancy in promising, *ipse*

stands on its own 'baring in a sense the selfhood of the self, severed from its base in sameness' (Ricoeur, 1992, p. 123).

5.3.2 Self-constancy and character

We will now further explore both poles of personal identity –character and self-constancy-, beginning with the latter, where *ipse* stands on its own. How can we describe the pole of self-constancy, baring pure selfhood, when it lacks objectively identifiable characteristics? As aforesaid, here, we enter the phenomenological realm, attempting to grasp an essential aspect of the subject from the first-person perspective that is not easily describable in objective terms. Therefore, Ricoeur notes, these insights aspire to a different sort of certainty and verification than that which is attached to the objective verification of (empirical scientific) descriptions, in terms of truth. Conversely, the only type of certainty and verification phenomenology can claim is that of *attestation*. This is not *doxa* in the sense of mere personal belief, but rather is closely linked to testimony and its attendant features of trust and credence. In a sense, the structure of our self as a subject is something we can only *attest to*.

Ricoeur (1992) follows this path by bringing forward the structure of the self that is implied in the act of promising and which he regards as emblematic of selfhood. Promising assumes a self that consists of self-constancy and self-maintenance -a work of self upon self-, expressing reflexivity and temporality. This self-constancy expressed in promising denotes a 'who' –a permanence in time or identity- that does not need, and in fact is completely detached and independent from, the recognizable and objective features of sameness expressed in character. In other words, it is not the simple continuation of character or preferences that one counts upon when a promise is made, but rather the persistence of self-constancy despite one's characteristics and the potential changes therein or changes in the situation. Indeed, keeping one's word appears 'as a challenge to time, a denial of change: even if my desire were to change, even if I were to change my opinion or inclination, "I will hold firm"' (Ricoeur, 1992, p. 124). It is precisely this aspect of identity that is put in position when one exchanges wedding vows, and which, in turn, the one to whom the vow is made counts upon. As Ricoeur (1992, p. 165) makes clear, selfhood thereby has a fundamental ethical dimension:

> Self-constancy is for each person that manner of conducting himself or herself so that others can *count on* that person. Because someone is counting on me, I am *accountable for* my actions before another, the term "responsibility" unites both meanings: "counting on" and "being accountable for." It unites them, adding to them the idea of a *response* to the question "Where are you?" asked by another who needs me. This response is the following: "Here I am!" a response that is a statement of self-constancy.

How then should we understand the opposite pole of identity, that is, the other answer to the question "Who am I?," where *idem* and *ipse* overlap, which is character? Ricoeur (1992, p. 121) describes character as 'the set of lasting dispositions by which a person is recognized' and by which he or she can be re-identified as the same. This already indicates the strong *idem* aspect of this pole of identity. On the one hand, these lasting dispositions are someone's *habits* that are formed and acquired across time through a process that Ricoeur compares to sedimentation. Importantly, the formation

of habits also implies a moment of *innovation*, which is covered up by the movement of sedimentation.

On the other hand, people's lasting dispositions are a result of *acquired identifications* with what is other than oneself, which is then internalized as one's own. According to Ricoeur (1992, p. 121), the identity of a person is largely 'made up of these identifications with values, norms, ideals, models, and heroes *in* which the person…recognizes itself' (emphasis in original). The processes of innovation and sedimentation and that of the internalization of that which is other to oneself, gives character a temporal dimension, i.e., *lasting* dispositions, which have a history. This temporal dimension is at once signified by stability, making it possible to recognize and re-identify someone as one and the same person, i.e., *idem*-identity, and, at the same time, by a certain movement and mutability. Given that it is always possible to acquire something that was not one's own before, i.e., something new, there is always the possibility for transformation. That is to say, someone's traits can change and obtain a different meaning and effect in different situations.

This dynamic aspect of personal identity, which entails that a person may change while remaining one and the same, is a product of the dialectical relationship between *ipse* and *idem*. A person's identity is not fully reducible to one's character traits: 'my character is me, myself, *ipse*' (Ricoeur, 1992, 121). I am not simply my character; rather, I *have* a character. Selfhood, although in a sense overlapping with sameness, thus stands its ground as an independent factor of identity, conveying upon character the fact that it belongs to a self. This constitutes a personal identity that relates to its dispositions through self-constancy, in turn, making change and varying interpretations and applications of these dispositions in particular situations possible, while continuing to recognize oneself as one and the same. At the same time, through *idem,* character provides the "who" –i.e., bare selfhood- with the flesh on the bones: 'character is truly the "what" of the "who".' (Ricoeur, 1992, p. 122). It expresses the "who" in terms of "what," without becoming completely indistinguishable from each other.

The relationship between *ipse* and *idem,* which constitutes a stable but dynamic concept of character, is well expressed in the act of "recognizing oneself in" certain values and heroes. It signifies a reflexive relation that needs to be re-established again and again in particular situations, implying room for change, new interpretations and, at some point, even the possibility of non-recognition. As Ricoeur (1992, p. 121) makes clear, assuming otherness (e.g., values) as our own implies a relation of loyalty (or self-constancy) through an effort of maintaining the self that is constitutive of our character. Our character is thereby not given, but –as a result of the dialectic between *ipse* and *idem*- subject to effort and liable to (gradual) change. This also has profound ethical implications, for if the other counts on me, then it is not simply my character traits that one counts on, but rather my loyalty towards them, my self-maintenance, so that I can respond to the other's call in a way that does justice to who I am as well as to the particular facts of the situation.

5.3.3 Narrative identity

Now the question can be asked of how someone's identity, with its stable and dynamic aspects, can be constituted as a result of the relation between the two described poles of identity -character and self-constancy. Here, Ricoeur (1992) introduces the notion of narrative identity, which mediates between the two poles and, ultimately, expresses

one's personal identity. Simply put, narrative identity is the interpretation of the self in terms of a narrative which is constitutive of the self. Ricoeur equates the person to the character in a story and gives central place to the structuring function of the plot. It is the plot that expresses the same dynamic relation of unity and mutability that is characteristic of the dialectic relationship between *idem* and *ipse* within personal identity. Hence, narrative identity and its structure of emplotment express the interplay between *idem* and *ipse* with regards to the identity of the character in a story as well as the identity of persons.

Emplotment is the 'synthesis of the heterogeneous' (Ricoeur, 1992, p. 141). The dynamic identity of the story told is constituted as the plot 'configures' 'discordant' events and facts into the unity of a 'concordant' whole (Ricoeur, 1992, p. 141-142). Through the plot, manifold contingent events are brought together within a singular temporal unity and a logical sequence, so that they appear necessary or at the very least probable within the narrative. A new event is therefore at the same time a 'source of discordance inasmuch as it springs up' –a source of disruption and change- and 'a source of concordance in as much as it allows the story to advance' as it is incorporated and integrated by the plot as part of the story (Ricoeur, 1992, p. 142).

Ricoeur argues that this 'unstable structure' of emplotment that signifies 'identity and diversity', unity and constant change, is transferred to the character of the narrative (Ricoeur, 1992, p. 142-143). The actions described in a narrative involve the character as both agent and patient, actor, and sufferer. The narrative thus can be said to define the identity of the character, insofar as it connects the "what", "who" and "why" questions of an action and ascribes them to a character in the story. As the character performs the narrated actions, the narrative provides the character with an identity through its plot. Consequently, story development is character development and vice versa (Ricoeur, 1992, p. 143-144). In this sense, the narrative *is* the character and 'characters... are themselves plots' (Ricoeur, 1992, p. 143), they are subjected to the same configuration between the whole and its parts, integrating identity and diversity.

Our personal identity is constituted in precisely the same way, Ricoeur (1992) argues. We understand ourselves as characters in the narrative that is our life. Configured through the plot, this narrative recounts the actions we performed as actors and sufferers as a relatively coherent and temporal whole, in turn, conveying upon us – its character- the discordant concordance or dynamic identity characteristic of the plot. Through this coherent narrative and temporal totality, we come to understand both our life as a singular unity distinct from others and ourselves as a permanence in time –our identity. At the same time, this concordance harbours a discordant diversity and is constantly threatened by a variety of new encounters, which need to be appropriated into our life story, further developing this construction of identity. In this way, our narrative identity constitutes a permanence in time, while, simultaneously, being subjected to constant change.

The dialectic between *ipse* and *idem* can be recognised in this dynamic of narrative identity, insofar as the narrative mediates between the sameness of character and self-constancy (Ricoeur, 1992, p. 141). On the one hand, by telling one's life narrative, a person receives their recognizable features, which typify the person and provide the means via which one can be re-identified. The narrative expresses a person's character, by virtue of the fact that recounting one's actions in a coherent manner also expresses one's habits and values. The temporal aspect of character, provided by sedimentation,

innovation, and identification with others, makes this narrative dimension of character comprehensible (Ricoeur, 1992, p. 122). On the other hand, narrative identity expresses self-constancy via the appropriation of different actions and events into a coherent whole. This coherence is subjected to constant change and the need to relate to variability yet maintains one relatively coherent story that belongs to someone[49].

This narrative concept of identity helps us to explain how one can at once recognize oneself and be recognized by others as being the same person with recognizable features, while, simultaneously, these can change and lead to a variability of behaviour in different particular situations. In this respect, it can be said to bring forward the aforementioned dynamic relationship we have with our lasting dispositions, which is nicely captured by Ricoeur (1992, p. 122) when he states that, 'What sedimentation has contracted, narration can redeploy.' and '...that character must be set back within the movement of narration...'. This points to the important fact that there is not just one narrative of someone's life and thereby not one characterization of someone: 'Along the known path of my life, I can trace out a number of itineraries, weave several plots' (Ricoeur, 1992, p. 161). Indeed, as Halsema (2019, p. 129) considers in an analysis of Ricoeur's concept of narrative identity: 'The self-narrative is not necessarily fully expressed and is not a singular life story...instead [it] should be seen as a lens, a perspective, that contains the story of others as well, for instance, the stories of my parents about my birth'. Your narrative can thus be told both more than once and in different ways, as you can relate to certain episodes instead of others, expressing different aspects of who you are and, in its wake, affecting how you interpret and act upon the things around you within a particular situation.

5.3.4 Two implications following from Ricoeur's concept of narrative identity

The latter points to the first important implication of this concept of narrative identity for our study, namely that it leads to a specific idea of decision-making and action. As one *appropriates* new events as part of one's life story, the interpretation of the new event -including one's actions- and the interpretation of who you are, are brought into a certain congruence with each another. On the one hand, this means that a person's self-narrative, imbued with certain norms, values, ideals, etc., forms a perspective through which one looks at and interprets the world. On the other, it suggests that a person may call into question such values, in the sense that one reinterprets what they mean in a particular decision-making situation, thereby developing and potentially

[49] Ricoeur (1992, p. 148-149) considers that the relation between *ipse* and *idem* can vary in one's narrative identity through recourse to examples from narrative fiction. In fairy tales, characters are recognizable throughout due to their definite characteristics and their performance of predictable actions. At this extreme, the sameness of character fully overlaps self-constancy. In classic fiction and novels of apprenticeship, there is greater room for mutability and change in characters, while identification in terms of the same does not fully disappear. This corresponds mostly with actual reality. At the other extreme, characters cease to have any recognisable features, like in Robert Musil's novel *The Man without Qualities*. Selfhood totally stands on its own here in that it has lost its support of sameness. Still, personal identity does not cease completely, as the person can still ask: "who am I?", thus attesting to the non-reducibility of *ipse*. In this kind of fiction, the disappearance of the sameness of the character corresponds to the loss of any coherent plot, turning the narrative into an essay. This demonstrates the correlation between character and plot as well as the underlying dialectic between sameness and selfhood that is mediated by the narrative.

changing their interpretation, and by re-appropriating them as part of the self, changes the self. Decision-making, here, involves relating to the particular situation through one's life story, attuning to the social context, and reinterpreting oneself in this context in order to figure out what one should do. Here again, the dynamic idea of character with regards to its influence on behaviour is emphasized. Instead of ossified personal dispositions determining what one should do, one relates to one's dispositions in the situation, connecting it to (certain versions of and certain episodes of) one's self-narrative. Self-constancy is expressed here, not by always reacting in exactly the same way or stubbornly holding on to certain behaviours or values, but rather precisely by recognizing oneself in the act of making a decision in a specific situation as someone who takes account of the values and norms that one finds important, i.e., one's *own* values.

Ricoeur (1992, p. 157-158) elaborates on this dynamic between self-narrative and action at the level of making 'life plans'. Life plans are 'mobile' and 'changeable' and are shaped by the 'back-and-forth movement between' 'the vague and mobile horizon of ideals and projects in light of which a human life apprehends itself in its oneness' 'and the weighing of advantages and disadvantages of the choice…on the level of practices'. Interestingly, as Ricoeur (1992, p. 158) notes, this dynamic resembles the mutual exchange between the whole and its parts one encounters when interpreting a text. Here, the part is understood in light of the whole, whilst the former contributes, advances, and changes the latter. This, in turn, affirms the narrative dimension of action and identity.

This leads to a second important implication following from Ricoeur's concept of narrative identity that is relevant for our study, namely that it is through the narrative that action and identity are given ethical meaning. Narratives are never ethically neutral, they are always evaluative: 'in the exchange of experiences which the narrative performs, actions are always subject to approval and disapproval and agents to praise or blame.' (Ricoeur, 1992, p. 164). The plot does not only structure a narrative logically, but it also constructs it in such a way that it reflects certain reasons, goals, and values. This also means that our self-narrative is imbued with such evaluative phenomena, such as values, norms, and judgments, which testifies to the close connection between narrative identity and moral identity. Indeed, as Halsema (2019) shows, for Ricoeur ethical identity implies the capacity to give a narrative account of one's actions and give reasons for them. Similarly, when one is held accountable for one's actions, it is in the form of a narrative, expressing evaluations, that one is expected to respond.

This ethical dimension of the self-narrative was of course already implied in the identification with values and heroes as part of our character, as well as when referring to the vague ideals by which we come to see our life as a unity and to which we fit our life plans in mutual interaction with daily practices. Ultimately, Ricoeur makes clear, our self-narrative expresses our idea of what constitutes "the good life". As narrative identity is not only characterized by the dialectic between selfhood and sameness, but also by the dialectic between the self and the *other*, this idea of the good life is not (merely) determined by a concern for ourself qua self-interest, but rather also takes into account the good treatment of and concern for others.

Here, Ricoeur also underscores the ethical dimension of *ipse*, insofar as my responsibility involves me being accountable to others who count on me (Ricoeur, 1992, p. 165). This is the ethical meaning of self-constancy, keeping one's promises to others,

regardless of changing circumstances: "'I will hold firm'" (Ricoeur, 1992, p. 124). Furthermore, it is ultimately the appeal by the other that constitutes us as a person and gives us a starting point or sense of direction when we have lost track of who we are, in the sense of an identity crisis, which is a loss of *idem*. In responding to the other, who counts on you, one turns the question "who am I?" –a question that never disappears- into the answer: "'I can try anything," to be sure, but "Here is where I stand'" (Ricoeur, 1992, p. 168). To conclude, our narrative identity, which consists of performed actions, narrated in a certain way while taking account of others, is also our *moral* identity or *moral* self.

5.4 The narrative moral self as a fundament to empirical approaches to perpetrators and resisters of international crimes

5.4.1 The reflexive and narrative structure of the moral transformation process

It is time to retrace our steps and take stock of what has been discussed thus far. In section 5.2, through an overview and discussion of the three main explanatory approaches –situational, dispositional, and interactionist- we demonstrated, first and foremost, the relevance of people's moral identity for explaining why people (do not) go along with the systemic context of international crimes, which is expressed in the personal conviction of "doing the right thing". Second, we claimed that the moral transformation, which emerges in the empirical approaches, is essentially a transformation of one's moral identity. Extant explanations of both perpetrators and resisters imply a moral identity that is dynamic and capable of change, while, simultaneously, remaining one and the same person. Third, we argued that this moral identity has an important reflexive component, insofar as the moral transformation is explained by psychological mechanisms that imply a certain relation of self towards oneself, thus leading to an altered moral self-understanding. The questions at the end of that section were, firstly, how is such a dynamic and reflexive moral identity, which arises from and underlying psychological and criminological theory, possible and, secondly, how can gaining a better grasp of this conception enhance our understanding of behaviour within the context of international crimes?

To answer the first question, we turned to Ricoeur's concept of the narrative self. This concept of moral identity makes the possibility of change in interaction with one's environment, while remaining one and the same person, intelligible. The reflexive structure of self-understanding plays a pivotal role here. The narrative understanding of the self implies a dynamic relationship with our moral traits, such as moral values and norms, which feature within our self-narrative. Furthermore, the configuring structure of emplotment, a key characteristic of narrative identity, suggests that moral decision-making involves a reinterpretation of the self within a given situation, rather than a straightforward application of norms and values. That is to say, one attunes to the social context and comes to terms with the moral values that one considers to be one's own. This recognition of the self within a particular context harbours both flexibility and constancy.

This brings us to the second question: how can this concept of narrative identity lead to a better understanding of perpetrators and resisters in the context of international crimes? Ricoeur's concept of narrative identity helps us grasp the notion

of moral identity that is assumed by the empirical approaches, but is not sufficiently explicated. This concept of moral identity underlies explanations that emphasize the extraordinary context of international crimes, where perpetrators transform as a result of internalizing the criminal ruling norms, whilst resisters manage to take a stand against them. Through recourse to Ricoeur's notions of *ipse* and *idem* and the mediation of narrativity, it becomes clear that the transformation of perpetrators indeed entails a transformation of a person's moral identity, that is, one's moral self-understanding. It also becomes evident how this is possible while continuing to maintain the idea of oneself as the one and same good person.

The transformation of a person's moral self-understanding is possible through a gradual process in which both the meaning of one's actions and the story of who one is are mutually fitted, thus leading to a narrative of the self in which these actions are acceptable and make sense. This process of mutual fitting is both facilitated and smoothened by ideologies and rationalizations, which are themselves narratives about others and about oneself. They give meaning to one's actions in ways that help appropriate them into one's self-concept and which at the same time changes this self-concept. In this way, rationalizing behaviour is then ultimately –or even primarily- a rationalizing of oneself. This involves, first and foremost, constructing and adopting a new narrative or understanding of the self that makes this behaviour morally acceptable or obligatory to pursue. When we say that the torturer "learns by doing," then what this actually means is the appropriation of the narratives of his acts into his self-narrative, which, simultaneously, constitutes a change in his understanding of the acts as well as his understanding of himself.

The configuration of the plot can be regarded as a further articulation of the self-constancy that is implied in cognitive dissonance theory. It is the relative flexibility and, at the same time, convergent force of emplotment that makes it possible to integrate apparently opposing moral actions, which express opposing moral values, into one understanding of the self (i.e., think of being a good father or taking care of strangers in need, whilst torturing prisoners in a death camp). This congruent understanding of the self entails that, overall, one is one and the same good person doing the right thing. The narrative of one's actions (e.g., "fulfilling my duty" or "protecting the state from enemies") help to satisfy certain abstract ideals as well as an overall vague, and often non-articulated, idea of the good that is considered as part of the self. Yet, at the same time, what it practically means to be a "good person" and "to do the right thing," as expressed in the narrated actions and appropriated as one's moral identity, has dramatically changed.

The dynamic between the whole of the self-narrative and the integration of narrated actions, then, expresses the dynamic relationship we have with our moral norms and values. The transformation of a person's moral identity can be regarded as a change in his relationship with or interpretation of his moral norms and values. As the self-narrative transforms, one does not fully abandon certain morally relevant values and beliefs, but rather relates to them differently and gives a different meaning to them in certain social situations. So, one's own dispositions are reinterpreted in light of a new social context and subsequently reintegrated into one's life story. The ordinary person who has turned into a torturer has not simply abandoned the moral norm of "not harming other human beings," which contains the moral value of "care", but rather has reinterpreted himself in his role of torturer as someone who either does not violate any

such moral norm or who adheres to it because the torturing of "enemies of the state" contributes to the protection of others. The "I can" is then an understanding of oneself as someone who is capable of committing such crimes, based on one's self-narrative which is imbued with one's values and norms.

5.4.2 The narrative moral self as a foundation and explication of the interactionist approach

As one may have discerned, the concept of narrative identity, as explicated by Ricoeur, underlies an interactionist approach to explaining behaviour. This becomes clear when we relate the key principles of interactionism considered by Newman (2002) to the aforementioned decision-making process that is suggested by the movement of emplotment in narrative identity. The interactionist idea that traits influence the meaning of situations and that not all situations trigger the same traits, is reflected in the mutual fitting within narrative identity of the self-narrative and new events. It entails that one gives meaning to a situation from the viewpoint of one's self-concept, and, moreover, that one attunes to the social context when reinterpreting the self in a given situation, which also means that in certain situations only certain aspects of the narrated self may come to the fore. Furthermore, narrative identity also posits that we select situations that fit our self-narrative and which further develop it. Finally, the interactionist idea that situations change people's characteristics is a core aspect of the constantly changing narrative self, due to encountering and incorporating new situations, that we have described above.

Besides providing a fundament to understanding behaviour in an interactionist manner, the presented narrative concept of identity also deepens this understanding by explicating the narrative structure of the interaction. The notion of emplotment specifies what is already suggested by interactionism, namely how the person and their surroundings interact in such a way that transcends the idea of behaviour being the simple product of general personal and situational factors. Furthermore, by making *self-understanding* the focal point of the intertwinement between individuality and situations, we suggest that it is also here that we should look for the possibility for individuals resisting and, in turn, get a better understanding of the differences in this respect between individuals.

5.4.3 A lesson for both the situationistic and individual dispositional approaches

In connection to this contribution to the interactionist approach, the concept of narrative identity also appears to provide additional nuance to the two approaches that form both ends of the extreme in extant debates on explaining behaviour within the context of international crimes –the situational and the dispositional approach.

Beginning with the former, following Ricoeur's concept of personal identity it becomes clear that i*ndividuality* forms an inevitable part of the constitution of behaviour within the context of international crimes. Through the irreducible question "who?", as in the question "who is the performer of the action?", the individual person enters the stage. The question thus functions as an antidote to excessive situationistic perspectives, where individuality vanishes as behaviour is fully explained in terms of social and situational factors, in that it reminds us that it is still *a person* that these situational factors *act upon*. More specifically, from the situationistic perspective, the

question becomes: "Who is the subject of these situational pressures?". The answer provided by this approach is, at a minimum: "a subject who is able to come to see its acts as the right thing to do in a specific situation", which is an expression of moral identity. Through this assertion, situationism presupposes a form of individuality or moral identity that is easily subsumed by the overwhelming emphasis on situational factors, but which we are able to draw out and further articulate through Ricoeur's conception of identity. In line with Ricoeur, we can namely state that there is *someone* to whom this moral belief belongs, and there is *someone* who is undergoing transformation. In other words, situationism also cannot evade the question "who?".

This raises the question of what precisely this identity consists of? Following Ricoeur, this answer is, at the very least, a self that, through its reflexive component given in selfhood, has a moral understanding of itself. That is, a self that can have an understanding of itself in terms of what is right or wrong for it to do. It is this moral self-understanding, where the self relates to itself, that forms the point of contact for situational pressures to direct individual behaviour. At the same time, it is this moral self-understanding that can form the starting point for resistance, insofar as one understands oneself as someone who is unable, in a moral sense, to commit such acts. It thus forms the basis for the individuality that is pointed to within empirical theory as being an inevitable aspect of resistance. In other words, it is the self-understanding of the agent, given by the reflexivity of selfhood, that the *other* counts upon, for good or for bad, whether it be a criminal authority or a helpless victim.

This brings us to the lesson for those explanations that emphasize dispositions. Here we want to highlight that individual dispositions are at the pole of personal identity where besides *idem* also *ipse* plays a role, signifying that we relate to our dispositions. Accordingly, dispositions should, Ricoeur argues, not be understood as rigid or ossified traits. Rather, our character is expressed by the narratives that we tell about ourselves in which these traits figure as habits and identifications with. Of course, it is possible to, at some point, distil general characterizations -such as a certain level of empathy- from our narrative. This is what actually occurs in personality questionnaires where one self-reports about certain habitual behaviour or beliefs, deduced from narratives of action and beliefs that one has about oneself, which result in a general self-concept. What should be kept in mind though, is that this is always a reduction of what one's identity actually is, that is, a reduction of the stories in which these traits feature within a specific social context and thus of the dynamic role that one's dispositions may play in these particular decision-making situations. Every time we act, our dispositions are at stake to some extent, as it is in the recognition of oneself within a given situation that one relates to them anew, which provides the possibility for (minor) changes in interpretation and action.

This understanding of dispositions as playing a dynamic part within one's self-narrative, is well illustrated by Munch-Jurisic's (2018) study on perpetrators' disgust. Here, the personal disposition of a perpetrator's disgust has a counterintuitive effect due to how the perpetrator *relates* to the trait in his understanding of himself. In other words, it is not the trait itself that directly determines someone's behaviour, but rather the trait receives a certain meaning within the narrative that the perpetrator tells about himself. This narrative is of course always embedded in the specific social context and role the perpetrator has within it, which are themselves narratively understood. Within

this context, the disposition of a perpetrator's disgust often leads to more aggressive behaviour as opposed to protective behaviour.

5.4.4 Methodological suggestions emerging from the analysis

The above leads to an important methodological suggestion in order to gain a better understanding of why some people become perpetrators and others are able to resist. Besides the experimental work that is conducted to find individual differences in behaviour, e.g., between those who go along and those who do not, it appears insightful to investigate the narratives that these actors tell about themselves and which made their actions morally understandable for themselves. It is within these broader morally imbued understandings of oneself, situated within a specific social context, that the interactions between dispositions and situations as well as their meanings can become clearer. In this regard, Munch-Jurisic (2018) is a nice example of a study which makes use of narratives from the perspective of perpetrators –it even uses a fictional story of an SS officer- to describe the complex understandings and interpretation of a disposition, in this case the emotion of perpetrators' disgust. Through the perpetrators' narratives, Munch-Jurisic (2018) makes it clear how these understandings function within the specific context of international crimes and are constitutive of behaviour.

Another methodological avenue supported by our findings is to make use of (auto-)biographies of perpetrators and resisters to gain a better understanding of their different behaviours in this context and, ultimately, deduce some commonalities that can help us to make sense of them. A good example of such a study is Smeulers's (2008) typology of perpetrators on the basis of ego documents and (auto-)biographies. By studying the narratives of these perpetrators, who are always in interaction with their environment, Smeulers (2008) is able to pinpoint differences between types of perpetrators, which harbour the mutual interactions between person and situation, advancing our understanding of the constitution of different behaviours within the context of international crimes. One way forward that appears to be especially promising in this regard is to study the narratives of resisters. As argued, one apparent certainty about this group is that they have at some point come to the conclusion "I can't go along with this", which is an expression of moral self-understanding. In this respect, their self-narratives would provide a primary source through which to improve our understanding of why this is the case.

5.5 Conclusion

In this study we have presented a concept of moral identity that can serve as a fundament to empirical approaches explaining behaviour within the context of international crimes. Through conducting a (non-exhaustive) overview of the empirical field, we showed that situational as well as dispositional and interactionist approaches presuppose a dynamic and reflexive concept of moral identity, which is not further articulated. Through recourse to Ricoeur's narrative concept of identity, we have articulated an agent that is both capable of moral change and finds its identity in the reflexive structure of self-understanding. Our analysis emphasizes both the narrative and reflexive structure of the moral transformation process perpetrators undergo, as well as of the moral development of those who are able to take a stand against the perverse ruling norms within the context of international crimes. This supports the idea

that an interactionist approach represents the best way forward to gain a better understanding of the different behaviours within the context of international crimes. Our study shows that this interaction goes by the understanding of self in narrative terms, which makes it a central aspect of transforming into a perpetrator, as well as the possibility to resist. By transcending explanations in terms of situational and personal factors, the question of perpetration and resistance becomes essentially about the stories that a person is able to tell about themselves. To move forward in our understanding of the different behaviours within this context, it is therefore vitally important for empirical studies to focus on these self-narratives of both perpetrators and resisters.

References

Alvarez, A. (2009). *Genocidal crimes*. Routledge.

Anderson, K. (2019). The Margins of Perpetration: Role-Shifting in Genocide. In A. Smeulers, M. Weerdesteijn, & B. Hola (Eds.), *Perpetrators of international crimes: theories, methods and evidence* (pp. 132-152). Oxford University Press.

Aronson, E. (2004) *The Social Animal* (9th ed.). London: Worth.

Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgment. In M. H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177-190). Pittsburgh: Carnegie.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3, 193-209.

Bauman, Z. (1989). *Modernity and the Holocaust*. Cambridge: Polity Press.

Beauvois, J. L., Courbet, D., & Oberlé, D. (2012). The prescriptive power of the television host. A transposition of Milgram's obedience paradigm to the context of TV game show. *European Review of Applied Psychology*, *62*(3), 111-119.

Bègue, L., Beauvois, J. L., Courbet, D., Oberlé, D., Lepage, J., & Duke, A. A. (2015). Personality Predicts Obedience in a Milgram Paradigm. *Journal of Personality*, *83*(3), 299-306.

Blass, T. (1991). Understanding Behavior in the Milgram Obedience Experiment: The role of personality, situations, and their interactions. *Journal of personality and social psychology*, *3*(3), 398-413.

Blass, T. (1993). Psychological perspectives on the perpetrators of the Holocaust: The role of situational pressures, personal dispositions, and their interactions. *Holocaust and genocide studies*, *7*(1), 30-50.

Blass, T. (2009). From New Haven to Santa Clara: A historical perspective on the Milgram obedience experiments. *American Psychologist*, *64*(1), 37-45.

Bocchiaro, P., & Zimbardo, P.G. (2010). Defying unjust authority: An exploratory study. *Current psychology*, *29*(2), 155-170.

Browning, C. R. (1992). *Ordinary men: Reserve Police Battalion 101 and the final solution in Poland*. New York, NY: Aaron Asher Books.

Burger, J. M. (1999). The foot-in-the-door compliance procedure: A multiple-process analysis and review. *Personality and social psychology review*, *3*(4), 303-325.

Burger, J. M. (2009). Replicating Milgram: Would people still obey today?. *American Psychologist*, *64*(1), 1-11.

Burger, J. M. (2014). Situational Features in Milgram's Experiment that Kept his Participants Shocking. *Journal of Social Issues*, *70*(3), 489–500.

De Swaan, A. (2015). *The killing compartments*. Yale University Press.

Drumbl, M. (2007). *Atrocity, Punishment and International Law*. Cambridge: Cambridge University Press.

Einarsen, T. (2012). *The Concept of Universal Crimes in International Law*. Oslo, Norway: Torkel Opsahl Academic Epublisher.

Elms, A. C., & Milgram, S. (1966). Personality characteristics associated with obedience and defiance toward authoritative command. *Journal of Experimental Research in Personality*, *1*, 282-289.

Fein, H. (1984). Scenarios of genocide: models of genocide and critical responses. In I.W. Charny (Ed.), *Toward the understanding and prevention of genocide – proceedings of the international conference on the Holocaust and Genocide* (pp. 3-31). Boulder and London: Westview Press.

Fenigstein, A. (2015). Milgram's Shock Experiments and the Nazi Perpetrators: A Contrarian Perspective on the Role of Obedience Pressures during the Holocaust. *Theory & Psychology*, *25*(5), 581–598.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of personality and social psychology*, *4*(2), 195-202.

Halsema, A. (2019). The Self is Embodied and Discursive: Tracing the Phenomenological Background of Ricoeur's Narrative Identity. In *A Companion to Ricoeur's Fallible Man* (pp. 125-144). Lexington Books.

Haritos-Fatouros, M. (2003). *The psychological origins of institutionalized torture.* London: Routledge

Harrendorf, S. (2014). How Can Criminology Contribute to the Explanation of International Crimes. Journal of International Criminal Justice, 12, 231-252.

Haslam, S. A., & Reicher, S. (2007). Beyond the banality of evil: Three dynamics of an interactionist social psychology of tyranny. *Personality and social psychology bulletin*, *33*(5), 615-622.

Haslam, S. A., & Reicher, S. D. (2017). 50 years of "obedience to authority": From blind conformity to engaged followership. *Annual Review of Law and Social Science*, *13*(1), 59-78.

Hilberg, R. (1961). *The Destruction of the European Jews*. Chicago: Quadrangle Press.

Jetten, J., & Mols, F. (2014). 50:50 Hindsight: Appreciating anew the Contributions of Milgram's Obedience Experiments. *Journal of Social Issues*, *70*, 587-602

Kelman, H. C. (2001). Reflections on social and psychological processes of legitimization and delegitimization. In J.T. Jost and B. Major (Eds.), *The psychology of legitimacy: Emerging perspectives on ideology, justice, and intergroup relations* (pp. 54-73). Cambridge: Cambridge University Press.

Kelman, H. C. (2009). The policy context of international crimes. In: A. Nollkaemper & H. Van der Wilt (Eds.), *Systemic Criminality in International Law* (pp. 26-41). Cambridge: Cambridge University Press.

Kelman, H. C., & Hamilton, V.L. (1989). *Crimes of Obedience: Toward a social psychology of authority and responsibility*. New Haven: Yale University Press.

Lamb, R. D. (2010). Mycrodynamics of Illegitimacy and Complex Urban Violence in Medellin, Colombia (PhD Dissertation, University of Maryland, MD). Retrieved from http://drum.lib.umd.edu/bitstream/handle/1903/10242/Lamb_umd_0117E_11062.pdf?sequence=1

Milgram, S. (1965). Some Conditions of Obedience and Disobedience to Authority. *Human Relations*, 57-76.

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.

Miller, A.G. (2014). The Explanatory Value of Milgram's Obedience Experiments: A contemporary appraisal. *Journal of Social Issues*, *70*, 558-573

Mullins, C. W. (2009). 'We Are Going to Rape You and Taste Tutsi Women' Rape during the 1994 Rwandan Genocide. *The British Journal of Criminology*, *49*(6), 719-735.

Munch-Jurisic, D. (2018). Perpetrator disgust: A morally destructive emotion. In T. Brudholm & J. Lang (Eds.), *Emotions and mass atrocity: theoretical and philosophical explanations* (pp. 142-161). Cambridge: Cambridge University Press.

Newman, L. S. (2002). What Is a "Social-Psychological" Account of Perpetrator Behavior? In L. S. Newman & R. Erber (Eds.), *Understanding genocide: The social psychology of the Holocaust* (pp. 43-67). Oxford University Press.

Newman, L. S. (2020). In what way is evil "banal"? Hanna Arendt's (interactionist) thesis. In L. S. Newman (Ed.), *Confronting humanity at its worst: Social psychological perspectives on genocide* (pp. 220–246). Oxford University Press.

Nollkaemper, A. (2009). Introduction. In A. Nollkaemper & H. Van der Wilt (Eds.), *Systemic Criminality in International Law* (pp. 1-25). Cambridge: Cambridge University Press.

Overy, R. (2014). "Ordinary men," extraordinary circumstances: Historians, social psychology, and the Holocaust. *Journal of Social Issues*, *70*(3), 515-530.

Perry, G. (2013). *Behind the Shock Machine: The Untold Story of the Notorious Milgram Psychology Experiments*. Victoria, Au.: The New Press.

Pettit, P. (2001). *A theory of freedom: from the psychology to the politics of agency*. Oxford University Press on Demand.

Punch, M. (2009). Why Corporations Kill and Get Away with it. In A. Nollkaemper & H. Van der Wilt (Eds.), *Systemic Criminality in International Law* (pp. 42-68). Cambridge: Cambridge University Press.

Ricoeur, P. (1992). *Oneself as another*. Chicago: The University of Chicago Press.

Shaw, M. N. (2008). *International Law* (6th ed.). Cambridge: Cambridge University Press.

Smeulers, A. (2004). What Transforms Ordinary People into Gross Human Rights Violators. In S. C. Carey and S. C. Poe, (Eds.), *Understanding Human Rights Violations: New Systematic Studies* (pp. 239-253). Aldershot: Ashgate.

Smeulers, A. (2008). Perpetrators of International Crimes: Towards a Typology. In A. Smeulers and R. Haveman (Eds.), *Supranational Criminology: Towards a Criminology of International Crimes* (pp. 233-266). Antwerp: Intersentia.

Smeulers, A. (2015). Female perpetrators: ordinary or extra-ordinary women?. *International criminal law review*, *15*(2), 207-253.

Smeulers, A. (2019). Historical overview of perpetrator studies. In A. Smeulers, M. Weerdesteijn, & B. Hola (Eds.), *Perpetrators of international crimes: theories, methods and evidence* (pp. 11-28). Oxford University Press.

Smeulers, A. & Grünfeld, F. (2011). *International Crimes and Other Gross Human Rights Violations: A multi- and Interdisciplinary textbook*. Leiden: Martinus Nijhoff Publishers.

Smeulers, A. L., Hola, B., & Van den Berg, T. G. C. (2013). Sixty-Five Years of International

Criminal Justice: The Facts and Figures. *International Criminal Law Review*, *13*, 7–41.

Smeulers, A., Weerdesteijn, M., & Holá, B. (2019). Theories, Methods, and Evidence. In A. Smeulers, M. Weerdesteijn, & B. Hola (Eds.), *Perpetrators of international crimes: theories, methods and evidence* (pp. 29-45). Oxford University Press.

Staub, E. (1989). The Roots of Evil: The Origins of genocide and other group violence. Cambridge: Cambridge University Press

Staub, E. (2011). *Overcoming Evil: Genocide, Violent Conflict and Terrorism*. Oxford: Oxford University Press

Staub, E. (2014). Obeying, joining, following, resisting, and other processes in the Milgram studies, and in the Holocaust and other genocides: Situations, personality, and bystanders. *Journal of Social Issues*, *70*(3), 501-514.

Sykes, G. M., & Matza, D. (1957). Techniques of neutralization: A theory of delinquency. *American sociological review*, *22*(6), 664-670

Taylor, C. (1989). *Sources of the self: The making of the modern identity*. Harvard University Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. basil Blackwell.

Tyler, T.R. (1997). The Psychology of Legitimacy: A Relational Perspective on Voluntary Deference to Authorities. *Personality and Social Psychology Review*, 1(4), 323-345.

Üngör, U. Ü. (2019). Perpetration as a Process: A Historical-Sociological Model. In A. Smeulers, M. Weerdesteijn, & B. Hola (Eds.), *Perpetrators of international crimes: theories, methods and evidence* (pp. 117-131). Oxford University Press.

van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2022). Why are general moral values poor predictors of concrete moral behavior in everyday life? A conceptual analysis and empirical study. *Frontiers in psychology*, 3834.

Waller, J. (2007). *Becoming Evil* (2nd ed.). New York, NY: Oxford University Press.

Whitehead, E. (2021). *BREAKING THE CHAIN OF COMMAND: Toward a theoretical framework: How and why soldiers refuse to obey illegal orders*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam]. s.n.

Williams, T. (2020). *The Complexity of Evil: Perpetration and Genocide*. Rutgers University Press.

Zimbardo, P. (2011). *The Lucifer effect: How good people turn evil*. Random House.

6. Conclusion and Discussion

6.1 Introduction

I set out on my investigation to get a better understanding of the relationship between people's morality and their behaviour. In our everyday experience, moral considerations often play an important role in our decision-making and our motivation for action. Also, many behaviours can be regarded as morally relevant, in the sense that moral considerations (good or bad, wrong or right) can play a role in the decision to act. Hence, people's individual morality – e.g., their moral beliefs, moral considerations, and moral values-, are important to take into account when we want to explain, predict, and understand people's decision-making and their actions. Also, for policymakers it is important to get a better understanding of how people's morality influences behaviour. Many of society's current challenges have a strong moral component and moral conviction is generally regarded as a durable source of motivation. The empirical and conceptual study of people's morality and its relation to behaviour is therefore a much needed direction.

In the last decades, the empirical study of morality has developed into a thriving practice, especially within moral psychology, leading to many interesting insights on morality's empirical reality (Alfano, 2015). However, one particular problem that this field faces is the predictive gap between empirically measured (aspects of) individual morality and morally relevant behaviour. *In this thesis,* I have focused on general moral values as a particular conceptualization of individual morality that has relatively recently received new impetus within the empirical study of morality. What I have called "empirical moral value theories", such as Moral Foundations Theory (MFT) (Graham et al., 2013), have been extensively used to capture people's differences in general moral preferences and have been mainly related to other attitudes and beliefs. Also, a number of behavioural studies have been including these measures (e.g., the Moral Foundations Questionnaire (MFQ)) in their models to test to what extent people's individual moral beliefs can explain certain forms of behaviour. However, so far, a systematic investigation into the relationship between general moral values and behaviour has been lacking. *It is the aim of this thesis to get a better understanding of the empirical and conceptual link between general moral values and behaviour. Secondly,*

a subsidiary aim is to make methodological suggestions following from my analysis to better capture individual morality and its influence on behaviour.

In order to reach these objectives I conducted the four studies presented in this thesis: two empirical studies with a substantial conceptual part (Chapters 2 and 3) and two fundamental conceptual studies, in which I make use of insights from the philosophical field of phenomenology (Chapters 4 and 5). The question now is what we have learned from these studies and how it contributes to a better understanding of the relationship between general moral values and behaviour. In the following I will first present my main conclusions. This is followed by further reflections on methodological directions and policy implications following from this thesis.

6.2 The main empirical conclusion: general moral values poorly predict concrete moral behaviours

The first main conclusion, drawn from chapter 2 and 3, is that general moral values are poor predictors of concrete moral behaviour. In the conducted empirical studies I used, overall, two different moral value scales to measure people's general moral values (MFQ and MAC-Q) and looked at a broad set of morally relevant behaviours. The great majority of the found standardized effects were under 0.1, a few between 0.1 and 0.2, and none above 0.2. This means that in my findings a general moral value could explain –at most- only 1 to 4% of the variance of the selected behaviours. The majority of the effects were found to be non-significant at the conventional 5%-level. Also, adding the moral value variables to the models resulted in a non-significant or only very small contribution to the explained variance compared to the models including only the demographic variables. The significant effects that I did find, were generally in line with what would be expected. Overall, I conclude, that the studies in both empirical chapters show a picture of few and very weak empirical relationships between the measured general moral values and behaviour.

Given that I found the same kind of effects in a range of different studies and the fact that the found effects are comparable to those found between other (moral) psychological constructs, such as general values and behaviour, I am confident about the robustness of the main empirical conclusion that general moral value measures predict concrete moral behaviour poorly. This also means that conceptualizing morality in terms of general moral values does not bring us much further in bridging the empirical gap between a person's individual morality and behaviour.

6.3 The main conceptual conclusion: the relationship between general moral values and behaviour is mediated by the moral self and therefore has an interpretive character and is dynamic

6.3.1 The concept of general moral values and its link with behaviour assumed by empirical moral value theories

Given the above conclusion and the fact that a theory such as MFT claims that people's moral foundations make up their morality, does that mean that these results imply that

people's morality or general moral values do not matter for their behaviour? Well, only when we would accept the conceptualization of general moral values –and, thus, of people's individual morality- that is implied by empirical moral value theories. As we have seen in chapters 3 to 5, this assumption is untenable.

As was argued in chapter 3 and 4, *empirical moral value theories* assume that general moral values are relatively stable personal dispositions, latently existing within the individual and causing a similar kind of behaviour across different contexts. A conceptualization that is implied by the chosen method of measuring moral values in a general way, outside of any context, and including them as general individual moral predictors in linear models of association and prediction. A measured general moral value can namely only function as a predictor when it is assumed that it has a similar effect across relevant decision situations and leads to the same kind of morally relevant behaviour. In other words, the general embracement of a moral value would (regularly) need to lead to a certain form of behaviour and the absence of such general embracement would (regularly) need to lead to not performing such behaviour. It is this assumption that makes it useful to measure people's moral values in a general way and relate them to other variables, such as attitudes and behavioural ones, in order to get a better understanding of how morality influences them. It is also this assumption that turns general moral values into assumed stable latent dispositions. As we have seen in chapter 4, when considering Moral Foundations Theory, this mechanical causal view of moral influence is theoretically sustained by presenting people's morality as psychological modules consisting of causal input-output regularities.

It is this conceptualization of general moral values, and individual morality more generally, as consisting of stable causal dispositions, that is challenged by the conceptual analyses in chapters 3, 4 and 5. Here, I argue that the relationship between moral values and behaviour is dynamic and has an interpretive character. A person's individual morality varies in which moral value becomes important within a certain situation and it depends on context what specific behaviour is considered to be right in light of it.

6.3.2 A cue from the positivistic paradigm of moral psychology

The idea that we have a dynamic and interpretive relationship with general moral values is first indicated by the conceptual analysis of the moral decision-making process as conducted in chapter 2 and 3. Making use of Rest's (1986) psychological model of Ethical Decision-making, I made clear that the process of moral decision-making is defined by at least four stages that can be conceptually separated: awareness, moral judgment, moral intention, and moral behaviour. As stated, for general moral values to be useful empirical concepts to use as predictors of behaviour it needs to manifest itself in these different stages in a similar manner across situations. However, when we look into the four stages of moral decision-making, linking general moral values to behaviour, this assumption and according conceptualization of general moral values as stable causal dispositions is unfounded. *Every stage of the moral decision process is influenced by numerous individual and situational specific influences, and ultimately, by the specific interpretations of and meanings given to the particular decision situation*.

In the awareness stage, it is the interpretation of the situation that determines whether and of which moral values one becomes aware as being relevant. When making a moral judgment and forming a moral intention, it are the specific meanings that moral values and other kinds of values receive within the decision-making context that determine how they are weighed vis á vis each other and the eventual moral judgment and moral intention it leads to. This directly relates to the fact that general moral values can lead to different interpretations within a specific situation and, therefore, to different specific moral considerations, judgments, and intentions. Interpretations even play a role in whether one thinks it is possible to follow-up on one's intention. This means that the importance one has given to a general moral value, through for instance a general questionnaire, does not seem to say much about the importance a moral value receives within a certain decision situation. Also, different specific behaviours can be considered to be right, invoked by the same moral value. This renders the influence of general moral values contextually dependent and indeterminate[50].

This analysis of the moral decision-making process has implications for how we view the relationship between general moral values and behaviour. Instead of a stable and causal relationship, it implies that this relation relies on the interpretation of our moral values within a specific context and that the relation with our moral values is dynamic, that is, it can vary across different situations. This also implies that the found weak effects are rather an expression of the inability of general moral value measures to grasp the specific influences of morality on behaviour than an indication of there being no influence.

This view, of course, needs further substantiation and articulation. In particular, the pressing question becomes how we can understand individual morality if it is not in stable cross-situational and general terms. For is not that what belongs to the individual defined by what remains constant across contexts and which is thereby distinguished from its changing environment? How is it possible to intelligibly talk about the individual moral person and individual morality if not in these general cross-situational terms? I sought an answer to these questions in chapter 4 and 5.

[50] To illustrate this, think of two people who would both score relatively high on the general moral value of care in MFQ. The first does volunteer work, while the second does not. Now the following seems very much conceivable. For the first person, moral considerations having to do with the value of care play an important role for doing volunteer work. For the second, though she generally cares about others, she is just not aware that doing voluntary work is a realistic option for her. A second possibility is that the second person did think about volunteering, but she has (moral) reasons not to do it. One could be that she wants to make extra working hours to take care of her family. Another would be that she wants to stay at home with her husband because he feels very lonely without her company. A third, that, in this case, she rather does something else, such as relaxing at home and watch TV. Does that mean that morality or moral values do not matter for doing volunteer work? No, for the first person they do and this can very well be the case for many. For the second person, moral values and moral considerations could also have influenced her decision-making not to volunteer: the moral value of care possibly made her decide to put her time in getting a payed job (an example of the fact that it is hard to tell which behaviour a general moral value may lead to). In case of staying at home with her husband, it is conceivable that in this particular situation considerations of loyalty influenced this decision, even if in general she would score higher on and find the moral value of care more important than loyalty. Regarding the third, more selfish reason not to volunteer, it is still perfectly conceivable that in other instances care considerations do prevail.

6.3.3 A turn to phenomenology

Though I was able to establish the dynamic and interpretive relationship with our moral values within positivistic moral psychology, its call for a further articulation and substantiation cannot be answered within this same paradigm. It needs to analysed on a more fundamental level, as it directly relates to questions about what constitutes a moral person, as the above questions signify. More particularly, it pertains to the fundamental question of moral identity, namely how one understands oneself and understands others as one and the same individual moral person.

I turned to phenomenology, a field that has extensively looked into the fundamental questions around (moral) personhood and (moral) identity or (moral) self. What I have sought there is a concept of moral identity that can incorporate this "oneness", together with the dynamic and interpretive relationship that we have with our moral values that was found in chapter 2 and 3.

6.3.4 A self containing moral multitudes

What we need to understand then is the moral version of what is so proudly expressed in Walt Whitman's poem *Song of Myself: "*Do I contradict myself?// Very well then I contradict myself,// (I am large, I contain multitudes.)"[51], namely *an individual person that can contain moral multitudes*. This was found in Paul Ricoeur's (1992) idea of personal identity and his narrative concept of the moral self.

With Ricoeur's important distinction between *ipse*- and *idem*-identity, it became clear that empirical moral value theories implicitly consider moral identity only in terms of *idem*-identity and do not incorporate the *ipse*-identity of the moral person[52]. As described in Chapters 4 and 5, *ipse*-identity denounces the reflexive relationship one has with oneself –a self-constancy that is expressed in the act of promising and that is an answer to the irreducible question 'who?'. *Idem*-identity defines a thing or a person in terms of sameness, as "this" or as "that".

Empirical moral value theories' identification of the moral person with general moral dispositions is then one that is fully in terms of "this" or "that", i.e., in terms of *idem*-identity. As Ricoeur makes clear, *idem*-identity with regard to persons is expressed by their character, including their moral character. That is, one's (moral) dispositions in the sense of developed habits and identifications with values and exemplars. However, following Ricoeur, this identification of the moral person is only half the story. Dispositions or character, are always *someone's* dispositions or character, i.e., *they belong to a self*. The reflexivity of *Ipse*-identity or selfhood implies that someone *is* not simply one's dispositions, such as moral values, but that one relates to them. Identifying oneself as a moral person in a certain situation entails recognizing oneself in the moral values that one calls one's own, attuned to the specific context. The relation with our

[51] Also see, or better, listen to the song by singer and songwriter Bob Dylan "I contain multitudes" which is a nice description of this idea and which was inspired by these lines.

[52] Though not the explicit focus of this thesis, it seems that the same can be said about other theories concerned with general moral psychological constructs. The moral psychological construct of moral identity, for instance, including the general manner by which it is measured as a concept of moral motivation, also points to an understanding solely in terms of *idem*-identity.

moral values is then not ossified, but rather one that needs to be maintained and that asks for self-constancy -self-constancy that is provided by the interpretation and recognition of oneself within a certain situation, in light of one's moral values. The interpretation of oneself as "this" or as "that" moral agent within a certain context, expressing *idem*-identity, then entails a dynamic and flexibility that is given by the dialectic relationship between selfhood and sameness constituting moral personhood. It can at the same time grasp the mutability of the moral person as well as its oneness.

The dialectic relationship between *ipse* and *idem,* constituting personal identity, and its aspect of mutability and permanence is further elaborated within Ricoeurs' concept of narrative identity. As explained in Chapters 4 and 5, narrativity functions as a mediator between the self in terms of selfhood and its expression in terms of sameness. The identification or conception of ourselves rests on the life story that we tell about ourselves -similar to the identification or understanding of characters in a fictional story. As in fiction, the oneness of our life story, and accordingly, of our identity, is constituted through the configuration of the plot, which connects the separate contingent situations we encounter, the thoughts we have, and the actions we perform into one meaningful storyline and understanding of ourselves. In other words, our story and our identity contain multitudes -a multiplicity of actions and thoughts, which may even sometimes seem to contradict each other- but also form one congruent understanding of oneself through the narrative that connects the questions of "who did what and why?" into one meaningful configuration.

6.3.5 Moral values as touchstones

It is in the stories we tell about ourselves, together constituting our life story and self-understanding, that our dispositions figure, such as our habits and values, and these stories are imbued with moral meaning. As explained in Chapter 4 and 5, narratives are never ethically neutral and giving an account of one's action is also explaining oneself – why it was the right thing to do. These stories express the "what" of the "who" and thus *characterize* us morally. They express a conception of how we morally think, behave, and value. But, at the same time, this expression of the "who" in terms of "what" is subjected to a constant reinterpretation and development as we can retell and reinterpret the stories we tell about ourselves, and, secondly, because we constantly encounter new experiences that are appropriated as part of the whole –our moral identity.

This movement of the plot -by which the part (a new experience) gets a meaning in light of the whole (one's life story) that appropriates it, while these new experiences inevitably further develop and change the life story and conception of self- leads to a specific conception of moral decision-making. The appropriation namely implies a mutual fitting between the new decision or action and the understanding of oneself as a moral person. The situation encountered is given meaning in light of one's life story or understanding of oneself, while figuring out what to do entails a reinterpretation and recognition of oneself when attuning to the specific decision situation.

Moral decision-making is then not characterized by a straightforward causal determination or application of an embraced moral value, but rather by a recognition

of oneself in one's decision and action in a specific context in light of the moral values that one calls one's own. *Moral values do not have the character of a causal determinant, but should rather be regarded as touchstones* that are called upon to figure out whether one's specific action is in line with the moral understanding of oneself. Which moral value becomes important and which specific meaning it gets within the decision situation is then very context specific, but this variance is still an integrated part of the congruent understanding of self. *Due to the underlying concept of moral self, individual morality is dynamic, context specific, and characterized by interpretation, but at the same time the expression of one congruent moral person.*

6.3.6 The narrative concept of the moral self as a fundament to the interactionist approach towards moral behaviour

In Chapter 5, I further theoretically validated the narrative concept of the moral self by investigating its relevance to understanding the relation between individual morality and moral behaviour within the extreme context of international crimes. In this field, explanations of perpetrators are dominated by situational factors. At the same time, the moral dimension of the individual is, though somewhat implicit, an important ingredient in social psychological and criminological theories to explain and understand the perpetration of and resistance to such crimes.

From a purely situational perspective, individual morality is presented as only a toy of situational influences, which transform ordinary men into killers who have come to believe that their actions are the right thing to do. However, this negation of the role of the individual does not do justice to the expression of the moral self which such a judgment still entails. There is still *someone* judging. This mistake becomes all the more apparent when we consider situations where the role of the individual has risen to great heights, namely in the case of resistance, where the individual stands his or her ground against such situational influences and accompanying perverted norms and is able to make a different moral judgment. Pure situationism has a hard time dealing with such cases.

With Ricoeur's concept of the narrative moral self, it is possible to go beyond the situation vs disposition dichotomy, that this contrast between perpetrators and resisters all too often invokes. It substantiates and deepens our understanding of a viable and fundamental interactionist position that can explain these acts. This narrative concept of moral self is able to incorporate a radical change in one's individual morality -a changing relationship to and interpretations of one's moral values- under influence of the situation, without letting the moral person vanish in thin air. The transformation that perpetrators undergo is namely understood as a reflexive process of self upon self, by which the person reinterprets itself through narrative. There is an individual who's narrative changes as it incorporates one's action under influence of situational pressure into one's self-narrative, but who also acts upon new situations on the basis of one's self-narrative, in which these experiences are appropriated. The individual person in its reflexive capacity has a crucial role here, if only in (not) rationalizing and accepting this new understanding of self when deciding to act from this self-understanding. This

interactive view makes the potential power of the individual moral person manifest, for good or for bad. Accordingly, resistance should be understood as the refusal to understand oneself as committing such acts.

An important conclusion of this study in Chapter 5 is then that even in the extreme context of international crimes in which situational factors are very dominant, individual morality still matters for behaviour. Individual morality should, however, be regarded as varying and dynamic and its relation to behaviour depends on its local and contextual interpretation. As we saw with the example of perpetrator's disgust in chapter 5, only knowing a person's moral dispositions is not sufficient to understand how an individual's morality influences behaviour. These moral dispositions get a certain meaning as part of one's narrative within a certain context. To grasp these interactions, we need to focus on these contextual interpretations.

This has direct consequences for how to understand the Model of Ethical Decision-making that was used and extended in Chapters 2 and 3 to conceptualize the relationship between general moral values and behaviour. The above shows that the usual causal presentation of this model, such as by Schwartz (2016), does not suffice. Something that is already indicated in these two chapters and now receives a more fundamental substantiation. Schwartz's elaboration of the model clearly presents decision-making as a causal process, where general individual and situational variables determine the outcome of the different process stages. Following our analysis, based on Ricoeur's narrative concept of the moral self, it becomes even clearer that such an approach fails to understand ethical decision-making, if we do not take into account the specific meanings that these individual and situational aspects receive within a certain context. The different process stages are constituted through a process of interpretation rather than by mechanic causation. Taking into account these specific meanings would be really taking a 'person-situation interactionist approach', as Schwartz (2016, p. 772) claims to do.

6.3.7 A constant presupposition of the self within empirical moral psychology

Finally, I want to emphasize that the articulation of the notion of self that I have explicated is only making explicit what is already implicit in the different empirical approaches that have been discussed in this thesis –so the model of Ethical Decision-making, Moral Foundations Theory, and the empirical approaches towards perpetrators of international crimes. This is most clearly seen in MFT, when its developers relate MFT to aspects of virtue theory –as we saw in Chapter 4. Attuning one's virtues to the local socio-moral context and relating to them through narrative, presupposes a reflexive and interpretive self. It is also such a self that needs to be presupposed when, with regard to the model of Ethical Decision-making, we speak of *someone* who is interpreting the situation and becoming aware of the relevance of certain moral values, who has the capacity to rationalize them away, and who weighs them against other (moral) values. Also in Chapter 5 it is made clear that the description of the transformation process of perpetrators by empirical theories presupposes a self, as I have articulated following Ricoeur.

What this comes down to is that these empirical moral theories cannot evade the common sense idea that persons are more than the empirical factors on which these theories generally focus. Persons relate to these factors within a particular context, giving room for interpretation, flexibility, adjustment and different understandings of oneself. In this way, this notion of self does justice to the central idea behind each of the two main empirical approaches towards human behaviour –situationism and personal dispositionism- in the sense that an individual's behaviour can vary across contexts, but also that it is an individual person with certain (moral) qualities. As such, it promises to be a useful fundament for interactionist approaches towards moral behaviour and other types of behaviour more generally.

6.3.8 Implications for empirical moral value theories: explaining the individual morality-moral action gap

At the end of this conceptual exposition, the question comes up what this means for the project of measuring individual morality through general moral value measures. Measures that are, without much conceptual thought, included by behavioural researchers into models of association and prediction with the aim to investigate the influence of morality on certain types of behaviour. I argue that this practice does not suffice and needs substantial revision.

In my conceptual analysis, I have emphasized the dynamic and interpretive relationship between general moral values and behaviour. The conceptualization of moral values as touchstones by which one recognizes oneself, while attuning to a specific decision situation, implies that our relationship with moral values cannot be expressed by a non-contextual given level of importance. The importance a moral value receives varies per context. Furthermore, if we interpret general moral values in light of our individual life stories, or particular episodes within these life stories that are relevant to the particular decision context, then the meaning of a general moral value can strongly differ between persons and between situations, and, therefore, lead to very different specific behaviours. *This leads to the conclusion that general moral value measures are unable to grasp the specific moral influences that play a role in the constitution of behaviour*. This makes the general measurement of moral values and their usage as moral predictors of behaviour rather useless. Based on my conceptual analysis I contend that this is the main reason for finding such few and weak effects between general moral values and moral behaviour.

Moreover, concerning the few effects that are found, it is unclear what these can really tell us. What does it mean that a general encompassment of the value of care contributes for a very small fraction to using less aggressive communication in traffic (Chapter 2)? Or the general encompassment of fairness to not visit vulnerable people during the Covid-19 pandemic (Chapter 3)? Care and fairness in what way? And what can we do with such information? Due to the general character of the moral value and its minimal effect, such findings do not contribute much to our understanding of the role of morality in constituting behaviour.

Concerning other moral dispositions, the weak effects that are generally found between them and behaviour may be explainable in the same manner. Given what has

been said about the importance of taking into account the specific meanings that moral dispositions can get within a certain context and the generality by which these are usually measured, suggests that this could be an important explanation for the individual morality-moral action gap found within moral psychology more generally. For instance, people's moral awareness and moral motivation may turn out to be quite context specific, depending on varying specific understandings of ourselves within different situations. This also means that to get a better understanding of how morality plays a role in the constitution of a certain form of behaviour, we do not need to extend our models with more general measures, but we need to zoom in and take the specific meanings reigning within a certain context into account.

Does the above mean that general moral values measures, such as MFQ, are good for nothing? No, not necessarily. Assuming that these measurements are a reflection of a very abstract and non-applied idea of which moral values one finds important, it is not surprising that they have been found useful to predict other general and abstract attitudes, such as political ideology. Even more so, because the general moral values included by MFT, categorizable in "individualizing" and "binding" values, fit the two major political streams of liberalism and conservatism like a glove. These, that we could call "politico-moral values" can possibly generate a deeper understanding of choices on this abstract politically charged level. However, they become less relevant when they are related to more concrete daily life behaviours, such as those studied in Chapter 2 and 3, and, presumably, even less so when we would consider our daily life moral interactions among each other. These moral foundations can therefore be hardly understood as what really matters for morality. The question then remains how we *can* make progress in capturing individual morality and its influence on behaviour. In the following section I formulate some suggestions.

6.4 Methodological implications following from the conceptual analysis

First of all, I need to emphasize that the analysis of this thesis does *not* imply that measuring individual morality and linking it to behaviour is a useless enterprise for getting a better understanding of the influence of morality on behaviour. On the contrary, this investigation has brought the moral person forward as indispensable within the process of constituting moral behaviour. The systematic empirical study of people's morality and its connection to behaviour, with its generalizable and objectifying abilities, is an important tool to get a better insight in certain meaningful broader tendencies concerning the influence of individual morality on behaviour.

What my analysis does show is that the individual morality measured by empirical moral value theories –and, for that matter, by many empirical approaches within moral psychology- misses an essential aspect of our individual morality. I have articulated that aspect with the narrative concept of moral self. It stipulates that we have *a relationship* with our individual morality and that this has a narrative structure, asking for constant interpretation and giving it a certain dynamic. With regard to measuring individual morality, this means *that we should look for ways in which we can somehow*

incorporate the dynamic and interpretive character of the moral self as best as possible in our measurements of individual morality. The fact that the moral self is hard to capture within the empirical paradigm, does not mean that its relevance for behaviour cannot be taken into account in designing empirical methodologies.

6.4.1 Measuring moral values within a delineated context

A first implication that I have argued for in this thesis is *to confine the empirical investigation of moral values to a delineated context of behaviour*. For instance, if we want to know how moral considerations play a role in the decision (not) to visit vulnerable others within a Covid-19 context, it makes more sense to look into the specific considerations that play a role there, than to measure moral beliefs on an abstract and general level. Only when we zoom into a certain context it becomes clear what kind of (specific) moral values play a role and what kind of specific meanings these receive. Chapter 3 provides some evidence for this train of thought. There I found that moral values that were tailored to the context of the behaviour studied (eating meat) performed better and explained away the effect of the general moral values. More extensive empirical research is of course needed to find out whether this is a robust finding. My empirical and conceptual analyses imply that making use of more specific moral values is a first good step to get a better understanding of the influence of morality on specific forms of behaviour.

Measuring context-specific values and relating them to behaviour to which they have been tailored may raise the worry of endogeneity and a reverse direction of causation (Kroesen & Chorus, 2018). However, it can be questioned whether this is really problematic. Following Ricoeur's concept of narrative self, it is one's actions that is constitutive of one's self-concept, as they are appropriated as part of one's life story. This self-narrative, in turn, directs further action. In other words, one's personal (moral) attitudes are, amongst others, constituted by one's actions and this will again influence further decision-making. Attitudes and actions are then inevitably more intertwined than is often supposed.

6.4.2 Start with a qualitative research phase

The second methodological implication following from my conceptual analysis, is that if we really want to know the specific moral considerations that play a role within a certain context of behaviour, *we need to include a qualitative research phase into our investigation*. Because of its exploratory character and focus on understanding the specific interpretations and meanings that constitute behaviour, it is the most apt form of research to express the different moral narratives that can play a role within a certain context. Qualitative research analysis itself can of course give valuable insight into the role of morality in constituting behaviour. In Chapter 5 I have emphasized the importance of studying the self-narratives of perpetrators and resisters through ego-documents to get insight into how one's moral self-understanding and the specific interpretation of the situation can lead to certain behaviour.

Furthermore, qualitative data collection can also be used as a preceding phase to more quantitative methods. As explained in Chapter 4, one possible form is, when

qualitative data is collected on a large enough number of persons, to turn these into context-specific quantitative scores (e.g., see for a non-context specific example the procedure used by Boyd et al., 2015). Another form is to design context-specific quantitative measures based on the collected qualitative data and use these to collect quantitative data on a new group of respondents. An example of a method that makes use of this procedure is Q-methodology[53]. An interesting feature of Q-methodology is that it assumes that within a certain context there is a limited number of different narratives to which all people's views can meaningfully be reduced. This makes it possible to categorize people accordingly and relate this to other variables, such as behaviour, while maintaining the rich and context specific perspective that can meaningfully explain the latter. In principle, such a method leads to much more meaningful results and understanding than only knowing that a certain general moral value associates with a certain form of behaviour.

Of course, such methods inevitably lead to reduction and place moral influence back in a causal framework. However, this is done while taking the varying character of individual morality and the specific meanings it receives as much as possible into account. The advantage of such empirical research is that it can give insight into broader meaningful tendencies with regard to the role of morality within a certain context of behaviour.

6.5 Policy implications

6.5.1 The relevance of individual morality for policymakers

Policymakers today are not to be envied. Dutch society and humanity as a whole are currently confronted with multiple crises -and not *just* because the word "crisis" is subject to inflation. In fact, we stand for major challenges that ask for some major transformations. Not surprisingly, the transformations that these problems push for are ultimately transformations of behaviour. One can think of moving towards more sustainable life patterns due to climate change, other ways of social interaction due to the Covid-19 pandemic, and other ways of treating and approaching other people in our strive to move towards a more inclusive society. Exactly because the behaviours involved are connected to the idea of "a good society", they by definition have a moral dimension. What makes it complicated is that the behaviours that we are to move *away* from have actually for a long time been considered as morally acceptable or even as the morally right thing to do and still *are* for the people who are not yet ready to move. In both directions, morality plays an important role and this is something policymakers need to take account of. Both in the sense of understanding why it is so difficult to move people toward new ways of behaviour, as in seeing possible opportunities there for behavioural change.

Of course, policymakers have some important traditional measures in their toolbox to accomplish behavioural change, such as legal tools and monetary incentives. These are very direct and tangible measures that have proven their effectiveness.

[53] See for a more extensive explanation of this method the discussion of Chapter 4.

However, given the strong moral relevance of the behaviours in question, *triggering people's moral motivation and changing their moral views about a certain subject are also important avenues for policymakers to commit to. Not* necessarily *instead* of the so-called "hard measures" –the law and money- but rather as an accompanying strategy. In the end, it seems that behavioural change is best accepted, is most enduring, and least expensive when people themselves believe that it is the "right thing to do". Moreover, non-governmental organizations that are committed to certain forms of behavioural change, often do not have hard measures available to them and mainly need to aim for the power of moral influence. In the next section I will explicate some important lessons and implications following from this thesis that can help to make progress here.

6.5.2 Implication one: target concrete moral considerations instead of general moral values

A general call for norms and values

First of all, my analyses imply that targeting people's *general* moral values in order to change certain patterns of behaviour will not lead to much success. For instance, the famous general call for more attention for "norms and values" by former Dutch prime-minister Jan-Peter Balkenende and his 2002 election slogan "Decency: that is something you do"[54], are not very effective ways to really change the tougher and rougher social manners he said he was fighting. Such abstract and somewhat hollow phrases can be interpreted in multiple ways, they can actually be utilized to rationalize all kinds of behaviour, and it is very doubtful that many people in their daily interactions with others have such a call on their minds, leading them to "more decent" behaviour. Though this call and slogan can have multiple purposes, it is not very effective for what seems to be its ultimate purpose, namely to actually change people's behaviour. In light of the analysis of this thesis, the mentioning of the general value directly followed by the inducement to action in the slogan even sounds somewhat ironic -for: what decency? And how?

The OneLove captain's armband

Besides such general calls by politicians, the targeting of general moral values have also been part of more concrete projects to change people's behaviour. One interesting recent example is the national "OneLove-armband"-action, initiated by, a.o., the National Dutch Football Association (KNVB) in October 2022[55]. For this action, the captains of every football team playing in the highest Dutch football league were asked to wear an armband with the colourful OneLove-logo, in the context of "international coming-out day". This action was explained as a statement against discrimination within football and an expression of respect, i.e., that everyone is welcome in football, no matter what one's race, gender, or sexual preference is (Ons voetbal is van iedereen,

[54] English translation of the Dutch "fatsoen moet je doen"

[55] Note that this example is specifically about the national action involving the OneLove-armband and not about the international one at the World Cup 2022 in Qatar, which also led to commotion ("Nederland en andere landen", 2022)).

2022a; Eredivisie, 2022). More specifically, this action was part of the broader *Ons voetbal is van iedereen*[56]-campaign of the KNVB that focusses on the prevention of discriminating and racist *behaviour* in football -such as shouting discriminating slogans in the stadium- by raising awareness and setting the norm that discrimination is not OK (Ons voetbal is van iedereen, 2022b, 2022c; Ons voetbal is van iedereen, 2020). The target here is the greater majority of people who 'do not deliberately want to hurt other people's feelings' (Ons voetbal is van iedereen, 2022b, para. 2).

However, as it turned out, not all captains wanted to participate. Two out of eighteen captains eventually refused to wear the OneLove-armband. Because the project was linked to "national coming out day" and because they associated the colourful logo with the rainbow symbol of the lhbtiq+ -community, the two captains felt uncomfortable to wear it out of religious considerations. They did emphasize that they respect all people, including lhbtiq+ persons ("Kokcu weigert regenboogband", 2022; "teleurstelling over maar ook bijval", 2022). This led to some considerable societal discussion, which made the KNVB cancel a planned repetition of the action. The football association found that the discussion revolved too much around "who is wearing the armband" instead of focusing on the central message of it.

From the insights of this thesis there are a couple of things to say about how this project worked out and how it could be improved. First of all, again, following my analyses and conclusions, it seems implausible that a general call for respect and a general statement that "everyone is welcome in football" has much effect on the concrete discriminating outings that happen in football. The general moral value of 'respect' is something that many will embrace in general, but which is interpreted and applied in many different ways, even when it is connected to the lhbtiq+ community. It is doubtful, to say the least, that a supporter who weekly scolds at the referee by calling him a homosexual of some sort -just like the rest of his friends and the rest of his section- will suddenly realize that he potentially hurts other people's feelings because his team's captain calls for "respect" by wearing an armband. It will often not be realized that such a general value is actually applicable to behaviour that has become part of someone's custom and that is regarded as normal within the specific context. Also, its breach is easily rationalized and trivialized due to its general meaning. To refer back to my analysis using the model of Ethical Decision-making, the link between such a general value and the concrete behaviour within context is too indeterminate. Or, in terms of my analysis of the moral self, due to the flexible understanding of oneself as a moral person, one can incorporate being a fanatic supporter that scolds with homosexuality with a general notion of oneself as being respectful to others, including lhbtiq+ persons. A general call for respect, even a general call for the respect of lhbtiq+ persons, will not easily breach that.

Secondly, also the refusal of the two captains to wear the OneLove-armband can be led back to the generality of the moral value that the armband expresses. This namely leaves a lot of room for interpretation. An interpretation that always stems from one's own perspective and understanding of self. It is quite conceivable that this room for

[56] Translated in English it means something like: "our football is for everyone".

interpretation led the two refusing captains to interpret the general expression of respect for homosexuality in a way that is against the more conservative religious and family values by which they identify themselves. Think of promoting the marriage between people of the same sex. This, while for others the central meaning of the armband lies elsewhere.

If the ultimate goal is to counter certain specific forms of behaviour then this clash was unnecessary. It is my guess that the two refusing captains -also because of their stated respect for lhbtiq+ people as persons- would have been much more receptive towards supporting a specific message on the condemnation of scalding with homosexuality during football matches. The central message would be that it hurts the feelings of people who have this sexual preference and who are in the stadium or watching at home. At least, such a message leaves far less room for differential interpretations and is, in fact, therefore much harder to refuse one's support to than a general and abstract moral message. My analysis then points towards the opposite of what was claimed by some of the criticasters of the refusing captains, namely that it was incomprehensible that the captains were unable to support such a *general* value of respect. To say it in the terms of the analysis in this thesis, the captains would have more likely been able to appropriate the more specific moral action as part of their self-narrative than their interpretation of the general one. *So, if the ultimate aim is to reduce certain forms of harmful behaviour, a project or campaign should be explicitly focused on the specific moral considerations that are involved with the behaviour that it wants to drive back.*

I expect that such an approach would also be more successful in changing the behaviour of the supporters themselves. At least for those who are assumed to be susceptible to such moral reasoning as they do not want to deliberately hurt other people. Following our analysis, an action or campaign would have to make people aware of the moral considerations that scolding with homosexuality can hurt other people's feelings and it should be brought in such a way that it is clear that this outweighs the fun, or any other consideration or urge for doing it. Ultimately, it should lead to the idea that continuing to scold in these terms can no longer be reconciled with one's understanding of self. One possible way by which this could be done is to hang large posters throughout the stadium with different supporters who identify as lhbtiq+ and which expresses what it does to this person when others in the stadium use their sexual preference as a way of scolding at people.

6.5.3 Implication two: use the model of Ethical Decision Making as a guideline and the understanding of self as an aim when targeting moral motivation

My analysis on the basis of Rest's (1986) model of Ethical Decision-making shows that someone's moral value can only become effective when it goes through all identified stages of moral decision-making: awareness, judgment, intention, and behaviour. These stages can therefore be considered as conditions that need to be met before someone can be morally motivated to perform some sort of socially desirable behaviour. For policymakers this means that they need to commit to each stage when wanting to steer

towards behaviour through people's moral motivations. The model can be used as a practical guide for this purpose.

Different policies or campaigns can be specifically focused on one of the different stages. So for instance, when trying to steer farmers towards more sustainable ways of farming by triggering their moral motivation, the first thing would be to raise awareness about the moral considerations that play a role here. One possible way to do this is through developing educational programmes and campaigns on the effects of different modes of farming. When directing one's attention towards the moral judgment and moral intention stage, it is important to increase the chance that moral considerations motivating a more sustainable farm mode prevail. Besides emphasizing the moral rightness of a certain mode of farming, this also entails making moral and egoistic considerations for less sustainable farm modes less powerful. An important cue here can be the rationalizations that are invoked by farmers to not move towards more sustainable modes and see if these can be made less attractive. Practically, it could mean making less sustainable modes of farming less profitable and offering compensation, to make a transformation towards sustainable farming possible without risking to lose one's own life and bread and that of one's family. Finally, policy should be directed towards turning good intentions into action. Policymakers must be extra alert on practical barriers that deny farmers who want to transform the chance to do this. Also in designing new policies there should be specific attention for the accessibility of opportunities of change for farmers that want to transform.

Of course, in order to know what kind of policies could work within a certain context to establish the different stages of moral decision-making, further investigation will often be needed. Following my methodological suggestions, such an investigation would have to start with a qualitative phase in which the specific (moral) meanings that play a role within a certain behavioural context are brought to the surface. Here, the stages of the model of Ethical Decision-making can again serve as a guideline. Relevant questions concerning the awareness stage are: How do actors interpret their situation? What moral meaning do they give to their behaviour? To what extent are they aware of relevant moral considerations? What is the reason for not being aware of certain moral meanings? Concerning the moral judgment and moral intention phase, possible relevant questions are: What are the specific moral and egoistic considerations that play a role in the different choices of behaviour? What are rationalizations that are often used? Etc.

What one is actually doing here, is in a systematic way mapping out the specific interpretations and meanings of situations and behaviour that are closely connected to the moral understandings of self that play a role within a certain context. In fact, from Ricoeur, it can be said that the coherence and connection between the different stages of the model of Ethical Decision-making is given by one's self narrative. Though analytically separable, these stages and their expressions are all part of one congruent moral understanding of oneself within a concrete situation where an ethical decision needs to be taken. As described in Chapter 4 and 5, these self-narratives always involve ethical ideals –abstract ideas about what it means to be a "good this or that", or, in the most general sense: "to live a good life" or "be a good person"- and into which, on a

lower practical level, ideas about "good practices" and "good actions" are nested. The ultimate aim then of mapping out the specific meanings and interpretations in a congruent way is to map out these moral self-understandings that are relevant within a certain context. Getting better insight in these moral understandings of self gives better insight in the views and behaviours of the different actors and it is also on this level that behavioural change on the basis of moral motivation can be made.

For policymakers it is essential to realize that trying to influence people's moral behaviour, is endeavouring for change on the level of people's self-narrative or moral identity. This, on the one hand, explains why such influence can encounter considerable resistance. On the other, it means that change on the level of self-narrative forms potentially an effective point of engagement for behavioural change. The heavy protests by farmers in the Netherlands in 2022, as a reaction to the announced nitrogen policy, which for many farming companies would mean to radically change their way of farming, to move, or to quit, can at least partially be explained by the fact that the demanded change in behaviour directly conflicts with the farmer's moral identity. That is, the moral self-understanding of farmers which involves ideas of what it means to be a good farmer and the concrete practices and actions that are part of it. Raising moral awareness and influencing moral considerations of farmers, will have to take place on this level of self-narrative and identity in order to be really effective. This means that the narratives about what constitutes good farming and to which the farmer mirrors him- or herself needs to be changed. The narrative of "good farming" would not only have to entail the idea of taking good care of one's animals and make profitable use of one's land, but also to take care of the biodiversity in one's region and make the land liveable and useable for future generations.

One way of offering alternative narratives is by setting other persons as an example. Such exemplars can be regarded as life stories that express certain values and ideals through their concrete practices and actions. An exemplar can then offer an alternative narrative about what constitutes a good practice and practically show how this can be performed. New ideas about what constitutes a "good practice" and ultimately "a good life" is thereby made tangible and eligible for appropriation into one's own narrative. A policy measure in this regard in the context of moving towards more sustainable modes of farming, would be to set up a programme by which mainstream farmers are stimulated to visit and participate in more sustainable farm companies.

With regard to the classical "hard" measures, following cognitive dissonance theory, legal and monetary incentives can also have a role in targeting and changing self-narratives. Namely, by simply pushing people towards new forms of behaviour. However, to make this new behaviour an enduring part of who one is, it needs to be accompanied by the right moral narrative. Appropriating the new behaviour as a part of one's life story entails justifying it, and this needs to be in the right terms if it is to become part of an enduring moral motivation to perform this new practice. The farmer who is partly pushed to a more sustainable mode of farming by hard measures needs to have a moral story available to him or her by which it can be understood as "good practice' and by which he can interpret him or herself as a "good farmer". It is therefore essential that hard measures are accompanied by a broader moral story, embedded in

moral values and ideals, that offers perspective and which is made concrete in certain good practices and actions. Appealing to moral values is in this way important for moving people towards certain behaviour, but always on the condition that it becomes clear what they mean for concrete action. The use of exemplars, who are themselves the concretization of such ideals and values, is a helpful tool here.

6.6 A final reflection

The analysis of the context of international crimes in Chapter 5 reminds us that targeting people's moral beliefs in order to constitute behavioural change can also have a dark side. In a very sinister way, these situations demonstrate the relevance and potential of individual morality for behavioural influence. Action is always accompanied by a certain moral perspective on oneself and on what one does. These moral-self understandings are as vulnerable as they are powerful. On the one hand this puts great responsibility on the shoulders of those who are in the position of influencing them to not abuse this power, e.g., policymakers, politicians, and influencers. On the societal level this asks for institutions that warrant a multiplicity of narratives about "the good", their critical discussion, and the freedom to abandon one in favour of another. On the other, self-narrative lends the individual the power to say "no" when someone else's perspective is unacceptable. It is this relationship that a person has with him or herself on which the resistance against perverted norms depends -and the conduct of moral behaviour more generally- and that should be part of our analyses and explanations to improve our understanding of moral behaviour.

References

Alfano, M. (2016). *Moral psychology: An introduction*. John Wiley & Sons.

Boyd, R. L., Wilson, S. R., Pennebaker, J. W., Kosinski, M., Stillwell, D. J., & Mihalcea, R. (2015). Values in words: Using language to evaluate and understand personal values. *Proceedings of the 9th International Conference on Web and Social Media, ICWSM 2015*, 31–40.

Kroesen, M., & Chorus, C. (2018). The role of general and specific attitudes in predicting travel behavior – A fatal dilemma? *Travel Behaviour and Society*, *10*(June), 33–41. https://doi.org/10.1016/j.tbs.2017.09.004

Eredivisie. (2022, October 13). *Voetbalwereld vraagt komend weekend aandacht voor LHBTI+ acceptatie*. https://eredivisie.nl/nieuws/voetbalwereld-vraagt-aandacht-voor-lhbti-acceptatie/

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental Social Psychology, 47*, 55-130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Nederland en andere landen zien af van OneLove-band na aangekondigde gele kaart. (2022, November, 21). NOS. https://nos.nl/artikel/2453245-nederland-en-andere-landen-zien-af-van-onelove-band-na-aangekondigde-gele-kaart

Kökcü weigert regenboogband: 'Voel mij vanuit geloof niet aangewezen persoon' (2022, October 16). NOS. https://nos.nl/artikel/2448597-kokcu-weigert-regenboogband-voel-mij-vanuit-geloof-niet-aangewezen-persoon

Teleurstelling over, maar ook bijval voor weigeren OneLove-aanvoerdersband Kökcü. (2022, October, 17). https://nos.nl/artikel/2448761-teleurstelling-over-maar-ook-bijval-voor-weigeren-onelove-aanvoerdersband-kokcu

Ons voetbal is van iedereen. (2022a, October, 13). *Het komend weekend maakt het voetbal opnieuw het statement dat iedereen welkom is en niemand mag worden uitgesloten. Een boodschap die helaas nog steeds actueel en noodzakelijk is*. https://onsvoetbalisvaniedereen.nl/statement-voetbal-iedereen-is-welkom/

Ons voetbal is van iedereen. (2022b, December 19). *ONZE AANPAK: 'ONS VOETBAL IS VAN IEDEREEN– SAMEN ZETTEN WE RACISME EN DISCRIMINATIE* BUITENSPEL'. https://onsvoetbalisvaniedereen.nl/onze-aanpak#voorkomen

Ons voetbal is van iedereen. (2022c, December 19). *ONELOVE: FOOTBALL HAS THE POWER TO UNITE PEOPLE*. https://onsvoetbalisvaniedereen.nl/one-love/

Ons voetbal is van iedereen. (2020). *Ons Voetbal is van Iedereen: samen zetten we racisme en discriminatie buitenspel*. https://onsvoetbalisvaniedereen.nl/wp-content/uploads/2020/09/Bijlage-4-aanvalsplan.pdf

Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.

Ricoeur, P. (1992). *Oneself as another*. The University of Chicago Press.

Samenvatting

Inleiding en achtergrond van het onderzoek

Dit proefschrift probeert een beter begrip te krijgen van de relatie tussen moraliteit en gedrag. In onze dagelijkse ervaring speelt moraliteit, in de zin dat iets het juiste of goede is om te doen, een belangrijke rol in ons handelen. Veel dagelijkse sociale interactie waarin we iets voor de ander over hebben is moeilijk te begrijpen als morele overwegingen geen rol zouden spelen. Denk aan opstaan voor een minder valide persoon in de trein, een vreemde de weg wijzen, of jezelf toch nog even testen met een corona snel-test voordat je naar een feestje gaat. Ook in onze ideeën over een betere samenleving en in het daaraan verbonden gedrag speelt moraliteit een belangrijke rol. Mensen eten minder vlees om dierenleed te verminderen en er zijn mensen die minder gaan vliegen om een duurzame aarde door te geven aan volgende generaties. Tijdens dit schrijven wordt een omstreden wereld kampioenschap voetbal in Qatar georganiseerd waar sommige Nederlandse voetballiefhebbers niet naar kijken op TV vanwege de mensenrechtenschendingen, zonder dat ze verwachten dat dit direct veel invloed zal hebben op het regime in Qatar, de FIFA of het wereldkampioenschap zelf. Ook met betrekking tot de verschillende crisissen waarmee we op dit moment worden geconfronteerd worden burgers aangesproken om het juiste te doen: om bijvoorbeeld je kwetsbare oma tijdens de Covid-19 pandemie even niet te bezoeken of om de kachel lager te zetten in de strijd tegen de agressieve en onwettige oorlog van Rusland tegen Oekraïne. In deze zin is het ook voor beleidsmakers van belang om beter inzicht te krijgen in de rol van moraliteit bij het maken van gedragskeuzes. Mensen moreel aanspreken is namelijk een mogelijke en regelmatig gebruikte manier om hen te motiveren tot bepaald sociaal wenselijk gedrag.

Dat morele overwegingen een belangrijke rol spelen in de keuzes die mensen maken is ook in de gedragswetenschappen doorgedrongen. Met name de morele psychologie heeft zich in de laatste decennia gericht op empirisch onderzoek naar moraliteit en moreel gedrag. Een belangrijk onderdeel hiervan is het meten van de individuele moraliteit van mensen. Met individuele moraliteit bedoel ik morele aspecten die aan het individu kunnen worden toegeschreven, zoals iemands morele overtuigingen, oordelen, waarden, normen etc. Deze zijn het object van empirisch onderzoek om deze metingen vervolgens te kunnen gebruiken om bepaalde attitudes en gedrag beter te begrijpen, te verklaren en te voorspellen. Echter, een probleem waar dit onderzoeksveld tegenaan loopt, is dat de verschillende manieren om individuele moraliteit te meten, bijvoorbeeld in termen van een bepaald ontwikkelingsniveau van moreel redeneren (Kohlberg, 1969), of van morele identificatie en motivatie (Aquino & Reed II, 2002), slechte voorspellers blijken te zijn van gedrag. Dit wordt ook wel de *moral judgment-action gap* genoemd.

Relatief recentelijk is het populair geworden om individuele moraliteit te conceptualiseren en empirisch te meten in termen van *algemene morele waarden* –zowel binnen als buiten de morele psychologie. Deze theorieën noem ik “empirische morele

waarde theorieën". De bekendste en meest populaire theorie op dit gebied is de *Moral Foundations Theory* (MFT) (Graham et al., 2013). Deze theorie stelt dat moraliteit bestaat uit een aantal -tenminste vijf- fundamentele morele waarden die bij een individu kunnen worden gemeten door middel van een algemene vragenlijst, de zogeheten *Moral Foundations Questionnaire* (MFQ). Hiermee wordt verondersteld dat hoe belangrijk een individu een bepaalde morele waarde vindt relatief stabiel is en van betekenis is in verschillende situaties. De morele waarden die MFT onderscheidt zijn: zorgzaamheid, eerlijkheid, loyaliteit, autoriteit, en vroomheid[57]. Een ander recent voorbeeld van een dergelijke theorie is de *Morality As Cooperation Theory* (*MAC*) (Curry et al., 2019). Deze theorie onderscheidt net weer andere morele waarden gebaseerd op onderlinge samenwerking.

Tot nu toe worden deze theorieën vooral gebruikt om morele voorkeuren van mensen in kaart te brengen en dit te koppelen aan andere overtuigingen en attitudes, zoals politieke voorkeur. Tegelijkertijd begint een groeiend aantal studies algemene morele waarden mee te nemen in voorspellende modellen om de invloed van moraliteit op een specifiek soort gedrag vast te stellen. Dit heeft echter nog niet geleid tot een meer systematisch en fundamenteel onderzoek naar de relatie tussen algemene morele waarden en gedrag. Het is vooralsnog onduidelijk of het meten van algemene morele waarden een goede manier is om de invloed van individuele moraliteit op gedrag te bestuderen.

Onderzoeksdoel en opzet

In dit proefschrift onderzoek ik de relatie tussen algemene morele waarden en gedrag. Het onderzoeksdoel is om tot een beter begrip te komen van de relatie tussen algemene morele waarden en moreel relevant gedrag op een conceptueel en empirisch niveau. Een tweede, ondergeschikt doel is om methodologische suggesties te formuleren die volgen uit deze analyse om individuele moraliteit en haar invloed op gedrag beter te bestuderen. Hiervoor heb ik vier verschillende studies uitgevoerd. De eerste twee zijn empirisch in combinatie met een grondige conceptuele analyse. De laatste twee zijn fundamentele conceptuele analyses waarin ik inzichten gebruik uit de *fenomenologie*; het filosofische vakgebied dat zich richt op de analyse van het eerste persoonsperspectief van het subject.

Inhoud van het proefschrift

De relatie tussen algemene morele waarden en agressief rijgedrag

Het onderzoek begint in *hoofdstuk 2* met een conceptuele en exploratieve empirische studie naar de relatie tussen algemene morele waarden en agressief auto rijgedrag. Rijgedrag is een alledaagse vorm van gedrag met een duidelijke morele dimensie omdat agressief en risicovol rijgedrag grote negatieve gevolgen kan hebben voor medepassagiers en medeweggebruikers. Op basis daarvan zou verwacht kunnen worden

[57] In het Engels luiden de *moral foundations* voluit: care/ harm, fairness/ cheating, loyalty/ betrayal, authority/ subversion, and sanctity/ degradation (zie MoralFoundations.org (2013))

dat men op basis van iemands algemene morele waarden dit gedrag (mede) kan *voorspellen*.

Echter op basis van mijn conceptuele analyse stel ik vast dat dit onwaarschijnlijk is. Aan de hand van het ethische beslismodel van Rest (1986) (*model of Ethical Decision Making)* laat ik zien dat voordat iemands algemene morele waarde van invloed kan zijn op een gedraging, deze zich eerst op een bepaalde manier in de verschillende stadia van een moreel besluit moet manifesteren. Ieder stadium (te weten: bewust worden van de relevantie van de morele waarde, moreel oordelen, morele intentie, en moreel gedrag) wordt beïnvloed door allerlei individuele en situationele factoren, alsook door de specifieke interpretaties die een rol spelen in een bepaalde beslissituatie. Dit maakt dat óf en wat voor invloed een algemene morele waarde heeft op gedrag contextafhankelijk is en je daardoor slecht op basis van iemands algemene morele waarden rijgedrag kunt voorspellen. Met andere woorden, het maakt de *empirische relatie* tussen algemene morele waarden en agressief rijgedrag onbepaald.

In de *empirische studie* wordt de verwachte zwakke empirische relatie tussen algemene morele waarden en agressief rijgedrag bevestigd. Algemene morele waarden zijn hier geoperationaliseerd door middel van de *Moral Foundation Questionnaire (*MFQ20*)* (MoralFoundations.org, 2013) en agressief rijgedrag doormiddel van de *Aggressive Driving Behavior Scale* (Houston et al., 2002). Op basis van de uitgevoerde meervoudige regressiemodellen vind ik zwakke en veelal niet-significante verbanden tussen de gemeten algemene morele waarden en gerapporteerd agressief rijgedrag. Hieruit wordt geconcludeerd dat iemands algemene morele waarden slechte voorspellers zijn voor iemands rijgedrag.

Een verbreding van het onderzoek naar de relatie tussen algemene morele waarden en gedrag: algemene morele waarden zijn slechte voorspellers van concreet gedrag

In *hoofdstuk 3* wordt het onderzoek verbreed, zowel empirisch als conceptueel. In het empirische deel van dit hoofdstuk, bestaand uit drie verschillende empirische studies, kijk ik naar een breed scala aan moreel relevante gedragingen, zoals het doen van vrijwilligerswerk, het eten van vlees en het zich conformeren aan de afgekondigde coronamaatregelen in Nederland tijdens de eerste *lockdown* van de Covid-19 pandemie. Verder maak ik in dit hoofdstuk gebruik van twee verschillende morele waarden schalen om het gedrag te voorspellen. Naast MFQ, wordt ook de meer recent ontwikkelde *Morality As Cooperation-Questionnaire* (MAC-Q) (Curry et al., 2019) gebruikt. In de studie naar het eten van vlees wordt er ook een meer specifieke morele waarde opgenomen in het model om te zien of waarden die zijn toegespitst op de specifieke context van het gedrag betere voorspellers zijn dan algemene morele waarden.

Daarnaast werk ik de *conceptuele analyse* van de relatie tussen algemene morele waarden en gedrag verder uit en wordt het inzicht in de contextuele afhankelijkheid van deze relatie verdiept. Het wordt bijvoorbeeld duidelijk dat óf en van welke morele waarde iemand zich bewust wordt in een bepaalde situatie sterk afhankelijk is van de interpretatie van die situatie. Ook de afweging die in een bepaalde beslissituatie wordt gemaakt tussen verschillende morele overwegingen onderling en ook tussen morele

overwegingen en andere (meer egoïstische) overwegingen lijkt sterk contextafhankelijk. De invloed van een morele waarde lijkt daarom moeilijk te vangen met het meten van contextonafhankelijke algemene morele waarden. Dit leidt tot een onbepaaldheid van de empirische relatie tussen algemene morele waarden en gedrag die verder lijkt te gaan dan met het toevoegen van andere generieke variabelen (bijvoorbeeld algemene egoïstische waarden) kan worden opgelost. Deze onbepaaldheid ligt namelijk besloten in de *interpretatie* van de algemene morele waarde zelf in een specifieke context. Hoe belangrijk een bepaalde morele waarde binnen een beslissituatie wordt, is sterk afhankelijk van de specifieke betekenis die de waarde binnen een bepaalde context krijgt. Dit is gerelateerd aan het feit dat een algemene morele waarde kan leiden tot verschillende specifieke interpretaties in een beslissituatie en dus ook tot verschillende gedragingen. Op basis van de conceptuele analyse wordt daarom verwacht dat algemeen gemeten morele waarden slechte voorspellers zijn voor concreet moreel gedrag.

In de uitgevoerde *empirische studies* wordt deze verwachting bevestigd. Er worden wederom alleen lage tot zeer lage effecten tussen de algemene morele waarden variabelen en de verschillende soorten gemeten gedragingen gevonden. De opgenomen *specifieke morele waarde*, in de studie naar vlees eten, voorspelt beter dan de algemene morele waarde en verklaart ook het initiële effect van deze laatste weg.

Betekent het feit dat er lage effecten worden gevonden dat moraliteit geen invloed heeft op gedrag? Nee, op basis van de resultaten van de empirische analyse, in combinatie met de contextafhankelijkheid van het morele beslisproces dat naar voren komt in de conceptuele analyse, betoog ik dat individuele moraliteit wel degelijk invloed kan hebben op beslisgedrag, maar dat deze invloed specifieker en dynamischer is dan met algemene morele waarden schalen, zoals MFQ, kan worden uitgedrukt. Meetmethoden van individuele moraliteit moeten met deze contextafhankelijkheid rekening houden door bijvoorbeeld meer specifiek binnen een bepaalde gedragscontext individuele moraliteit te meten (denk aan 'op kantoor', 'op het schoolplein' 'gedrag met betrekking tot de coronamaatregelen' etc.). Deze standpunten worden verder uitgewerkt en onderbouwd in het volgende deel van het proefschrift.

Richting een meer fundamentele conceptuele analyse met behulp van de fenomenologie om tot een beter begrip te komen van bovenstaande resultaten

De bevindingen in hoofdstuk 2 en 3 suggereren dat de relatie tussen iemands morele waarden en gedrag dynamisch van aard is en gekenmerkt wordt door interpretatie. Verschillende morele waarden kunnen belangrijk worden binnen een bepaalde context en één en dezelfde waarde kan leiden tot verschillend gedrag, afhankelijk van de context en de specifieke betekenis die het daarin krijgt. Dit duidt op een morele persoon die als interpreterend subject zijn of haar morele overtuigingen en gedragingen kritisch bevraagt en, zo nodig, wijzigt. Dit gaat in tegen het meer mechanisch causale model van morele waarden dat door empirische morele waarden theorieën wordt verondersteld, door morele waarden te conceptualiseren als psychologische situatie-overschrijdende disposities, die op een contextonafhankelijke manier gemeten kunnen worden en die vervolgens gebruikt kunnen worden als morele variabelen in voorspellende modellen.

De vraag wordt dan hoe we een dergelijke dynamische en interpreterende morele actor moeten begrijpen? En hoe kunnen we spreken over een individuele morele persoon, als dit niet is in termen van algemene situatie-overschrijdende disposities? De vraag naar hoe we de geïdentificeerde dynamische en interpretatieve relatie tussen morele waarden en gedrag kunnen onderbouwen, leidt dus direct naar fundamentele vragen over de morele persoon en, preciezer, naar de fundamentele vraag over morele identiteit of het morele zelf[58]. Oftewel, hoe iemand zichzelf en ook anderen als één en dezelfde morele persoon begrijpt. Om deze vragen te beantwoorden wend ik mij tot een filosofische discipline die zich al langer vanuit het eerste persoonsperspectief met vragen rondom de morele persoon en morele identiteit bezighoudt: de fenomenologie. In dit tweede deel van het proefschrift maak ik gebruik van de inzichten van de filosoof Paul Ricoeur over persoonlijke identiteit en het narratieve zelf. Aan de hand hiervan probeer ik te begrijpen wat de veronderstelde eenheid van de morele persoon constitueert, terwijl deze tegelijkertijd wordt gekenmerkt door een dynamische en interpretatieve relatie met morele waarden.

.

Het narratieve concept van het morele zelf als een verbeterd onderliggend concept voor de empirische studie van moraliteit en gedrag

In *hoofdstuk 4* neem ik bovenstaande vragen op door het impliciete essentialistische concept van morele identiteit van *the Moral Foundations Theory*, bestaande uit situatie-overschrijdende causale disposities, te confronteren met het narratieve concept van morele identiteit van Ricoeur. Ik betoog dat we de contextsensitieve en interpretatieve relatie met onze morele waarden aan de hand van het eerste concept van morele identiteit niet goed kunnen begrijpen, maar aan de hand van het tweede wel.

Ricoeur (1992) maakt duidelijk dat persoonlijke identiteit moet worden onderscheiden in twee polen, die van *ipse* of "zelfheid" (*selfhood)* en die van *idem* of "hetzelfde". *Ipse*-identiteit wordt gekenmerkt door reflexiviteit, dus een relatie van het zelf tot het zelf, en is een antwoord op de vraag "wie?". *Ipse* houdt een onveranderlijke constantheid of permanentie van het zelf in de tijd in, die niet is uit te drukken in termen van een bepaald kenmerk van iemand, maar die is terug te zien in het doen van een belofte: "hoewel de situatie kan veranderen en zelfs mijn kenmerken kunnen veranderen, zal ik mijzelf houden aan deze belofte en zal ik later doen wat ik nu zeg". *Idem*-identiteit is juist de identificatie van een persoon in termen van kenmerken en drukt een persoon uit in termen van "wat". Het karakter van een persoon is hiervan een uitdrukking.

Ik laat in dit hoofdstuk zien dat *Moral Foundations Theory* de identiteit van de morele persoon geheel uitdrukt in termen van *idem*-identiteit –als algemene morele disposities of karaktertrekken- en daarbij het fundamentele reflexieve aspect (*ipse*) van de morele persoon vergeet. Dit is problematisch omdat, volgens Ricoeur, deze reflexiviteit juist inhoudt dat we een verhouding tot ons karakter, bestaande uit onze disposities (zoals onze waarden en gewoonten) innemen. Dit maakt het mogelijk deze

[58] De termen "morele identiteit" en "morele zelf" worden in het proefschrift door elkaar gebruikt.

te veranderen en in situaties opnieuw te interpreteren en toe te passen terwijl deze nog wel kunnen worden begrepen als de uitdrukking van één en dezelfde persoon.

Ricoeur werkt de relatie tussen *idem* en *ipse* en de aspecten van permanentie en veranderlijkheid van de persoon verder uit aan de hand van het narratieve zelf. Volgens Ricoeur identificeren wij onszelf op basis van het levensverhaal dat wij over onszelf vertellen -net zoals we fictieve karakters identificeren en begrijpen aan de hand van het vertelde verhaal. Op dezelfde manier als bij een fictief verhaal, wordt de samenhang van ons levensverhaal en daarmee die van onszelf -onze identiteit- gegeven door de configuratie van het *plot*. Het plot smeedt de verschillende en contingente gebeurtenissen die we tegenkomen en de handelingen die we verrichten samen tot één geheel, namelijk een betekenisvolle verhaallijn en daarmee tot een betekenisvol begrip van onszelf. Onze identiteit of begrip van onszelf, geconstrueerd door ons zelf-narratief, herbergt zo een veelzijdigheid en veranderlijkheid en vormt tegelijkertijd een eenheid en permanentie in de tijd.

De verhalen die we over onszelf vertellen en die samen ons levensverhaal en zelfbegrip vormen zijn nooit ethisch neutraal maar bevatten altijd ideeën over het "goede". Onze verhalen zijn dus altijd "morele verhalen" en geven gestalte aan een morele identiteit. Het is dan ook in deze verhalen, die ons morele zelfbegrip uitdrukken in termen van "wat", dat onze disposities zoals onze (morele) waarden en gewoonten naar voren komen. Onze disposities zijn daardoor niet vastgeroest, maar worden begrepen door middel van verhalen die veranderlijk zijn. Het reflexieve aspect van het zelf (*ipse*) houdt in dat we steeds opnieuw een houding ten opzichte van ons verhaal kunnen aannemen. We kunnen de verhalen over wie we zijn namelijk opnieuw vertellen en interpreteren én ons levensverhaal maakt steeds nieuwe ontwikkelingen door, door middel van de nieuwe gebeurtenissen die we tegenkomen en die we vervolgens in ons levensverhaal opnemen. Onze waarden en gewoonten kunnen daardoor veranderen en steeds opnieuw –en op een andere manier- worden toegepast en begrepen.

De beweging van het plot leidt ook tot een interpretatief, contextsensitief en daarmee dynamisch concept van beslissen op basis van je morele waarden. Door de beweging van het plot waarin nieuwe gebeurtenissen worden opgenomen in het geheel, krijgen deze gebeurtenissen betekenis in het licht van ons levensverhaal en moet tegelijkertijd ons verhaal of zelfbegrip ook steeds opnieuw geïnterpreteerd worden in het licht van de specifieke situatie. Het maken van een beslissing vraagt daarom om een herinterpretatie van het zelf in een bepaalde situatie. Door jezelf in een beslissing te herkennen, kan het als onderdeel van jezelf binnen het plot worden opgenomen. Beslissen op basis van je eigen morele waarden houdt dus geen causale determinatie of automatische toepassing van je morele waarde in maar is een herinterpretatie van jezelf en een herkennen van jezelf in de situatie, in het licht van je eigen morele waarden. De relatie met je morele waarden is daardoor flexibel, aanpasbaar aan de situatie en veranderlijk, terwijl je jezelf toch als één en dezelfde morele persoon herkent.

Dit concept van morele identiteit botst met het impliciete concept van morele identiteit van *Moral Foundations Theory* op basis van algemene vaststaande morele waarden. Het sluit wel aan bij een meer theoretisch paper van de ontwikkelaars van MFT waarin de deugden die volgen uit iemands *moral foundations* flexibeler lijken te

worden voorgesteld. Ik betoog dat als dit als serieus onderdeel van de theorie zou worden opgenomen dit vraagt om een herbezinning op de algemene theorie en de gebruikte meetmethode van individuele moraliteit in termen van algemene situatie-overschrijdende algemene morele waarden. Om een interpretatief en dynamisch moreel zelf tot op zekere hoogte in het meten van individuele moraliteit op te nemen is een exploratieve benadering binnen een bepaalde afgebakende context nodig. Dit kan door een kwalitatieve fase in het onderzoek op te nemen. Een voorbeeld van een methode die hier bij aan kan sluiten is Q-methodologie.

Een verdere validatie: het narratieve concept van het morele zelf in de context van internationale misdaden

In het laatste hoofdstuk voor de conclusie, *hoofdstuk 5*, onderzoek ik de relevantie van het narratieve concept van het morele zelf voor het verklaren en begrijpen van gedrag in de extreme context van internationale misdaden. Hiermee wordt dit concept verder gevalideerd als een belangrijk fundamenteel concept voor het onderzoek naar moreel gedrag.

Internationale misdaden worden in de regel gepleegd binnen een systeem dat wordt gekenmerkt door een geperverteerde moraal. Hierin gaan veelal "gewone" mensen -dat wil zeggen mensen zonder een psychische aandoening- over tot extreem en structureel geweld, door zich te conformeren aan de gewelddadige normen van het systeem en door de bevelen van een autoriteit te gehoorzamen. Deze daden worden binnen de literatuur veelal verklaard vanuit situationele factoren die ervoor zorgen dat gewone mensen een transformatie proces doormaken waarin zij extreme daders worden. Echter, het feit dat niet alle daders hetzelfde zijn en er ook mensen verzet plegen, duidt er op dat individuele disposities ook een rol spelen. Het onderzoek naar welke individuele disposities relevant zijn heeft tot nu toe geen uitsluitsel gegeven, maar het heeft wel geleid tot het idee dat een interactionistische benadering nodig is om dit gedrag te verklaren en begrijpen. Bij deze benadering wordt ervan uitgegaan dat situationele factoren en individuele disposities elkaar wederzijds beïnvloeden.

Ik betoog dat het narratieve concept van het morele zelf van Ricoeur als een goed fundament kan dienen voor een diepgaande interactionistische benadering van dit gedrag. Ik laat zien dat de transformatie van "gewone" mensen in daders primair een *morele* transformatie is waarbij het gebruik van geweld als het "goede" wordt gezien en dat dit een belangrijke aspect is in de verschillende verklaringen van dit gedrag. Deze morele transformatie is op het niveau van de persoon een verandering van iemands morele identiteit, terwijl deze persoon zichzelf nog steeds begrijpt als één en dezelfde morele persoon. Op basis van Ricoeur's narratieve concept van het morele zelf wordt duidelijk dat dit mogelijk is door de interactionistische beweging van het plot waarbij het zelf zichzelf geleidelijk opnieuw interpreteert onder invloed van de situatie. Het opnemen van nieuwe gebeurtenissen en handelingen in iemands levensverhaal -zoals eerste geleidelijke initiaties in geweldpleging- leiden tot een veranderend moreel verhaal en daarmee een veranderend begrip van het morele zelf. Tegelijkertijd is dit individuele morele zelfbegrip ook steeds de basis om nieuwe situaties te interpreteren

en geeft het daarmee de mogelijkheid om een gevraagde handeling niet (meer) te accepteren.

Dit maakt duidelijk dat het transformatieproces een narratieve structuur heeft - die tegelijkertijd zorgt voor veranderlijkheid en permanentie- en dat het zelfbegrip een cruciale rol speelt in het plegen (van verzet tegen) internationale misdaden. Door de rol van het interpreterende zelf kunnen morele disposities wederom niet begrepen worden als vaststaande algemene trekken, maar als individuele aspecten die flexibel en variërend worden begrepen binnen verschillende situaties en die kunnen veranderen. Aan de andere kant, maakt dit begrip van het morele zelf duidelijk dat situationele factoren het individu kunnen beïnvloeden, maar het nooit helemaal weg kunnen vagen. In het nemen van een beslissing verhoudt het individu zich tot de situatie en tot zijn of haar zelfbegrip en is zo altijd het startpunt van een nieuwe handeling.

Uit deze analyse volgt dat om een beter begrip te krijgen van de rol van individuele moraliteit het belangrijk is om het levensverhaal en zelfbegrip van daders in kaart te brengen. Dit kan bijvoorbeeld door een analyse van meer kwalitatieve bronnen zoals ego-documenten.

Conclusies, methodologische implicaties en beleidsimplicaties

De eerste algemene conclusie van het proefschrift is dat algemene morele waarden slechte voorspellers zijn van concreet moreel gedrag. De conceptualisering van individuele moraliteit in termen van algemene morele waarden leidt niet tot het oplossen van de *judgment-moral action gap*. *Ten tweede*, concludeer ik dat een belangrijke verklaring hiervoor is dat de relatie tussen algemene morele waarden en gedrag wordt bemiddeld door het narratieve morele zelf en daardoor een interpretatief karakter heeft en dynamisch is. In plaats van algemene causale morele disposities, moeten algemene morele waarden eerder gezien worden als morele toetsstenen aan de hand waarvan een persoon zichzelf in een situatie herkent en op basis waarvan deze persoon beslissingen maakt.

Op methodologisch niveau betekent de bovenstaande analyse dat de contextspecifieke invloed van moraliteit moeilijk te vatten is door middel van de algemene morele schalen gebruikt door empirisch morele waarden theorieën, zoals bijvoorbeeld de *Moral Foundation Questionnaire*. Om de invloed van morele individualiteit beter te meten moet deze altijd binnen een bepaalde afgebakende gedragscontext worden onderzocht. Verder stel ik dat als we willen weten welke morele overwegingen binnen een bepaalde context een rol spelen, er een kwalitatieve onderzoeksfase moeten worden opgenomen in het onderzoek. Hiermee kan op exploratieve wijze de specifieke morele betekenissen die een rol spelen bij de totstandkoming van gedrag bloot gelegd worden.

Met betrekking tot het formuleren van beleid, impliceren mijn bevindingen dat beleidsmakers moeten inzetten op het beïnvloeden van specifieke morele overwegingen van mensen om specifiek gedrag te kunnen beïnvloeden. Algemene morele waarden als 'respect' kunnen op allerlei manieren worden geïnterpreteerd en het inzetten hierop leidt daardoor niet gemakkelijk tot de gewenste gedragsverandering. Daarnaast kan het

ethische beslismodel van Rest (1986) dienen als een leidraad om met specifieke beleidsmaatregelen gewenst gedrag te stimuleren op basis van de eigen morele motivatie van mensen. Beleidsmakers moeten zich realiseren dat het beïnvloeden van moreel gedrag ook altijd verandering op het niveau van het levensverhaal van mensen en hun morele identiteit betekent. Dit kan aan de ene kant leiden tot weerstand, maar het vormt tegelijkertijd ook een belangrijke sleutel tot duurzame gedragsverandering. Er moet dan ingezet worden op het overbrengen van een nieuwe definitie -of beter- verhaal, van wat het betekent om bijvoorbeeld een "goede boer" of "goede leidinggevende" te zijn. Uiteraard altijd vertaald naar wat dit betekent voor concreet gedrag. Hierbij kan goed gebruik worden gemaakt van (de verhalen van) andere mensen die tot voorbeeld dienen, omdat zij precies de concretisering vormen van bepaalde abstracte idealen en waarden.

Referenties

Aquino, K., & Reed ll, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental Social Psychology, 47*, 55-130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4407236-7.00002-4

Houston, J.M., Harris, P.B. and Norman, M. (2003). The Aggressive Driving Behavior Scale: Developing a Self-Report Measure of Unsafe Driving Practices. *North American Journal of Psychology*, *5*, 269-278.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In Goslin, D. A. (ed.), *Handbook of Socialization Theory and Research (*pp. 347-480). Chicago, IL, Rand McNally.

MoralFoundations.org. (2013). Retrieved May 3, 2020, from https://moralfoundations.org

Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.

Ricoeur, P. (1992). *Oneself as another*. Chicago: The University of Chicago Press.

Summary

Introduction and background to the research

This thesis focuses on gaining a better understanding of the relationship between morality and behaviour. In our daily experience, morality plays an important role in our actions in the sense that something is the “right” or “good” thing to do. Many everyday social interactions in which we do something for others would be difficult to understand if moral considerations did not play a role. Think of: standing up for a less able-bodied person on the train, giving directions to a stranger, or testing yourself with a corona rapid test before going to a party. Morality also plays an important role in our ideas about a better society and in behaviour conducted for this purpose. People eat less meat to reduce animal suffering, and some will fly less to pass on a sustainable earth to future generations. In writing this, a controversial World Cup is being organized in Qatar that some football fans in the Netherlands are not watching on TV because of human rights violations committed there, without any expectation that this will directly have much impact on the regime in Qatar, FIFA, or the World Cup itself. Furthermore, during the various crises we are currently facing, citizens are also called upon to do the right thing: to refrain from visiting their vulnerable grandmother for a while during the Covid-19 pandemic, for instance, or to turn down the stove in the fight against Russia's aggressive and illegal war against Ukraine in 2022. Therefore, it is also important for policymakers to better understand what role morality plays in making behavioural choices. Morally addressing people is one possible -and in fact, regularly used- way to motivate them to certain socially desirable behaviour.

That moral considerations play an important role in the choices people make has also penetrated the behavioural sciences. Moral psychology, in particular, has in recent decades focused on the empirical investigation of morality and moral behaviour. An important part of this is measuring people's individual morality. By individual morality, I mean moral aspects attributed to the individual person, such as one's moral beliefs, judgements, values, norms, etc. These can be the object of empirical research, the measurements of which are used to understand, explain and predict certain attitudes or behaviour. However, one problem that this field of research has encountered is that different ways of measuring individual morality, for instance, in terms of a certain developmental level of moral reasoning (Kohlberg, 1969) or moral identification and motivation (Aquino & Reed II, 2002), turn out to be poor predictors of behaviour. This is also known as the moral judgment-action gap.

A relatively recently popular way of conceptualizing and empirically measuring individual morality -within and outside moral psychology- is in terms of general moral values. I call these theories "empirical moral value theories". The best-known and most popular theory in this area is Moral Foundations Theory (MFT) (Graham et al., 2013). This theory states that morality consists of a set of -at least five- fundamental moral values that can be measured in an individual through a general questionnaire, called the Moral Foundations Questionnaire (MFQ). This assumes that how important an

individual considers a particular moral value is relatively stable and significant across different situations. The moral values that MFT distinguishes are: care/ harm, fairness/ cheating, loyalty/ betrayal, authority/ subversion, and sanctity/ degradation. Another, more recent, example of such a theory is the Morality As Cooperation Theory (MAC) (Curry et al., 2019); this theory distinguishes different moral values based on the idea of morality as a solution to mutual social cooperation.

So far, these theories have been mainly used to identify people's moral preferences and link these to other beliefs and attitudes, such as political affiliation. In addition, a growing number of studies -also outside moral psychology- include general moral values in predictive models to determine the influence of morality on specific types of behaviour. However, this has not yet resulted in more systematic and fundamental research on the relationship between general moral values and behaviour. In this regard, the link between general moral values and actual behaviour remains under researched.

Research objectives and design

In this thesis, I specifically investigate the relationship between general moral values and behaviour. *The research goal is to come to a better understanding of the relationship between general moral values and morally relevant behaviour on a conceptual and empirical level. A secondary aim is to formulate methodological suggestions following from my analysis to better capture individual morality and its influence on behaviour*. To this end, I conducted four different studies. The first two are empirical in combination with in-depth conceptual analysis. The last two are fundamental conceptual analyses in which I use insights from phenomenology, the philosophical discipline that focuses on the analysis of the first-person perspective of the subject.

Content of the thesis

The relationship between general moral values and aggressive driving behaviour
The study begins in *Chapter 2* with an exploratory conceptual and empirical study of the relationship between general moral values and aggressive driving behaviour. Driving behaviour is an everyday form of behaviour with a clear moral dimension because aggressive and risky driving behaviour can have major negative consequences for fellow passengers and fellow road users. On this basis, it might be expected that one could (partly) predict this behaviour based on one's general moral values.

However, based on *my conceptual analysis*, I argue this is unlikely. Using Rest's (1986) Model of Ethical Decision Making, I show that before one's general moral value can influence behaviour, it must first manifest itself in a certain way in the different stages of moral decision-making. Each stage (i.e., moral awareness, moral judgment, moral intention, and moral behaviour) is influenced by all kinds of individual and situational factors as well as by the specific interpretations that play a role in a given decision situation. This means that whether and what influence a general moral value

has on behaviour is context-dependent, and that driving behaviour is hard to predict based on someone's general moral values. In other words, the empirical relationship between general moral values and aggressive driving behaviour is indeterminate.

The empirical study confirms this expected weak empirical relationship between general moral values and aggressive driving. General moral values were operationalized here through the Moral Foundation Questionnaire (MFQ20) (MoralFoundations.org, 2013) and aggressive driving behaviour through the Aggressive Driving Behaviour Scale (Houston et al., 2002). Based on the multiple regression models performed, I find weak and mostly non-significant relationships between measured general moral values and reported aggressive driving behaviour. From this, it is concluded that one's general moral values are poor predictors of one's driving behaviour.

A broadening study of the relationship between general moral values and behaviour: general moral values are poor predictors of concrete behaviour

Chapter 3 broadens the research, both empirically and conceptually. In the empirical part of this chapter, consisting of three different empirical studies, I look at a wide range of morally relevant behaviours, such as volunteering, eating meat, and conforming to the proclaimed corona measures in the Netherlands during the first lockdown of the Covid-19 pandemic. Furthermore, I use two different moral values scales to predict behaviour in this chapter. Next to MFQ, the more recently developed Morality As Cooperation Questionnaire (MAC-Q) (Curry et al., 2019) is also used. In the meat-eating study, a more specific moral value is also included in the model to see whether values tailored to the specific context of behaviour are better predictors than general moral values.

In addition, I elaborate on *the conceptual analysis* of the relationship between general moral values and behaviour and deepen the understanding of the contextual dependence of this relationship. For example, I make it clear that whether and which moral value a person becomes aware of in a given situation depends heavily on the interpretation of that situation. Also, the trade-off made in a given decision situation between different moral considerations among themselves and between moral considerations and other (more selfish) considerations seems to be strongly context-dependent. Therefore, the influence of a moral value seems difficult to capture by measuring context-independent general moral values, and this leads to an indeterminacy of the empirical relationship between general moral values and behaviour that seems to go beyond what can be resolved by adding other general variables (e.g., general egoistic values). Indeed, this indeterminacy lies in the interpretation of the general moral value itself in a specific context. How important a given moral value becomes within a decision situation is highly dependent on the specific meaning the value acquires within a given context, and this is related to the fact that a general moral value can lead to different specific interpretations in a decision situation and, thus, to different behaviours. Based on the conceptual analysis, it is therefore expected that generally measured moral values are poor predictors of concrete moral behaviour.

The empirical studies confirm this expectation. Again, I find only minimal effects between the general moral values variables and the different types of behaviour measured. The included specific moral value in the meat-eating study predicts better than the general moral value and also explains away the initial effect of the latter.

Does finding minimal effects mean that morality has no impact on behaviour? No, based on the results of my empirical analysis, combined with the context-dependence of the moral decision process that emerges in our conceptual analysis, *I argue that individual morality can indeed influence decision behaviour, but that this influence is more specific and dynamic than can be expressed with general moral values scales, such as MFQ*. Measurement methods of individual morality should take into account this context dependence by, for example, measuring individual morality within a delineated context of behaviour (think "in the office", "in the schoolyard", "behaviour related to corona measures," etc.). These views are further developed and substantiated in the next part of the thesis.

Toward a more fundamental conceptual analysis -using phenomenology- in order to reach a better understanding of the relationship between moral values and behaviour

The findings in Chapters 2 and 3 suggest that the relationship between one's moral values and behaviour is characterized by interpretation and is dynamic. Different moral values can become important and each can lead to several different behaviours, depending on the context and the specific meaning the moral value receives within it. This analysis indicates a moral person who, as an interpreting subject, critically questions and, if necessary, modifies their moral beliefs and behaviour. This idea goes against the more mechanistic causal model of moral values assumed by empirical moral value theories by conceptualize moral values as psychological cross-situational dispositions that can be measured in a context-independent way to be used as moral variables in predictive models.

The question then becomes how to understand such a dynamic and interpretive moral person? And how can we talk about an individual moral person if not in terms of general cross-situational dispositions? The question of how to ground the identified dynamic and interpretive relationship between moral values and behaviour thus leads directly to fundamental questions about the moral person and, more precisely, to the fundamental question about *moral identity or the moral self*[59]. In other words, *how a person understands him or herself and others in terms of one and the same moral person*. To answer these questions, I turn to a philosophical discipline that has long been concerned with questions surrounding the moral person and moral identity from the first-person perspective: phenomenology. In this second part of the thesis, I draw on philosopher Paul Ricoeur's insights on personal identity and the narrative self. From there, I try to understand what constitutes the supposed unity of the moral person, while it is at the same time characterised by a dynamic and interpretive relationship with moral values.

[59] In this thesis, we use the terms "moral identity" and "moral self" interchangeably.

The narrative concept of the moral self as an improved underlying concept for the empirical study of morality and behaviour

In *Chapter 4*, I take up the above questions by confronting the Moral Foundations Theory's implicit essentialist concept of moral identity, consisting of cross-situational causal dispositions, with *Ricoeur's narrative concept of moral identity*. I argue that we cannot properly understand the context-sensitive and interpretive relationship with our moral values using the first concept of moral identity, but using the second, we can.

Ricoeur (1992) makes it clear that personal identity must be distinguished into two poles, that of *ipse* or "selfhood" and that of *idem* or "sameness". *Ipse*-identity is characterized by reflexivity, i.e., a relation of self to self, and is an answer to the question "who?". *Ipse* implies an unchanging constancy or permanence of the self in time, which cannot be expressed in terms of a particular characteristic of a person. It is rather the permanency in time that is reflected in making a promise, "although the situation may change and even my characteristics may change, I will keep myself to this promise and will do later what I say now". *Idem*-identity is precisely the identification of a person in terms of characteristics and expresses a person in terms of "what". A person's character is an expression of this.

I show in this chapter that Moral Foundations Theory expresses the identity of the moral person entirely in terms of *idem*-identity -as general moral dispositions or character traits- forgetting the fundamental reflexive aspect (*ipse*) of the moral person. This is problematic because, according to Ricoeur, this reflexivity means precisely that we occupy a relation to our character, consisting of our dispositions (such as our values and habits). This reflexive relationship allows these to be changed and reinterpreted and reapplied in situations, while, at the same time, they can still be understood as the expression of one and the same person.

Ricoeur further elaborates on the relationship between *idem* and *ipse* and the aspects of permanence and mutability of the person through the *narrative self*. According to Ricoeur, we identify ourselves based on *the life story we tell about ourselves* -just as we identify and understand fictional characters based on the story told. In the same way as with a fictional story, the coherence of our life story and, thus, of ourselves -our identity- is given by *the configuration of the plot*. The plot forges together the various and contingent events we encounter and the actions we perform into a single entity, namely a meaningful storyline and thus a meaningful understanding of ourselves. Our identity or understanding of ourselves, constructed by our self-narrative, thus harbours versatility and mutability while at the same time forming a unity and permanence in time.

The stories we tell about ourselves that make up our life story and self-understanding are never ethically neutral but always contain ideas about the "good". Thus, *our stories are always "moral narratives" and shape a moral identity*. In these stories, which also express our moral self-understanding in terms of "what", our dispositions, such as our (moral) values, and habits emerge. Our dispositions are, therefore, not fixed but understood through narratives that are changeable. *The reflexive*

aspect of the self (ipse) implies that we can continually take a stance towards our story. Indeed, we can retell and reinterpret the stories of who we are, and our life story is always going through new developments due to the new events we encounter. Our values and habits can, therefore, change and be applied and understood again and again -and in a different way.

The movement of the plot also leads to an interpretative, context-sensitive and, therefore, dynamic concept of decision-making based on one's moral values. Because of the movement of the plot in which new events are appropriated, these events acquire meaning in light of our life story, while, at the same time, our story or self-understanding must also be reinterpreted again and again in light of the specific situation. Making a decision, therefore, requires a reinterpretation of the self in a particular situation. By recognizing oneself in a decision, it can be appropriated as part of oneself within one's life story. Deciding on the basis of your moral values, therefore, does not involve causal determination or automatic application of your moral value, but is rather a reinterpretation of yourself and a recognition of yourself in the situation in light of your own moral values. The relationship with your moral values is therefore flexible, adaptable to the situation, and changeable while still recognizing yourself as one and the same moral person.

This concept of moral identity clashes with Moral Foundations Theory's implicit concept of moral identity based on generally fixed moral values. It does align with a more theoretical paper by the developers of MFT, in which the virtues that follow from one's moral foundations seem to be presented more flexibly. Following my analysis, I argue that if this is included as a serious part of the theory, it calls for a rethinking of the overall theory and, especially, the measurement method used to capture individual morality. Incorporating the interpretive and dynamic moral self in measuring individual morality requires an exploratory approach within a defined context, and this can be done by including a qualitative phase in the study. An example of a method that can fit this approach is Q-methodology.

A further validation: the narrative concept of the moral self in the context of international crimes

In the last chapter before the conclusion, *Chapter 5*, I explore the relevance of the narrative concept of the moral self for explaining and understanding behaviour in the extreme context of international crimes. This further validates this concept as an important fundamental concept for the study of moral behaviour.

International crimes are generally committed within a system characterized by a perverted morality. In this, mostly "ordinary" people -that is, people without a psychological disorder- engage in extreme and structural violence by conforming to the system's violent norms and obeying an authority's orders. These acts are mostly explained within the literature from situational factors that cause ordinary people to go through a transformative process in which they become extreme perpetrators. However, the fact that not all perpetrators are the same and that people resist indicates that also individual dispositions play a role. Research into which individual dispositions are

relevant has been inconclusive. Still, it has led to the idea that an interactionist approach is needed to explain and understand this behaviour. This approach assumes that situational factors and individual dispositions influence each other mutually.

I argue that Ricoeur's narrative concept of the moral self can serve as a good foundation for an interactionist approach to this behaviour. I show that the transformation from "ordinary" people to perpetrators is primarily a moral transformation in which the use of violence is seen as “good” and plays an important role in the various explanations of this behaviour. At the level of the person, this moral transformation entails a change in one’s moral identity, while this person still understands himself as one and the same moral person. Based on Ricoeur’s narrative concept of the moral self, it becomes clear that this is possible through the interactionist movement of the plot, where the self gradually reinterprets itself under the influence of the situation. The appropriation of new events and actions -such as the first gradual initiations into violence- leads to a changing moral narrative and, thus, a changing understanding of the moral self. At the same time, this individual moral self-understanding is also always the basis for interpreting new situations .It ,thus, gives rise to the possibility of not (or no longer) accepting a requested action.

This makes it clear that the transformation process has a narrative structure - which simultaneously ensures changeability and permanence- and that self-understanding plays a crucial role in committing and resisting international crimes. The interpretive nature of the self again allows moral dispositions to be understood not as fixed general traits, but as individual aspects that are understood flexibly and variably within different situations and that can change. On the other hand, this understanding of the moral self makes it clear that situational factors can influence the individual, but can never completely erase it. In making a decision, the individual relates to the situation and to his or her self-understanding and thus is always the starting point of a new action.

It follows from this analysis that to better understand the role of individual morality, it is important to map the life story and self-understanding of perpetrators. This can be done, for example, by analyzing more qualitative sources such as ego-documents.

Conclusions, methodological implications and policy implications

The first general conclusion of this thesis is that general moral values are poor predictors of concrete moral behaviour. Conceptualizing individual morality in terms of general moral values does not solve the judgment-moral action gap. *Second*, I conclude that an important explanation for this is that the relationship between general moral values and behaviour is mediated by the narrative moral self and is therefore interpretive in nature and dynamic. Rather than general causal moral dispositions, *general moral values should rather be seen as moral touchstones* against which a person recognizes him- or herself in a situation and on the basis of which this person makes decisions.

On a methodological level, the above analysis means that the context-specific influence of morality is difficult to capture through the general moral scales used by

empirical moral values theories, think for example of the Moral Foundation Questionnaire. To better measure the influence of moral individuality, it should always be examined within a particular delimited behavioural context. Furthermore, I argue that if we want to know which moral considerations play a role within a given context, a qualitative research phase should be included in the study. This can expose in an exploratory manner the specific moral meanings that play a role in shaping behaviour.

With regard to policy formulation, my findings imply that policymakers should focus on influencing people's specific moral considerations in order to influence specific behaviour. General moral values such as “respect” can be interpreted in many ways and targeting these will not easily lead to the desired behavioural change. In addition, Rest's (1986) model of Ethical Decision Making can serve as a guide to develop specific policies to encourage desired behaviour based on people's own moral motivation. Policymakers should realize that influencing moral behaviour always also means endeavouring for change at the level of people's life story and moral identity. Such endeavouring may lead to resistance on the one hand, but at the same time it also forms an important key to sustainable behavioural change. Efforts should then be made to convey a new definition -or better story- of what it means to be, for instance, a "good farmer" or "good manager". Of course, always translated into what this means for concrete behaviour. Making use of (the stories of) other people as examples can be a valuable tool here. Such exemplars are namely precisely the concretisation of certain abstract ideals and values.

References

Aquino, K., & Reed ll, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of ‘morality-as-cooperation’ with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental Social Psychology, 47*, 55-130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4407236-7.00002-4

Houston, J.M., Harris, P.B. and Norman, M. (2003). The Aggressive Driving Behavior Scale: Developing a Self-Report Measure of Unsafe Driving Practices. *North American Journal of Psychology*, *5*, 269-278.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In Goslin, D. A. (ed.), *Handbook of Socialization Theory and Research (*pp. 347-480). Chicago, IL, Rand McNally.

MoralFoundations.org. (2013). Retrieved February 16, 2018, from https://moralfoundations.org

Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.

Ricoeur, P. (1992). *Oneself as another*. Chicago: The University of Chicago Press.

Curriculum Vitae

13 oktober 1986 Born in Beverwijk, The Netherlands

Research and Teaching

2022-present **Assistant Professor**
Department of Criminology, Faculty of Law
Vrije Universiteit, Amsterdam, The Netherlands

2017-2023 **Ph.D. researcher**
Faculty of Technology, Policy, and Management
Delft University of Technology, The Netherlands

2016-2017 **Junior Researcher**
Nederlands Studiecentrum Criminaliteit en Rechtshandhaving
Amsterdam, The Netherlands

2013-2014/ **Junior Lecturer**
2015-2016 Department of Criminology, Faculty of Law
Vrije Universiteit, Amsterdam, The Netherlands

2009-2013 **Student and Research Assistant**
Master International Crimes and Criminology
Department of Criminology, Faculty of Law
Vrije Universiteit, Amsterdam, The Netherlands

Education

2011-2016 **MA, Philosophy of Law and Governance**
Department of Philosophy, Faculty of Humanities
Vrije Universiteit, Amsterdam, The Netherlands

2009-2014 **MSc., International Crimes and Criminology** (*cum laude*)
Department of Criminology, Faculty of Law
Vrije Universiteit, Amsterdam, The Netherlands

2006-2009 **BSc., Criminology**
Department of Criminology, Faculty of Law
Vrije Universiteit, Amsterdam, The Netherlands

1999-2005 **Gymnasium**
Gymnasium Felisenum
Velsen-Zuid, The Netherlands

Activities

2017-2018 **Supervising 2nd year bachelor students**
Faculty of Technology, Policy, and Management
Delft University of Technology, The Netherlands

List of Publications

Only publications that are related to the Ph.D. research will be mentioned here

Journal Publications

1. van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2020). Does morality predict aggressive driving? A conceptual analysis and exploratory empirical investigation. *Transportation research part F: traffic psychology and behaviour*, *74*, 259-271.

2. van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2022). Why are general moral values poor predictors of concrete moral behavior in everyday life? A conceptual analysis and empirical study. *Frontiers in Psychology*, 3834.

3. van den Berg, T. G. C. & L. D. A. Corrias (in press). Moral Foundations Theory and the Narrative Self: Towards an Improved Concept of Moral Selfhood for the Empirical Study of Morality. *Phenomenology and the Cognitive Sciences*.

4. Szep, T., van den Berg, T. G. C., Cointe, N., Daniel, A. M., Martinho, A., Tang, T., & Chorus, C. (2023).Give and take: Moral aspects of travelers' intentions to participate in a hypothetical established social routing scheme. *Cities*, *133*, 104132.

Main Conference Presentations

1. van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2019). Does Morality Predict Aggressive Driving? An Empirical Investigation among Dutch Drivers. Symposium of the European Association for Research in Transportation (Budapest, Hungary), September 2019.

2. van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2019). Putting moral foundation theory to the behavioral test. (presented by co-author M. Kroesen). International Choice Modeling Conference (Kobe, Japan), August, 2019.

3. van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2019). Een empirisch onderzoek naar het verband tussen morele waarden en agressief rijgedrag. Colloquium Vervoersplanologisch Speurwerk (Leuven, Belgium), November 2019.

4. van den Berg, T. G. C. & L. D. A. Corrias (2021). Moral Foundations Theory and the Narrative Self: Towards an Improved Concept of Moral Selfhood for the Empirical Study of Morality. OZSW Conference. (Virtual), December, 2021.

5. van den Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2022). Why General Moral Values do not Predict Specific Moral Behaviour in Real Life (Poster). Annual Meeting of the Society for Judgment and Decision Making. (Virtual), February 2022.

Awards

The paper 'Why are general moral values poor predictors of concrete moral behavior in everyday life?' was nominated for the best ESS Ph.D. paper award 2023.

Acknowledgements

It sometimes still seems a bit surreal that I have finished a Ph.D. at the Delft University of Technology. Before coming to Delft I never thought I would be able to fit pursuing and finishing a doctorate at a technical university into my narrative identity. But it did fit -talking of "discordant concordances"- attesting to the extraordinary capabilities of our narratives.

The first two persons I want to thank for this accomplishment are my promotors. From the beginning you made me feel in place, in particular through the open and vivid discussions we had -on and off topic-, and your sincere curiosity and interest. Caspar, thank you for always moving me forward when I was lingering and, most of all, for standing right behind and in front of me when I needed it. Maarten, you've been an inspiring exemplar of how to combine academic rigorousness with pragmatism –I hope to achieve that one day. Thank you for always freeing time for me, no matter what and for constantly radiating that things will be all right. I am also very grateful to both of you for giving space for and having faith in my philosophical aspirations in this project.

Luigi, this is where you come in. I can't express how happy I am that you were my co-promotor. I could not have done this research project without you and working with you on this subject made that I've enjoyed the project until the very end. Your phenomenological teachings will last me a life-time and so will the many wise insights you've given me during the many long conversations we've had. Thank you for all your time, encouragements, and your inexhaustible kindness.

Then, I want to thank my BEHAVE-partners in crime: Teo, Tanzhe, Andreia, Nicolas, and Aemiro. You are the reason that I did not only feel in place, but also at home at the TU. Thanks for all the fun in the BEHAVE-room and being such nice friends. I dearly hope for life long reunions! In addition, I want to thank my other TLO and ESS-colleagues for the nice lunches, meetings, casual conversations and pieces of advice you've given me during my time in Delft.

I am also grateful to other academics. First of all, to Alette Smeulers and Joris van Wijk who started me on the academic road and guided me during my first steps. My gratitude also goes out to the editors, anonymous reviewers and receptive audiences, whose critical questions and comments have helped me a lot to improve my work. And, of course, I want to gratefully thank my doctoral committee members for reviewing and commenting on this thesis.

Then to my friends, Koerskers, Droelers, Prinsjes, VU buddy's and beyond, you are the best bunch, thank you for being there and for so much fun. You make that live is life nanananana, and above all, of course, that it is gezellug gezellug. Many of you have mustered the patience to listen to me rambling on about some "great insight" at a late night hour and I owe many of you for shaping my thoughts and beliefs over the years,

of which this thesis is of course a reflection, thanks for this! With regard to the thesis process I want to thank Freek, my fellow Ph.D'er, who gave me "working-from-home-shelter" when a 1-month-old was screaming the whole apartment down. Also, I want to thank Jappie, my study buddy and friend with whom I have shared the process of academic development since we founded the VPC-club. Furthermore, I thank Bellaar and Annika for their critical insights on the thesis and helping me to prepare for the defence.

A very special thanks goes to my paranymph and dear friend San-li. There hasn't been one conversation with you that did not leave me more cheerful than before (while most of the time I already feel quite cheerful!). Thanks for being such a faithful companion.

Finally, a heartful thanks to my family. First, to my unofficial in-laws, Kris, Wies, Martijn, Rietje, and Ruud, for your love and support for me, Liza and the kids. I couldn't have wished for a greater family not-marrying into. Ruud and Rietje, thank you so much for always jumping in whenever it is needed, I appreciate that a lot.

Then, to my dear brother (and paranymph) and sister, Bas and Brigitte, both of you are an inspiration and a touchstone, and just the coolest siblings one can imagine. If I wake up one morning to find that we have indeed conquered the world, it would not surprise me.

Who I cannot possibly thank enough are my parents. You've given me the best start of my life imaginable. I realize how extremely lucky I am to have parents who have given me and continue to give me their endless unconditional support and love. During corona and in the final phases of my thesis this was underlined again by all your help. Thanks a million, you are the best.

Lastly, my own little family, for the Ph.D. wasn't the only ambitious project that started about 5 years ago. Lou, Robin and Lex, mijn lieve spookies, you are my daily triple reminder that there are more important things in life. Liza, my love, raising a family together with you and being with you and our boys and girl gives me so much joy that I did not know this was possible. Needless to say, your belief in me, your support in the many ways that you do and your beautiful imperative eyes were invaluable for getting this finished. Thank you for being my beacon and for being my girl.