



Kent Academic Repository

**Jansen, Claudia (2018) *Moral objectivity: Kant, Hume and psychopathy.*
Master of Philosophy (MPhil) thesis, University of Kent,.**

Downloaded from

<https://kar.kent.ac.uk/66548/> The University of Kent's Academic Repository KAR

The version of record is available from

This document version

UNSPECIFIED

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

**Moral objectivity:
Kant, Hume and psychopathy**

Claudia Jansen

Thesis submitted for the degree of Doctor of Philosophy

Department of philosophy

University of Kent

2014

Abstract

Moral objectivity: Kant, Hume and psychopathy

Moral objectivity is about genuinely better or worse courses of action and states of affairs in the moral domain. It seems good to aim at an identification of objective moral justifications that is maximally independent of subjectivity (at least if the threat of relativism is to be avoided). Having said that, it seems problematic to accept objective discriminations or justifications that are devoid of subjectivity. Every account of objective moral justifications seems in need of some sort of relationship with naturalistic human minds. How else could such justifications enter the universe?

In this study I build towards arguments for deciding when claims about the status of moral objectivity are overambitious. I offer three lines of argument that point to moral objectivity being essentially anti-realist and (as such) mind-dependent. The first is grounded in Hume's (exclusively psychological) conception of 'reason'. It is paradigmatically well illustrated by Kant's philosophy.

The second and third lines of argument are grounded in research about the nature and etiology of psychopathy. The second is about conceptual relativity regarding normative judgements about good practical lives. The third is about libertarian freedom over innately given components, components crucial to the psychological possibility of taking account of others in evaluative decision-making. Due to conceptual and empirical problems about (possible worlds of) human nature, which will be laid out, these two lines of argument need further conceptual and empirical attention.

Additional to my constructive theory about the limits of moral objectivity, my study contains a critical reflection on methodological aspects of the contemporary meta-ethical debate. Overall, my study is a critical call for better reflection on the concept 'reason' and a deeper involvement with theoretical claims about human nature.

Table of Contents

<i>Acknowledgements</i>	6
<i>Notes to Kant and Hume’s work: references and translations</i>	7
<i>Introduction</i>	13
 Chapter 1 ‘An empirically informed account of psychopathy’	
1. Introduction.....	23
2. The popular conception of psychopathy.....	26
3. The history of psychopathy.....	28
4. Psychopathy as a distinct cluster of personality traits and behaviours.....	30
5. ‘Factor models’.....	43
6. The prevalence of psychopathy.....	45
7. More scientific approaches to psychopathy.....	57
8. Psychopathy in women.....	79
9. The etiology of psychopathy.....	82
10. Looking ahead.....	91

Chapter 2 ‘David Hume on reason and morality’

1. Introduction.....	93
----------------------	----

2. Hume’s philosophical project.....	95
3. Hume’s views on ‘reason’.....	101
4. Reason and mathematics.....	118
5. ‘Matters of fact’ and evaluative distinctions.....	124
6. ‘Reason’, ‘truth’ and contradiction.....	129
7. Humean passions.....	135
8. Hume’s view on morality.....	147
9. Further elements of Hume’s moral philosophy.....	156
10. Hume’s views on ‘belief’.....	163
11. Hume on human nature.....	186

Chapter 3 ‘On the risk of misreading Hume for contemporary meta-ethics’

1. Introduction.....	193
2. Hume on reason and belief: recalling some essential thoughts.....	194
3. Meta-ethical neo-Humeanism.....	197
4. On the relationship between neo-Humeanism and Hume.....	199
5. Intellectual drawbacks and hindrances as a result of misreading Hume.....	203

Chapter 4 ‘Immanuel Kant’s moral philosophy’

1. Introduction.....	219
2. An introduction to Kant’s work.....	221
3. Kant’s account of the noumenon.....	223
4. On the relationship between morality and ‘the empirical’ in Kant’s work.....	230

5. 'Vernunft'	240
6. The <i>Formula of Universal Law</i>	244
7. <i>FUL</i> , Kant and the embodied nature of the rational judge.....	248
8. A comprehensive understanding of <i>FUL</i>	252
9. Kant on conscience and other aspects of the mind of the scoundrel.....	259
10. Looking ahead.....	263

Chapter 5 'Building towards three anti-realist criteria'

1. Introduction.....	264
2. Hume (and Kant): the first criterion.....	267
3. Recalling psychopathy.....	273
4. Pre-intentional emotions.....	275
5. Easy and hard cases of understanding another person.....	276
6. Bio-psychological conditions for not taking the interests of others into account.	281
7. <i>Intermezzo: FUL</i> : not completely empirically free?.....	289
8. A shared argumentative building block for a second and third criterion.....	291

In closing: summary study and extension of second and third line of argument..... 293

Bibliography..... 313

Acknowledgements

I would like to thank my supervisor Simon Kirchin for all his support. Without him this study would have been a much poorer piece of work. I am also grateful to my secondary supervisor David Corfield for his comments on my work and to the University of Kent for their scholarship that enabled me to write this PhD.

I also owe many thanks to Jan Bransen, Phyllis Illari and Federica Russo for their continuous extensive support. Furthermore, I would like to express my gratitude towards Bob Brecher, David Cooke, Michael Franz, Jeanette Kennett, Jesse Prinz, Paul Russell, Eric Schliesser, and Kenneth Westphal for the help they have given me. I have also benefited from intellectual correspondence with Simon Blackburn, Rachel Cohon, Bob Hare, Kerrin Jacobs, Mark Nelson, Russ Shafer-Landau, Michael Smith and Jacqueline Taylor.

My final words of gratitude are for those who have supported me in less intellectual ways. Special thanks go to my parents as well as to Alvise, Angelos, Federica, Mary, Phyllis, Sue, Xiaoran and Wissia.

Notes to Kant and Hume's work: references and translations

Hume

Notes on citations

In this study, I have made use of the Selby-Bigge and Nidditch second edition of *A Treatise of Human Nature*. I have used the Selby-Bigge and Nidditch third edition of *An Enquiry concerning Human Understanding* and *An Enquiry concerning the Principles of Morals*. Full references are as follows:

Hume, D. *A Treatise of Human Nature* [2nd edition, 1978, edited, with an analytical index, by L.A. Selby-Bigge. Text revised and notes by P.H. Nidditch]. Oxford: Clarendon Press.

Hume, D. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals* [3d edition, 1975, edited with introduction, comparative table of contents, and analytical index by L.A. Selby-Bigge. Text revised and notes by P.H. Nidditch]. Oxford: Clarendon Press.

In the running main text, I generally use the abbreviation *Treatise* for *A Treatise of Human Nature* rather than the full title. In the running main text, I generally use the abbreviation *EPM* for *An Enquiry concerning the Principles of Morals* and *EHU* for Hume's book *An Enquiry concerning Human Understanding*. In the running main text, I use the full name of a particular work by Hume if and when I am concerned with other works by Hume.

When it comes to typical in-text referencing (for paraphrasing and quotations), in the main text, I use ‘T’, ‘EPM’ and ‘EHU’ for the works mentioned above. An example would be (T: 267) to refer to page 267 of the *Treatise*.

The following shortened titles and abbreviations are used for in-text referencing to refer to the following works by Hume:

CL: ‘Of civil liberty’

DP: ‘Dissertation on the passions’

DT: ‘Of the delicacy of taste and passion’

IN: ‘Of interest’

LG: ‘Letter from a gentleman’

MS: ‘Of the middle station of life’

DNR: *Dialogues Concerning Natural Religion*

OC: ‘Of the original contract’

RP: ‘Of the rise and progress of the arts and sciences’

SC: ‘The sceptic’

SE: ‘Of superstition and enthusiasm’

ST: ‘Of the standard of taste’

The abovementioned works can all be found on the scholarly website www.davidhume.org, a website I have used in this study for investigation of the abovementioned works.

Typically I have kept the original references of *secondary* literature, sometimes slightly adjusted for consistency in referencing style. Unless otherwise noted, there are no major discrepancies between original references to primary work of Hume as used in secondary literature and my presentation of the secondary literature with its appeal to primary work.

Kant

Notes on citations

Quotations from Kant's works are cited in the body of the text by abbreviation deduced from the name of the original German work (see below for a list of abbreviations, e.g. 'G' for Kant's *Grundlegung zur Metaphysik der Sitten* (*Groundwork of the Metaphysics of Morals*)) and page number. An example would be (G: 394). The page numbers are those as in *Kants gesammelte Schriften*, edited by the Royal Prussian (later German, then Berlin-Brandenburg) Academy of Sciences (Berlin: Georg Reimer, later Walter de Gruyter, 1900–). In terms of general use of abbreviations in in-text referencing, an exception is made for references to the *Critique of Pure Reason*, which are cited by the customary use of the pagination of its first (A) and second (B) editions. An electronic version of the Academy editions can be found on <http://www.korpora.org/Kant/> (which I have generally used with additional help from hard copy books).

The following German shortened titles and abbreviations are used to refer to specific works by Kant.

Anfang: 'Mutmaßlicher Anfang der Menschengeschichte' (Conjectural beginning of human history), 8¹: 107–23².

Ende: 'Das Ende aller Dinge' (The end of all things), 8: 325–39.

G: *Grundlegung zur Metaphysik der Sitten* (*Groundwork of the Metaphysics of Morals*), 4: 385–463.

Idee: 'Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht' (Idea for a

¹ This number is the volume number that reflects one of the 29 volumes of the Academy edition and indicates here in which one of those the text at issue can be found. In terms of the text at issue, Kant's text 'Anfang' can be found in volume 8 of the Academy edition.

² These indicate the page numbers of the text in a particular volume.

universal history with a cosmopolitan aim), 8: 15–31.

KpV: *Kritik der praktischen Vernunft (Critique of Practical Reason)*, 5: 1–163.

KrV: *Kritik der reinen Vernunft (Critique of Pure Reason)*, 3 (B / 2. Auflage): 1–553;
4 (A / 1. Auflage): 1–238.

KU: *Kritik der Urteilskraft (Critique of the Power of Judgment)*, 5:
165–485.

MAN: *Metaphysische Anfangsgrunde der Naturwissenschaft (Metaphysical
Foundations of Natural Science)*, 4: 465–565.

MdS: *Metaphysik der Sitten (Metaphysics of Morals)*, 6: 203–493.

ProI: *Prolegomena zu einer jeder kunftigen Metaphysik die als Wissenschaft
wird auftreten können (Prolegomena to Any Future Metaphysics That
Will Be Able to Come Forward as Science)*, 4: 253–383.

Rel: *Die Religion innerhalb der Grenzen der bloßen Vernunft (Religion within
the Boundaries of Mere Reason)*, 6: 1–202.

In the main body of the text, lecture transcriptions (that are also part of the Academy edition) are referred to by the name of the transcriber (e.g., *Collins*) followed by volume and page number. References to handwritten notes by Kant are indicated by the compound German term ‘Handschriftlicher Nachlass’ plus volume number and page number of the Academy edition.

I use, with occasional modifications, the following English translations of these major works by Kant:

G: Paton, H.J. (1964). *Immanuel Kant. Groundwork of the Metaphysics of Morals*. New York: Harper & Row.

MdS: Ladd, J. (1965). *The Metaphysical Elements of Justice*. Indianapolis, New York, Kansas

- City: The Bobbs-Merrill Company. [first part of *The Metaphysics of Morals*]
- Gregor, M.J. (1964). *Immanuel Kant. The Doctrine of Virtue*. New York: Harper & Row. [second part of *The Metaphysics of Morals*]
- KpV: White Beck, L. (1949). *Critique of Practical Reason*. Chicago: The University of Chicago Press.
- KrV: Kemp-Smith, N. (1933). *Immanuel Kant's Critique of Pure Reason*. Hongkong: The MacMillan Press LTD.
- KU: Guyer, P. & Wood, A. (general editors); Guyer, P. (specific volume editor and translator) and Matthews, E. (specific volume translator) (2000). *Critique of the Power of Judgement. The Cambridge Edition of the Works of Immanuel Kant*. Cambridge: Cambridge University Press.

Translations from Kant's lectures on ethics come from:

Guyer, P. & Wood, A. (general editors); Heath, P. (specific volume editor and translator) and Schneewind, J.B (specific volume editor) (1997). *Lectures on Ethics. The Cambridge Edition of the Works of Immanuel Kant* (general editors Paul Guyer and Allen W. Wood; Cambridge: Cambridge University Press.³

All other translations are my own.

References to *quotations* from translated works have the following form: (G-Paton: 91 / G: 424). First the translated work is given together with the page number where the referenced sentence or passage can be found in the translated work. Subsequently, the page number of the sentence or passage as in the original German academy edition is given. NB: for

³ I am grateful to Robert Louden for having inspired me as to the above notes on translations of and referencing to Kant's work. I rely heavily on Louden's notes as in his (2000) and (2011). I have adapted them to reflect translations of Kant's work and references to Kant's work in this study.

paraphrased sentences and passages I typically provide 'just' the page number(s) to the original German Academy edition. The traditional Academy volume and page numbers (and also the A and B pagination from the *Critique of Pure Reason*) are reprinted in the margins of most contemporary editions and translations of Kant's writings.

Introduction:

Moral objectivity: Kant, Hume and psychopathy

Intellectual questions about ‘moral objectivity’ concern the existence, non-existence and nature of genuinely better and worse courses of action, states of affairs, and the like. In contemporary meta-ethics, debates about moral objectivity are frequently debates about realism and anti-realism, indeed *between* realists and anti-realists. Moral anti-realists stress the dependence of the ordinary world on our minds. They hold that morality is not built into or otherwise essentially part of the fabric of the world. That being an essential commitment of all anti-realists, anti-realists vary in their view as to whether or not and if so how the essential anti-realist commitment just expressed is compatible with the existence of genuinely better and worse courses of action, i.e. ‘moral objectivity’.

As for the position ‘moral realism’, as Stephen Finlay (2007) notes, the contemporary debate over ‘moral realism’, a century after it was launched by G. E. Moore’s *Principia Ethica*, is a tangled and bewildering web. As he explains:

This is largely due to dramatic differences in what philosophers assume it is about....A pivotal problem is the lack of consensus over what ‘realism’ should mean in the context of ethics; we shall see that the variety of metaethical claims labeled ‘realist’ cannot be collectively characterized any less vaguely than as holding that ‘morality’, in some form, has some kind or other of independence from people’s attitudes or practices. We look in vain for a reference for ‘morality’ and a kind of attitude-independence common throughout the debate. (Finlay, 2007)

I will come back to moral realism soon. Let me first briefly say something about realism about the external world.

According to Alex Miller, the everyday world of (a) macroscopic objects and (b) their properties is plausibly thought of as having two general aspects. One concerns an *existence* claim: “[t]ables, rocks, the moon and so on, all exist, as do the following facts: the table’s

being square; the rock's being made of granite, and the moon's being spherical and yellow.” (Miller, 2009: introduction). The other aspect concerns a claim about *independence*: the fact that the rock is being made of granite and that the moon exists and is spherical is independent of anything people say or think about the matter.

Miller suggests that realism as a generic position (covering different positions about different subject matters) takes the following form:

a, *b*, and *c* [the distinctive objects of a particular subject matter] and so on exist, and the fact that they exist and have properties such as *F-ness*, *G-ness*, and *H-ness* is (apart from mundane empirical dependencies of the sort sometimes encountered in everyday life) independent of anyone's beliefs, linguistic practices, conceptual schemes, and so on. (Miller, 2009; introduction)

According to Miller, there are at least two ways in which a non-realist can reject the existence dimension of realism about a particular subject matter. The first way rejects the existence dimension by rejecting the claim that the distinctive objects of that subject matter exist. The second way admits that those objects exist, but denies that they instantiate any of the properties distinctive of that subject matter. The first can be illustrated by means of Field's error-theory about arithmetic and the second via Mackie's error-theory about morality. Briefly, according to Field's error-theory of arithmetic, the objects distinctive of arithmetic do not exist, and it is this which leads to the rejection of the existence dimension of arithmetical realism, at least as platonically conceived. Mackie on the other hand proposes an error-theoretic account of morals, not because there are no objects or entities that could form the subject matter of ethics (it is no part of Mackie's project to deny the existence of persons and their actions and so on), but because it is implausible to suppose that the sorts of properties that moral properties would have to be are ever instantiated in the world (Mackie, 1971 Chapter 1) (see Miller, 2009). An alternative, seemingly compatible way of reading Mackie would be to say that according to Mackie moral values have a conceptual nature – that is, they are necessarily and essentially both objective and prescriptive – that cannot be

found anywhere in the world. What moral values amount to conceptually speaking is not part of the fabric of the world; there is no place in the universe where this concept of moral values is instantiated.

Let me now return to *moral* realism. I have already said that in meta-ethics moral realist views come in a bewildering and tangled variety. There are forms of moral realism which claim that moral reality transcends all forms of mind-dependence. But *descriptively* speaking (i.e. looking at what actually exists in the meta-ethical debate) a commitment to that idea is not a feature of all moral realist theories. And it *might* also not be a *normatively* necessary condition. Further to the *descriptive* issue, there is a wide variety of actual realist views in meta-ethics and sometimes these views are really very different. Therefore it is difficult to say what moral realism generically amounts to *descriptively*. And because of the fact that many self-ascribed moral realist views seem *prima facie* entitled to describe their view as a realist view, it is difficult to say what moral realism –as a general position– *normatively* amounts to.

Returning again to the *descriptive* issue, it is beyond doubt that there are large differences between some self-ascribed moral realist views, but there are also some features moral realist theories share. First, there is the obvious feature of making a claim about the existence of moral reality (facts, values, properties). Moral reality *somehow* exists. Furthermore, it seems to be the case that nearly all characterizations of contemporary forms of moral realism include some version of the following two core claims:

- (i) Ethical discourse is *assertoric* and *descriptive*; ethical claims purport to state ethical facts by attributing ethical properties to people, actions, institutions, etc. and are thus true or false depending on whether their descriptions of things are accurate or not (and similarly, the ethical beliefs expressed by such claims are true or false depending on whether their representations of things are accurate or not).

- (ii) At least some ethical claims, when literally construed, are *true* in the above sense.
(Fitzpatrick, 2009)¹

Often there is a focus on semantic and other language issues in meta-ethics when people advocate or oppose a theory about moral realism. The question however is whether, strictly speaking, there can be a legitimate relationship between semantic and other language matters and debate about the correctness of moral realism. There seems to be something plausible about Devitt's remark that "Realism says nothing semantic at all beyond...making the negative claim point that our semantic capacities do not constitute the world." (Devitt, 1991: 39², cited in Miller, 2009: § 6).³

My aim in this study is to offer argumentative evidence for the view that moral objectivity, if it exists, is essentially non-realist. I offer three different sorts of argumentative contributions for the view that there is an essential non-realist component to the nature of morality. One of them is based in Hume's philosophy, and well-illustrated by Kant's philosophy. The other two are based in arguments about human nature. These in turn arise from the science of psychopathy, which in my study partly functions as a negative test case: reflections on psychopathy lead to conclusions about the nature of typical practical agents as well as the nature of the moral judgements they make.

¹ One plausible and standard way to make distinctions in between the class of moral realist theories as they exist in meta-ethics is by distinguishing mind-*independent* from mind-*dependent* views. Another way is to distinguish between non-naturalist and naturalist forms of realism. As Cuneo correctly argues, "according to the 'standard definition', moral naturalism is the view that there are things that display moral properties and that all such properties are natural. Moral nonnaturalism, by contrast, is the position that there are things that have moral properties, at least some of which are not natural" (Cuneo, 2007: 851-2). Cuneo argues that he follows Nicholas Sturgeon (2006) here. 'Moral Naturalism'. In Copp, D. (ed.) *The Oxford Handbook of Ethical Theory* (Oxford: Oxford UP), pp. 91–121. As Cuneo notes, "clearly, if the standard definition is to be informative, we need some understanding of what it is for a property to be 'natural'". There are multiple interpretations here in meta-ethics.

² Devitt, M. (1991). *Realism and Truth*. Princeton: Princeton University Press.

³ I point out that Devitt is not specifically pointing to *moral* realism.

Having just mentioned that I intend to offer criteria for the view that moral objectivity, if it exists, is essentially non-realist, this study is *not* best understood as a polemical attack on moral realism. Rather, this study is best seen as a project in which I want to embark on a secure philosophical pathway by identifying criteria by means of which preposterous theoretical claims about the objective nature of moral objectivity can be avoided. Because this study is not best seen as a polemical attack on realism, I prefer to use the term '*non*-realist' rather than '*anti*-realist' for the argumentative criteria I offer. What comes on top of the just mentioned reason for preferring the term 'non-realist' over 'anti-realist' is that at least *prima facie* there is reason to think that not anything goes in the moral domain; that there are somehow genuinely better and worse courses of action. I take it that the idea that there are genuinely better and worse courses of action runs the risk of being overshadowed by the polemical term 'anti'.

As indicated, in this study I want to make a contribution to a framework of non-realist boundaries within which there might well be genuine standards of moral rightness and wrongness. Having said that, it is *beyond the purposes of this study to provide a substantive positive meta-ethical theory* of the *nature* of moral knowledge/moral objectivity. That is, it is beyond the purposes of this study to give an account of what in a deeply *meta-ethical* theoretical way it means that there are better and worse courses of action in the moral domain. Such an account may or may not have a relationship with *first-order normative theories*. In any case, offering any first-order normative theories is also beyond the purposes of this study.

So, in this study I aim for theoretical criteria that make clear that there are limits to an often proclaimed realist nature of moral objectivity. How should one go about with such a project? Where should one start one's investigations? Many will assume that *if* there are any limits to the realist nature of moral objectivity, then these limits have something to do with the relationship between sentiments and moral evaluations. Suppose that is true, then the question is what type of affective involvement matters, why and how exactly. Will any type

of emotional outburst of any sort of (human adult) agent (resulting in any moral evaluation) do? Why, or why not? What is so essential and crucial about the identified relevant types of sentiment? Are they perhaps a reflection of some type of lack of freedom we have, where this type of freedom is crucial as a constraint on the hardness of moral objectivity? If not that, what then is it that is so important about these sentiments?

I have just supposed that sentiments are the cause of some constraints on the hardness of objectivity. But perhaps it is rather the case that the bounds of moral objectivity come from there being limitations as to the power of an *intellectual* faculty called 'reason' (compare some normative institution called 'reason'). Or perhaps there is an issue with *both* sentiments and an intellectual faculty of reason when it comes to constraints on the hardness of objectivity. If both, are the two somehow two sides of the same limiting coin or do they both constrain in an independent way?

My investigations as to the boundaries of moral objectivity proceeds largely through a dissection of Kant's and Hume's conception of 'reason'. Textual investigation will deliver us a Kantian conception of reason ('Vernunft') that in fact is a complex union of two different Kantian sub-conceptions of 'Vernunft': something reflecting a capacity of the soul and something normative (that is, *ratio subiectiue sumta* and *ratio obiectiue sumta*). Textual investigation will also reveal Hume's conception of 'reason', a fully psychological-intellectual conception that is very different from Kant's conception(s). It is an identification of important differences between Kant's and Hume's conceptions of 'reason' ('Vernunft' for Kant) and these conceptions' relationship with Kant's a priori *justification* of practical reason, i.e. the *Formula of Universal Law (FUL)* that is illustrative for conclusions about the limits of moral objectivity. 'Illustrative' here means that strictly speaking a fine-grained dissection of Hume's conception of reason will suffice to give us one criterion for there being a non-realist component to moral objectivity. This first criterion however is helpfully comprehended when presented in the context of an interaction with *FUL*.

FUL plays an important role in my study as an object of investigation. What characterizes this Kantian moral principle is that it probably is the most ‘empirical averse’, including ‘sentiment-averse’, principle available in moral philosophy. The other side of the coin is that *FUL* is an *a priori* principle. What is possibly more controversial but nevertheless seems striking about *FUL* is that it has a scent of being a mind-independent one. Having said that, at the same time, *FUL* seems to be a mind-*dependent* principle in the sense of it being grounded in an intellectual reasoning procedure. The union of all these features makes it that *FUL* is an ultimate challenge to as well as a paradigmatic object of investigation for my account about the limits of moral objectivity. Because of its alleged mind-*independent* aspect, *FUL* is an ultimate challenge in the sense of it being an ultimate object to surpass. Namely, mind-independence of the principle would mean that there is no theoretical dependence on whatever form of subjectivity. This rules out limiting influences on the hardness of moral objectivity, including those paradigmatic candidates for such limiting influences called ‘sentiments’. *FUL* is a paradigmatic object of investigation rather than challenge, because of the thought that it is mind-dependent. Because of this, *FUL* at face value seems to keep itself in touch with the empirical world, which seems a desirable thing. (NB: my investigation will show that Kant’s philosophy has clear problems meeting this apparently desirable thing.) As to a connection with the empirical world, a starting point for my investigation as to the limits of moral objectivity is a *prima facie* agreement with Simon Blackburn’s view that moral discriminations that deserve a status of objectivity in one way or another should find room for ethics in the natural order we inhabit, and of which we are part (Blackburn, 1998: 49). The only alternatives to this plausible idea seem to be some sort of theological account of moral objectivity or else an appeal to fully mind-independent factual entities. The latter run a huge risk of being obscure entities. The former face the old and famous Euthyphro dilemma.

This study is set-up as follows. In chapter 1, I start with a discussion of psychopathy; largely an empirical discussion, but one accompanied by philosophical reflections. In my

study, psychopaths partly serve as a negative test case. If and when they do so, that means that the focus is on psychopaths to discover something about *us* typical practical agents involved in the moral domain through thought and deeds. By focusing on psychopaths' *atypical* affective and conative actual mental life and predispositions for these mental states, the goal is to learn something about *typical* human beings' capacities and predispositions that in our daily life play a role in the acknowledgement of moral discriminations. If and when psychopaths do *not* function as a negative test case, then that means that their presence (their existence) in the world is in some way directly important to the identification of criteria for the boundaries of moral objectivity.

In chapter 2, I dissect Hume's philosophy about reason, belief, passion and morality. Through this, I arrive at one important ingredient for my own philosophical theory as to the limits of moral objectivity: Hume's conception of 'reason'. Having said that, there is a further upshot of my study of Hume's work. Attention for Hume's work also provides us the substantive arguments for the claim that the *neo-Humean Split* and the *Inertia of Belief Thesis* that are essential to contemporary meta-ethical neo-Humeanism are unsupported when one looks at Hume's texts.⁴ Note that Hume's *Inertia of reason* thesis, is of crucial importance as to my conclusions about the limits of moral objectivity, while the *Inertia of belief* thesis is not.

Chapter 3 is an intermezzo chapter. In it, I discuss certain aspects of the methodological culture of meta-ethics, one of which concerns a bad reading of certain aspects of Hume's work, especially Hume's concept of belief. This chapter is different from the others in the sense that it does *not* make a constructive contribution to my argument about the limits of moral objectivity. Rather, in this chapter I explore unfruitful consequences of misreading certain aspects of Hume's work –especially his views of belief – for the contemporary meta-ethical debate *at large* with its core attention for the problems of

⁴ By the *neo-Humean Split* I mean neo-Humeanism's sharp and essential distinction between beliefs and desires. The *Inertia of Belief thesis* reflects neo-Humeanism's characteristic claim that beliefs have no intrinsic motivational force.

the nature and status of moral objectivity; the nature and status of moral judgements and the nature of moral motivation. Through this discussion other less fortunate aspects of the methodological culture of meta-ethics will reveal themselves. Aspects that seem to obstruct a more fruitful collective intellectual pathway to a solution of the three core meta-ethical objectives I have just mentioned.

In chapter 4, I dissect Kant's practical philosophy. This dissection paradigmatically includes a dissection of Kant's concept of 'Vernunft' as well as a thorough discussion of the first version of Kant's categorical imperative: *FUL*. Importantly (from the perspective of the aim of my study), I connect Kant's concept of Vernunft to *FUL* as a moral principle in the way I believe Kant wanted them to be connected.

In the final chapter, chapter 5, I present the first criterion based in Hume's philosophy. I also present a shared argumentative basis for two further theoretical criteria for there being an anti-realist component to moral objectivity. In the general conclusion I split this shared argumentative basis into two different argumentative branches for two further criteria for there being a non-realist component to moral objectivity.

This study will focus a lot on issues of human nature. My own contribution to that comes mainly in the form of reflections on the nature and etiology of psychopathy (but I also discuss Kant's and Hume's views of human nature). My study is meant to be a call for more engagement with theoretical claims about human nature. I intend to make clear that deeper engagement is important to solve the problem of moral objectivity. Indeed, it seems the case that such an engagement is first necessary to *comprehend* the meta-ethical intellectual problem of moral objectivity in a sufficiently deep sense, a sense that leads to the deeply satisfactory philosophical conclusions about the matter. By means of this study, I want to make clear that more and deeper engagement with issues about human nature should be an integral component of meta-ethical discussion.⁵

⁵ I believe that the following remark by Ian Shapiro about human nature and *political* philosophy is one meta-ethicists should start taking seriously for their discipline.

“Every political philosophy takes for granted a view of human nature, and every view of human nature is controversial. Some political philosophers have taken the view that human nature is an immutable given, others that it is shaped (in varying degrees) by culture and circumstance. Differences about the basic attitudes of human beings toward one another – whether selfish, altruistic or some combination – have also exercised political philosophers. Although none of these questions has been settled definitively, various advances have been made in thinking systematically about them.” (Shapiro, 1998)

Chapter 1: Psychopathy

1. Introduction

Human Nature is the only science of man; and yet has been hitherto the most neglected. 'T will be sufficient for me, if I can bring it into a little more fashion.

(Hume, T: 273)

Let us suppose such a person ever so selfish; let private interest have engrossed every so much his attention; yet in instances, where that is not concerned, he must unavoidably feel some propensity to the good of mankind, and make it an object of choice, if everything else be equal. Would any man, who is walking along, tread as willingly on another's gouty toes, whom he has no quarrel with, as on the hard flint and pavement?

(Hume, EPM: 226)

No man is entirely without moral feeling, for were he completely lacking in capacity for it he would be morally dead. And if ... the moral life-force could no longer excite this feeling, then humanity would dissolve ... into mere animality and be mixed irrevocably with the mass of other natural beings.

(Kant, MM: 398, MM-Gregor: 60)¹

¹ In Gregor's perception of the academy version, the original German passage is to be found on page 399. In my perception it is on page 398.

How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it...That we often derive sorrow of others, is of a matter of fact too obvious to require any instances to prove it; for this sentiment, like all the other original passions of human nature, is by no means confined to the virtuous and humane, though they perhaps may feel it with the most exquisite sensibility. The greatest ruffian, the most hardened violator of the laws of society, is not altogether without it.

(Smith, *The Theory of Moral Sentiments*: 13)

Though psychopaths know, in some sense, what it means to wrong people, to act immorally, this kind of judgment has for them no motivational component at all. They do not care about others or their duties to them, have no concern for others' rights and feelings, do not accept responsibility, and do not know what it is like to defer one's own gratifications out of respect for the dignity of another human being. Quite significantly, they feel no guilt, regret, shame, or remorse (though they may superficially fake these feelings) when they have engaged in harmful conduct. They are paradigms of individuals whom Kant would call "morally dead". (Murphy, 1972: 286-7)

In this chapter, I start my project of investigating the limits of moral objectivity by means of a focus on psychopathy. This chapter on psychopathy should give us information about what it is that makes it *impossible* to sincerely take the interest of others into account in thought. This information then will be a basis for the second and third line of reasoning that speak in favour of their being a non-realist component to moral objectivity.

I proceed as follows. In § 2, I point out two conceptions of psychopathy that are held by a rather large number of people unfamiliar with the current diagnostic criteria of psychopathy. I explain why they are out of line with the dominant scientific opinion on psychopathy.

In § 3, I give a brief introduction into the history of (research on) psychopathy. This will help us understand contemporary constructs of psychopathy and current measurement tools used to diagnose psychopathy.

In § 4, I discuss the dominant academic conception of psychopathy: psychopathy as a distinct cluster of behaviour and inferred personality traits (as advocated and developed by Robert Hare²). I also discuss the Psychopathy Checklist-Revised (the PCL-R) which is currently the gold standard tool (used by clinicians) for measuring psychopathy. Importantly, I explain the PCL-R's relationship with Hare's conception of psychopathy just mentioned. At the end of § 4 I present some reflections on psychopathy and the concept 'disorder'.

In § 5, I focus on the statistical concept/method called 'factor analysis'. I include an explanation of the relationship between this widely applied statistical method and (on the other hand) data coming from the use of measurement instruments (such as the PCL-R). We need to understand something (elementary) about how factor analysis works to comprehend one ground on basis of which psychopathy researchers might make certain conceptual claims about the nature of psychopathy.

In § 6, I focus on prevalence rates of psychopathy. I first discuss the general prevalence rate of psychopathy and then proceed with prevalence rates for a variety of sub-populations (e.g. the forensic population and 'the household population'). I express my doubts about the dominant view that the general prevalence rate of psychopathy amounts to 1% of the human population. In this section I also explain something about the relationship and differences between psychopathy and anti-social personality disorder (ASPD). This will help us to critically reflect on the claimed prevalence rates of psychopathy.

In § 7, I leave Hare's widely shared conception of psychopathy behind for a while and switch to a discussion of five further conceptions of psychopathy. These are: (i) psychopathy as an extreme variation of normal personality traits; (ii) psychopathy as a neurological deficit;

² Note that there *do exist* other theories in the field that could be grouped under the abstract heading 'psychopathy as a distinct cluster of behaviour and inferred personality traits'. Note furthermore that I believe that one (say, as a philosopher) can also *think* of other reasonable conceptual substantive variants that deviate in some conceptually reasonable way (in order to be reasonable they probably cannot deviate 'too much'), from Hare's conception.

(iii) psychopathy as a genetic deficit; (iv) psychopathy as an inhibition problem and (v) psychopathy as an adaptive reproductive strategy. These conceptions reflect different descriptions or at least different emphases of empirical researchers as to the nature of empirical psychopathy. Many of them are *not* mutually exclusive and most of them are also compatible with psychopathy as a personality syndrome as the first conception presented.

In § 8, I present some reflections and research on psychopathy in females. Currently research on psychopathy in females is still quite in its infancy, but there is some.

In § 9, I focus on the etiology of psychopathy. This section is especially important for the theory about the bounds of moral objectivity.

In the final § 10, I look ahead by drawing some connections between the current chapter and upcoming considerations in other chapters.

2. The popular conception of psychopathy

Let me now start with presenting the phenomenon of psychopathy. James Blair, Derek Mitchell and Karina Blair open their *The Psychopath: Emotion and the Brain* in the following way:

Humans have long been concerned by or fascinated with the concept of evil and the people thought to personify evil. Say the word psychopaths and most people can easily conjure up an image of someone they believe to embody the word. Some may think of characters from movies: Hannibal Lecter from *The Silence of the Lambs*, Mr Blonde from *Reservoir Dogs*, Norman Bates from *Psycho*, and Freddy Krueger from *A Nightmare on Elm Street*. Others may gain inspiration from the world of politics and claim that Adolf Hitler, Saddam Hussein, Margaret Thatcher, George W. Bush or even Bill Clinton is psychopathic. Yet more may consider their current employer or ex-partner to be the ultimate psychopaths. (Blair, Mitchell & Blair, 2005: 1)

And at the very beginning of his seminal work *Without Conscience: the Disturbing World of the Psychopaths among us*, psychopathy pioneer Robert Hare tells us:

When I agreed to write this book I knew it would be difficult to present hard scientific data and circumspection in a way that the public could understand. I would have been quite comfortable remaining in my ivory tower, having esoteric discussions with other researchers and writing technical books and articles. However, in recent years, there has been a dramatic upsurge in the public's exposure to the machinations and depredations of psychopaths. The news media are filled with dramatic accounts of violent crime, financial scandals, and violations of the public trust. Countless movies and books tell the stories of serial killers, con artists, and members of organized crime. Although many of these accounts and portrayals are of psychopaths, many others are not, and this important distinction is often lost on the news media, the entertainment industry, and the public... Th[e] failure to distinguish clearly between offenders who are psychopaths and those who are not has dire consequences for society, as this book makes clear. (Hare, 1993: xi-xii)

Taking these quotations in a wider empirical context (i.e. both a clinical context and a research context), what these two quotations represent is the widespread view of experts that lay people's understanding of what psychopaths are is suboptimal or even profoundly wrong. The dominant view amongst empirical experts is that the construct of psychopathy amounts to a certain distinct cluster of personality traits and behaviours, including e.g. a lack of empathy and extremely manipulative behaviour. While experts see these traits as traits that paradigmatically characterize psychopaths, they hold that only under certain strict clinical criteria (to which we come in § 4) the presence of this cluster of traits and behaviours reflects psychopathy. So according to these researchers, the absence of empathy and the presence of manipulative behaviour we might very reasonably ascribe to our boss or ex-partner rarely - very rarely - means that our boss or partner is a psychopath. Not everyone who presents with certain traits and behaviour that paradigmatically characterizes

psychopaths is a psychopath. As mentioned, more information about the clinical criteria for psychopathy follows later.

Another thing that needs to be said from an academic perspective and in response to the popular conception of psychopathy is that most scientists hold that the cluster of traits and behaviours that characterize psychopaths *need not include extremely violent behaviour* as paradigmatically shown by movie characters such as Hannibal Lecter.³ Having mentioned that scientists typically hold that the class of psychopaths does not exclusively consist of extremely violent people, it is also important to note that another typical view amongst experts is that *not all violent criminals are psychopaths*. This is so because not all violent criminals present with a distinct cluster of traits and behaviour that function as criteria for a diagnosis of psychopathy.

3. The history of psychopathy

It may very well be the case that there have been psychopaths since the beginnings of human history. That could be explained by psychopathy being an adaptive evolutionary strategy (more on that in § 7.6 below). In any case, any affirmative or negative answer to that issue is dependent on what the construct of psychopathy amounts to. And that might be a topic that is never beyond reasonable conceptual discussion.

When it comes to the *existence of actual psychopaths*, suggestions from Hervey Cleckley lead us back to around 400 BC to the Greek statesman and general Alcibiades as a human being who might really have embodied (the right use of) the word ‘psychopath’ (Cleckley, 1988: 325-336). When it comes to *early conceptualizations* we might have one in a

³ Note however that psychopaths commit more than twice as many violent and aggressive acts, both in and out of prison, as do other criminals (Hare, 1993: 88).

conceptualization given by Theophrastus in 280 BC when he describes ‘the unscrupulous man’:

The man without moral feeling is the kind who will take an oath with no sense of responsibility....By nature he is a base kind of person, lacking the most elementary sense of decency and capable of absolutely nothing. He leaves his mother without support in her old age....knows the inside of the town jail better than his own house....In court, he is capable of playing any role: defendant, plaintiff, or witness. He knows a good many rascals. (see Babiak & Hare 2006: 320)

It seems best to say that as an academic field of research psychopathy (only) really developed from 1941 onwards. In 1941 Hervey Cleckley published the first edition of his book *The Mask of Sanity*, a pioneering work on psychopathy. From that point onwards academics became heavily interested in studying psychopathy (see Vitale & Newman, 2001). Cleckley was a psychiatrist. Cleckley’s book offered the first truly comprehensive clinical descriptions of human beings that Cleckley identified as very different from his other patients; human beings Cleckley himself labelled as psychopaths and human beings who would now be really good candidates for a diagnosis of psychopathy according to what now are dominant clinical criteria, criteria that have been heavily influenced by Cleckley’s work. Cleckley found the following traits in this special group of his patients, or ‘clients’ as clinicians now typically call them.

- superficial charm and good intelligence
- absence of delusions and other irrational thinking
- absence of ‘nervousness’ or other psychoneurotic manifestations
- unreliability
- untruthfulness and insincerity
- lack of remorse or shame

- inadequately motivated anti-social behaviour
- poor judgment and failure to learn by experience
- pathologic egocentricity and incapacity for love
- general poverty in major affective reactions
- specific loss of insight
- unresponsiveness in general interpersonal relations
- fantastic and uninviting behaviour with drink and sometimes without
- suicide rarely carried out
- sex life impersonal, trivial, and poorly integrated
- failure to follow any life plan⁴

4. Psychopathy as a distinct cluster of personality traits and behaviours

4.1. An introduction

In the decades after the publication of Cleckley's psychopathy criteria, academics have used these criteria to make reliable measurement tools for the clinical assessment of psychopathy (Vitale & Newman, 2001). This counts first and foremost for psychopathy expert Robert Hare. Hare has developed a conception of psychopathy that is now the dominant one in the academic and clinical field. According to Hare, psychopathy is a distinctive cluster of behaviours and inferred personality traits (1993: ix; cf Hare & Neumann, 2008: 222). The candidate character traits are all summed up in the most frequently used diagnostic tool for the measurement of psychopathy, the *Psychopathy Checklist-R- Revised (PCL-R)*, developed by Hare himself. The PCL-R contains 20 items. These are:

⁴ These are the traits (16 in total) as they can be found in the fifth and final edition of *The Mask of Sanity* (see 1988, esp. 338-9). The initial (1941) version of the book contained 21 features of the clinical profile of psychopaths. By the time of the publication of the fifth version, Cleckley had removed a few items, split one item into two and combined a few items (see Hare & Neumann, 2008: 224-5).

- Glib/superficial charm
- Grandiose sense of self-worth
- Pathological lying
- Conning/manipulative
- Lack of remorse or guilt
- Shallow affect
- Callous/lack of empathy
- Failure to accept responsibility for own actions
- Need for stimulation/proneness to boredom
- Parasitic lifestyle
- Poor behavioural controls
- Early behavioural problems
- Lack of realistic long-term goals
- Impulsivity
- Irresponsibility
- Juvenile delinquency
- Revocation of conditional release
- Promiscuous sexual behaviour
- Many short-term marital affairs
- Criminal versatility

The clinician scores those items by means of a semi-structured interview and additionally case-history information. Each item is scored '0', '1' or '2' according to the extent to which it applies to the individual. The total score can therefore range from 0-40. The cut-off point for psychopathy is 30 (Hare, 1996; Edens, 2001). The *PCL-R* is considered to be the most valid and reliable measurement tool for measuring psychopathy in adults (Fulero, 1995). Casual

inspection of the published literature shows that since 2001, more than 50 articles (not including those articles from the royalties-receiver of the PCL-R: Robert Hare), describe it as the 'gold standard' (Hare & Neumann, 2010).

As mentioned, besides a semi-structured interview, there is a case-history part of the assessment. This part contributes to the item scores by looking at the subject's background, including such things as work and educational history. It is very important to adopt this case-study part - that does not rely on the words of the potential psychopath - as part of the assessment, because psychopaths lie frequently and often do it rather well; it is not uncommon for even experienced interviewers to be tricked by psychopaths in the sense of coming to believe their untrue and insincere stories (see Hare, 1993:124).⁵

In her (2009) Mette Kreis argues that the PCL-R has shaped current conceptualization of psychopathy to the point where the measure has become synonymous with the construct (Kreis, 2009: 30). This claim needs some attention.

In their (2010a and 2010b), Jennifer Skeem and David Cooke argue that the psychopathy construct should never be conflated with its measurement tool (s) (*cf* Kreis, 2009: 30). Skeem and Cooke also make the specific point that the psychopathy construct as advocated by Hare and the PCL-R are intertwined in a vicious way. Hare and his colleague Craig Neumann agree with the first general point made by Skeem and Cooke, but disagree with the latter specific accusation as to the relationship between the PCL-R and the

⁵ There are two psychopathy measurement tools that can be considered as close family members of the PCL-R. Both are derived from the PCL-R. One of them, the PCL-R: Screening Version (SV) is in fact a somewhat shorter version of the PCL-R. Since the PCL-R:SV can be completed in the absence of criminal record information, it is more appropriate than the PCL-R for use in non-forensic settings and is particularly well-suited for use in civil psychiatric evaluations and studies of community samples (Ullrich, S. Farrington, D.P., Coid, J.W. (2008). 'Psychopathic personality traits and life success'. *Personality and Individual Differences*, 44, pp. 1162-1171). The PCL-R: SV has 12 items. The cut off score for psychopathy as measured by the PCL-R: SV is 18 (out of 24). (Babiak & Hare, 2006: 28; Babiak, Neumann, Hare, 2010).

The second derivative of the PCL-R is the PCL-R-YV. This measurement tool is just as the PCL-R based on a semi-structured interview and collateral information. And just as the PCL-R it consists of 20 items designed to measure interpersonal, affective and behavioural features related to psychopathic traits (see Hare, 1996; see also Finger, Marsh, Mitchell, Reid, Budhani, Kosson, 2006). The most significant difference is that the PCL-R-YV (youth version) is not designed for adults; it is meant to measure psychopathic traits in youth, aged 12-18 (see e.g. Forth, 2005).

construct that accompanies it (Hare & Neumann, 2010). I think Hare and Neumann are right in that the construct of psychopathy as (first and foremost) embraced by Hare and the PCL-R as a measurement tool are fundamentally distinct. Reflection on the construct of psychopathy - heavily informed by Cleckley's conclusions- has preceded the development of the PCL-R. Furthermore, as far as I can judge, Hare has always been aware of how important it is to distinguish the construct from the measurement tool. Having said that, I believe that the empirical field of psychopathy currently stands still and has been standing still for quite a while in terms of conceptual reflection on the construct *because of the dominance of the PCL-R*. And since the PCL-R construct (rather than the measurement instrument) seems to stand still too in terms of critical reflection since its initial establishment, Kreis' claim that "the PCL-R has shaped current conceptualization of psychopathy to the point where the measure has become synonymous with the construct" (Kreis, 2009: 31) seems to contain a significant bit of truth.

4.2. Is psychopathy a 'disorder'?

Rather often clinicians and empirical researchers describe psychopathy as a 'disorder'. This is a striking and rather awkward claim since clinicians and empirical researchers generally do *not* regard psychopaths as personifying a normative sense of abnormality. Or, a bit more careful, psychopaths are considered to be normatively normal in at least the following sense: in contrast to people who present with certain other 'personality disorders', psychopaths are *not* subject to delusions or other states are a straightforward candidate for the catchy term 'crazy' or 'mad' or 'insane'. Robert Hare states:

Psychopaths are not disoriented or out of touch with reality, nor do they experience the delusions, hallucinations, or intense subjective distress that characterize most other mental

disorders. Unlike psychotic individuals, psychopaths are rational and aware of what they are doing and why. (Hare, 1993: 22)

And Jeffrey Murphy claims:

[U]nlike the psychotic, the psychopath seems to suffer from no obvious cognitive or volitional impairments.⁶ He knows what he is doing (he has no delusions); and, since he typically does just what he wants to do, it would be odd to call him compulsive or to claim that he acts on irresistible impulses. Thus, he is by no means clearly 'insane' by currently accepted medical or legal standards. (Murphy, 1972: 285)

As Hare implicitly argues, a psychopath who has killed someone has not done that because of orders received from a Martian in a spaceship or something like that (see Hare, 1993: 22). Psychopaths are not crazy in that sense and therefore (I argue) possibly not crazy at all. The fact that psychopaths are capable of controlling their behaviour (note that they are aware of the potential consequences of their acts) also speaks against seeing them as being somehow 'mad', 'insane', or 'crazy' (see Hare, 1993: 143).

The question we are left with then is why it is empirical researchers and clinical practitioners so frequently call psychopathy a 'disorder' despite them not being inclined to describe psychopaths as 'insane', 'mad' or 'crazy'? There are two straightforward candidate *explanations* (rather than justifications) for this. The two might very well be related. First, the frequent use of the word 'disorder' might have to be explained by the fact that although psychopathy is not in the current version (*DSM-V*) of the bible for clinicians, the *Diagnostic and Statistical Manual of Mental Disorders*, it has been in the *DSM* in the past. In these past times, psychopathy, in the *DSM*, was included in the group of 'personality disorders (in § 6 I explain why psychopathy is not in the *DSM* now)⁷

⁶ On the absence of cognitive defects, see e.g. also Cleckley (1988) and Mealey (1995).

⁷ One of the more recent versions of the *DSM*, the *DSM-IV-TR* ('text revision') tells us the following about the compound term 'mental disorder':

Secondly, the current use of words might (also) have to be explained by the fact that typically clinical psychological and psychiatric labels refer to mental conditions that *much more uncontroversially*⁸ could be called ‘disorders’, because typically diagnostic labels of

“Although this volume is titled the *Diagnostic and Statistical Manual of Mental Disorders*, the term *mental disorder* unfortunately implies a distinction between ‘mental’ disorders and ‘physical’ disorders that is a reductionistic anachronism of mind/body dualism. A compelling literature documents that there is much ‘physical’ in ‘mental’ disorders and much ‘mental’ in ‘physical’ disorders. The term raised by the term ‘mental’ disorders has been much clearer than its solution, and, unfortunately, the term persists in the title of DSM-IV because we have not found an appropriate substitute.

Moreover, although this manual provides a classification of mental disorders, it must be admitted that no definition adequately specifies precise boundaries for the concept of ‘mental disorder’. The concept of mental disorder, like many other concepts in medicine and science, lacks a consistent operational definition that covers all situations. All medical conditions are defined on various levels of abstraction—for example, structural pathology (e.g. ulcerative colitis), symptom presentation (e.g. migraine), deviance from a physical norm (e.g. hypertension), and etiology (e.g., pneumococcal pneumonia). Mental disorders have also been defined by a variety of concepts (e.g. distress, dysfunction, dyscontrol, disadvantage, disability, inflexibility, irrationality, syndromal pattern, etiology, and statistical deviation). Each is a useful indicator for a mental disorder, but none is equivalent to the concept, and different situations call for different definitions.

Despite these caveats, the definition of *mental disorder* that was included in DSM-III and DSM-III-R is presented here because it is as useful as any other available condition and has helped to guide decisions regarding which conditions on the boundary between normality and pathology should be included in DSM-IV. In DSM-IV, each of the mental disorders is conceptualized as a clinically significant behavioral or psychological syndrome or pattern that occurs in an individual and that is associated with present distress (e.g., a painful symptom) or disability (i.e., impairment in one or more important areas of functioning) or with a significantly increased risk of suffering death, pain, disability or an important loss of freedom. In addition, this syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example, the death of a loved one. Whatever its original cause, it must currently be considered a manifestation of a behavioral, psychological, or biological dysfunction in the individual. Neither deviant behavior (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above.

A common misconception is that a classification of mental disorders classifies people, when actually what are being classified are disorders that people have. For this reason, the text of DSM-IV (as did the text of DSM-III-R) avoids the use of such expressions as ‘a schizophrenic’ or ‘an alcoholic’ and instead uses the more accurate, but admittedly more cumbersome, ‘an individual with Schizophrenia’ or ‘an individual with Alcohol Dependence.’ (DSM-IV-TR, 2000: xxxi).

That much about the term ‘*mental disorder*’. The *DSM-IV-TR* states the following about the concept ‘*personality disorder*’ as used in that work:

“A Personality Disorder is an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual’s culture, is pervasive and inflexible, has an onset in adolescence or early adulthood, is stable over time, and leads to distress or impairment.” (DSM-IV-TR, 2000: 685)

⁸ But thereby not uncontroversially; one only needs to look at cross-cultural variability in terms of evaluations of certain personality traits and behaviour to start wondering whether some at face value normatively abnormal trait, particular bit of human behavior or cluster of traits and/or behaviours can really be called ‘defective’ or can be regarded as reflecting a ‘disorder’. Having said that, there also seem to be some clear cross-cultural convergence in terms of what is seen as mental illness. See Murphy (1976) for an interesting and informative read.

mental illness are applied to psychological conditions that in one or more ways interfere with the well-being of the individual to whom the diagnostic label is ascribed. Psychopathy by contrast, typically is not a phenomenon where the personality traits of the agent have a significant negative impact on his (or her) well-being. As to 'well-being', the mental condition/character of psychopaths typically does not cause them any mental distress (recall the quotation from Hare above). Apart from that, the character traits of a psychopath also have striking potential to be really advantageous to him. Take the trait 'manipulative'. This trait allows the psychopaths to fulfill his (or her) desires for e.g. money and power very well. Having said this, it must be said that regularly psychopaths' character puts them in a disadvantageous position in terms of their well-being by inviting for behaviour that results in having to live one's life temporarily or forever in prison rather than in the pleasant outside world.

4.3. Different styles

We have seen that Hare sees psychopathy as a distinct cluster of personality traits and behaviours. Hare and his colleague Paul Babiak have argued that this distinct cluster of traits and behaviours can present itself in three different guises, or *styles*. All of these three styles are for an important part characterized by 'a-typical' (or, as often said 'deficient') affective experience, i.e. a lack of empathy, a lack of guilt and remorse; and shallow emotions), but they differ somewhat on other dimensions. Below, I briefly discuss the three styles.

The 'classic style' consists of those who besides them being highly a-typical in affective experience, score high as well on three other dimensions that can be used to characterize the personality and behaviour of psychopaths. That is, classic style psychopaths also score high on (a) the dimension *lifestyle*, for example represented by a need for stimulation/proneness to boredom and a parasitic attitude. Furthermore (b) these

psychopaths score high on the dimension *anti-social*, represented for example by poor behaviour controls and juvenile delinquency. And finally (c) they score high on the dimension *interpersonal*, represented for example by glibness and demonstration of superficial charm.

The ‘manipulative style’ group of psychopaths is made up by those with a high score on the interpersonal and affective dimensions, and somewhat lower scores on the lifestyle and antisocial dimensions. Rather than being very impulsive and violent, they manipulate, deceive and charm actively to reach self-gratification at the cost of other –mostly human-beings.

Finally, there is the ‘macho style’ (Babiak & Hare, 2006: 185-186). Psychopaths who represent this style score high on the affective, lifestyle and antisocial dimension, but low on the interpersonal dimension. These psychopaths are aggressive, bullying, and abrasive beings. They are less charming and manipulative than the other types.

4.4. One common tactic

There is a kind of natural outgrowth of psychopath’s personality that manifests itself in some predator strategies and tactics. It is quite common for psychopaths to achieve their goal of self-gratification by a *three step-strategy* that in practice usually implies high emotional, financial or other practical trouble for other people.

What psychopaths often do *first* is assessing the value of individuals to their needs, and identify their psychological strengths and weaknesses (Babiak & Hare, 2006: 44). Note that it is unlikely that for psychopaths people exist in any other way than objects, targets and obstacles (see Babiak & Hare, 2006: 46). Psychopaths then size up the potential usefulness of an individual as a source of money, power, sex or influence (Babiak & Hare, 2006: 44). Besides assessing the potential gain from others, psychopaths assess their emotional weak

points and psychological defences in order to work out a plan of attack. Individual psychopaths do this in different ways and to varying degrees because of their psychopathic style (see previous section), experience and preference (Babiak & Hare, 2006: 44).

Secondly, psychopaths manipulate their future victims by feeding them carefully crafted messages, while constantly using feedback from them to build and maintain control. This is not only an effective approach to take with most people, it also allows psychopaths to talk their way around and out of any difficulty quickly and effectively when confronted or challenged (Babiak & Hare, 2006: 44). Many psychopaths are quite good at charming, a skill which they happily use to gain control over other people (e.g. Babiak & Hare, 2006: 48-50).

Thirdly, once psychopaths are bored or otherwise through with their drained and bewildered objects of attention, psychopaths abandon them happily. They can do that very easily and in a callous and harmful manner, because their emotional and social attachments to others are poorly developed (Babiak & Hare, 2006: 54).

4.5. On being emotionally blind

As we have seen, one of the items measured by the PCL-R is 'lack of remorse and guilt'. Verbal expression such as "I don't regret [anything]" (Hare, 1993: 41) are not uncommon for psychopaths, though even more often they pretend that they do feel remorse or some similar moral emotion in order to manipulate their victims. They often attempt to mimic emotions when necessary. Sometimes it is hard not to believe them, since for many psychopaths it is the case that their play-acting capacity is excellent. That is why they are so good at manipulation: they simply adapt themselves as a chameleon to the victim in front of them. Having said that, sometimes, when the topic of emotions and especially moral emotions is addressed, some psychopaths clearly show that at least sometimes they are B-grade rather than A-grade actors. What happens in these psychopaths is that they for example verbalize

remorse, but then contradict themselves in words when challenged. If, for example, a psychologist exercises a bit of pressure on a claim made by a psychopath that he feels remorse, the psychopath –when further pressed–subsequently may state that ‘he does not feel bad inside’. It is not the case however, that we can ascribe such bad acting to all psychopaths, not even when we only focus on ‘emotions’. Some psychopaths clearly seem to deserve the status of A-grade actors.

One other characteristic feature of psychopaths’ emotional life is that they can talk about their malevolent actions in the same emotionally flat manner as most other people normally talk about their breakfast. This emotionally flat fact-talk is often accompanied by a rather remarkable sense of logic. As one of Robert Hare’s objects of research expressed:

Anybody could have seen I was in a rotten mood that night. What did he want to go and bother me for? ...Anyway, the guy never suffered. Knife wounds to an artery are the easiest way to go.
(Hare, 1993, p.42)⁹

Research has demonstrated that psychopathic individuals experience emotions differently, showing qualitative and/or quantitative differences in their ability to experience emotion (see Hastings, Tangney and Stuewig, 2008). When it comes to emotions, psychopaths seem to be subject to some sort of emotional peculiarity that limits both the range and depth of their feelings. “[W]ords do not have the same emotional or affective colouring for psychopaths as they have for other people. ...[P]sychopaths lack some of the ‘feelable’ component of language” (Hare, 1993: 131). Psychopath Jack Abbott made this revealing comment: “There

⁹ Here is one further example of the extraordinary logic of psychopaths:

“Tyler has never been married, but has had several livings-in partners. In each case, he moved in with them after ‘sweeping them off their feet’, as he puts it. The longest relationship lasted 6 months, but each was marked by violence and instability. He speaks of countless instances where he was seeing other women while living with another. When asked whether he was ever monogamous Tyler says that he has always been monogamous. When his apparent inconsistency is brought out to him, he denies any contradiction: “I’ve always been monogamous, because it is physically impossible for me to be in two different places at exactly the same time. Understand?” (Blair, Mitchell & Blair, 2005: 5-6)

are emotions- a whole spectrum of them-that I know only through words, through reading and in my immature imagination. I can *imagine* I feel these emotions (know, therefore, what they are), but I *do not*. ” (Hare, 1993: 53).

While at times psychopaths are cold and unemotional, they are prone to dramatic, shallow and short-lived displays of feeling (see Hare, 1993: 52). These feelings may be no more than *proto-emotions*: primitive responses to immediate needs (see Hare, 1993: 53).

Besides what Karpman (1948: 479) calls the ‘generous emotions’: sympathy, gratitude, appreciation or any binding affection, psychopaths seem to have a striking absence of fear. But it does not stop there; psychopaths also seem to have difficulties to experience sadness. As pointed out by Vuilleumier et al¹⁰, there are strong suggestions that the sadness (and also fear) of others activates an automatic aversive response in observers (cited in Blair et al. 2005: 40). That does not seem to be at issue in the mind of psychopaths. Adults with psychopathy (and children with psychopathic tendencies) present with reduced autonomic responses to the sadness of other individuals. While psychopaths seem to have atypical responses to sadness and fear, it should be noted that psychopaths do not present with impaired responding to angry, happy or surprised facial or vocal expressions (see Blair et al, 2005: chapter 4; *cf* Blair, 2005).

4.6. A lack of moral/social conscience

Psychopaths have no difficulty in making use of people whenever they think using others serves their interest. Their motives are to manipulate and take, ruthlessly and without remorse (Hare, 1993: 145). Psychopaths also tend to see any social exchange as a ‘feeding’

¹⁰ Vuilleumier, P., Armony, J.L., Driver, J., Dolan, R.J. (2003). ‘Distinct spatial frequency sensitivities for processing faces and emotional expressions’. *Nature Neuroscience*, 6, pp. 624-631.

opportunity, a contest, or a test of wills, in which there can be only one winner (Hare, 1993: 145).

Psychopaths exercise their life style of self-gratification at other people's expense in the absence of empathy and without a moral conscience. It is highly plausible to see a cause and effect relationship here. That is, it is plausible to assume that psychopaths' ruthless behaviour does take place and can take place *because* psychopaths are not inhibited by empathy and a moral conscience. As Hare argues, without the shackles of a nagging conscience, psychopaths feel free to satisfy their needs and wants and do whatever they think they can get away with. Any anti-social act, from petty theft to bloody murder, becomes possible (Hare, 1993: 76).

I think Hare is quite right when he suggests that "Our collective amazement and fascination with the psychopath's utter disregard for rules suggests, by comparison, the power our 'inner policemen' actually have over us" (Hare, 1993: 75). As Hare claims, in typical human agents "Together [the] inner voice [of conscience] and the internalized norms and rules of society act as an 'inner policeman', regulating our behavior even in the absence of the many *external* controls, such as laws, our perceptions of what others expect of us, and real-life policemen" (Hare, 1993: 75). Psychopaths, by contrast, do not seem to have an internal policeman of that kind. In contrast to typical practical human agents, psychopaths carry out their evaluation of a situation - what they will get out of it and at what cost - without the usual anxieties, doubts, and concerns about being humiliated, causing pain, sabotaging future plans, in short, the infinite possibilities that people of conscience consider when deliberating possible actions (Hare, 1993: 78).

4.7. Irresponsive to treatment

In contrast to what is the case with other mental conditions having to do with anti-sociality, the behaviour of psychopaths is notoriously resistant to change (see Hare, 1993: 94). Current treatment programs designed to change the character traits of psychopaths into something less anti-social are unsuccessful. And that obviously is a problem given the disastrous emotional and financial costs psychopaths typically bring to others. Let us have a look at what Hare - who was asked by the Canadian government to develop a new sort of treatment programme - says about matters psychopaths and treatment. Hare argues:

Depressing though the evidence is, there are several things that we should consider before writing psychopaths off as untreatable or unmanageable...Recognizing the urgent need for new ways to deal with criminal psychopaths, and aware of the prevailing pessimism about traditional treatment programs, the Canadian government recently challenged me to design an experimental treatment/management program for these offenders...Although it is not possible to provide a detailed description of the program in this book [i.e Hare's book *Without Conscience*], some broad principles can be outlined. To a large extent, these principles are based on the view that the premise of most correctional programs-that most offenders have somehow gone off track and need only to be resocialized-is faulty when applied to psychopaths. From society's perspective, psychopaths have never been on track; they dance to their own tune. This means that the program for psychopaths will be less concerned with attempts to develop empathy or conscience than with intensive efforts to convince them that their current attitudes and behavior are not in their own self-interest, and that they alone must bear responsibility for their behavior. (Hare, 1993: 202-4)

Time will tell us whether the abovementioned Hare-type treatment-programmes are more effective than the typical treatment programmes that aim at emotional development. It could well be the case that psychopaths' behaviour can be changed into something less anti-social,

but possibly only if and when such change reflective of less anti-social behaviour reflects a real benefit for the psychopath. In any case, the emotional anomalies, in the form of an absence of a sense of remorse; a lack of empathy; and a lack of a moral conscience do not seem to be responsive to treatment at all. That is what clinical practice shows up until now. What this seems to mean is that the core of psychopathy might be untreatable, since many assume that it is emotional anomalies that exhaust or at least are part of the very core of psychopathy (see e.g. Blair et al. 2006). Perhaps 'anti-sociality' and/or some other (probably more peripheral) features of the syndrome are responsive to treatment - most likely only if and when change pays off for the psychopaths - but it may well be the case that all or at least some of the emotional anomalies are of such a nature that they will never respond to any kind of treatment. And that might be because there is a strong genetic component to these emotional anomalies, while environmental factors contribute more strongly to less essential features of the syndrome. More on this in § 9 below.

5. 'Factor models'

Empirical scientists have made attempts to group together certain individual behavioural features (e.g. 'pathological lying') and personality traits (e.g. 'manipulative') on basis of correlations between data generated by items of the PCL-R. The statistical term for the process that results in such groupings is 'factor analysis'. When statistical analysis shows that a certain number of behavioural features and/or traits form a certain group, these features and traits *seem to be conceptual* aspects of the *construct* of psychopathy (that is typically conceived of as consisting of some combination of character traits and behaviours). *Seem to be*. That is unfortunately not entirely clear. Researchers use 'factor analysis' to determine how interrelated some referents of terms such as 'pathological lying' and 'manipulative' are, but it is not entirely clear whether that actually means 'how interrelated some sort of

conceptual features of psychopathy are'. First, given that there is unclarity about the relationship between the *construct* of psychopathy and the *items* of the PCL-R (recall § 4.1). Secondly, also given that there is the problem in what realist, constructivist, anti-realist etc. way we should conceive of the ontological nature of some sort of syndrome called 'psychopathy' (this is a general problem for psychological disorders).¹¹ That is, if psychopathy indeed should be regarded as some sort of amalgam of traits and behaviours. Factor analysis might suppose some sort of realist ontology of symptoms on basis of which then judgements about factors are made based on data generated by a measurement tool for the measurement of psychopathy.

I have mentioned that factor analysis is used to determine how interrelated some referents of terms such as 'pathological lying' and 'manipulative' are. Another reason because of which researchers might use factor analysis is because they benefit from a reduction of the number of 'variables' (as to psychopathy, be they measurement tool items and/or some sort of conceptual attributes) by combining two or more variables into a single factor.

In a nutshell, the practice of factor analysis works as follows. Researchers collect data in the form of PCL-R scores. Having these scores, they enter them into a computer program. Subsequently researchers then give the computer program certain commands to find out what on basis of the PCL-R scores one can say about any relationship between items and thereby (so the assumption seems to be) about the relationship between psychopathy symptoms that are supposed to be measured by the items. Roughly speaking, the computer program responds to the researcher's commands by finding or not finding correlations between items pointing to these items belonging together and thereby reflecting a certain aspect of the psychopathic syndrome (so researchers seem to think).

In 1988, the grouping of features started off with the efforts of Harpur, Hakstian and Hare. They came up with a 2-factor model of psychopathy (see also Harpur et al, 1989; Hare et al. 1990): the interpersonal/affective factor and the impulsive/antisocial lifestyle factor.

¹¹ For lack of a better general term. Recall my comments on why it is problematic to call psychopathy a 'as laid out in § 4.2.

Some scientists nowadays work with a 3-factor model consisting of the factors: ‘arrogant and deceitful interpersonal items’ (the individual items involved are: glibness/superficial charm; grandiose sense of self-worth; pathological lying; conning/ manipulative); ‘deficient affective experience’ (items: callous/lack of empathy; shallow affect; lack of remorse or guilt; failure to accept responsibility for own actions), and ‘impulsive and irresponsible items’ (items: need for stimulation/proneness to boredom; parasitic lifestyle; lack of realistic long-term goals, impulsivity; irresponsibility) (Cooke and Michie, 2001). There is also a 4-factor model consisting of the factors: interpersonal, affective, lifestyle and anti-social (Neumann, Hare & Newman, 2007 for findings pointing to four factors; for a short explanation of the 4-factor model see Babiak and Hare, 2006: 27).¹² We have seen these four factors reflected in the description of the different psychopathic styles in § 4.3.

6. The prevalence of psychopathy

6.1. Psychopathy and Antisocial Personality Disorder (ASPD)

In this section I focus on the prevalence rates of psychopathy. Below we will see that one can specify prevalence rates for different populations within the class of psychopaths. Not only can prevalence rates be specified for different areas of a society, such as ‘prison population’, ‘forensic psychiatric settings’, ‘the business world’ and (what is commonly referred to as) ‘the household population’. One can do so as well for different cultures and for sex.

Although researchers frequently focus on different populations for reasons of academic interest, they sometimes do so as well for pragmatic reasons. When one –as a researcher- is based at a university in the United Kingdom, it makes sense to investigate psychopathy populations in the UK rather than for example the United States, unless one has

¹² There are two items of the PCL-R, i.e. ‘promiscuous sexual behaviour’ and ‘many short-term marital relationships’ that do not load on any of the 4 factors of the 4-factor model while these items do contribute to the total PCL-R score (Hare & Neumann, 2008: 220).

a specific and strong interest in gaining information about the United States population. And if one as a researcher is sufficiently satisfied with psychopathy information about which one can hypothesize that it may well be generalizable over the entire psychopathic population it makes sense to go to the prisons and forensic psychiatric settings rather than for example a big company. It does so, because it is in prisons and forensic psychiatric settings where one can reasonably expect to find a high number of psychopaths and it is there where one can investigate them rather easily.

Below, before we come to specific prevalence rates for a variety of subpopulations, I comment first on several important differences and relationships between the syndrome ‘psychopathy’, that is currently *not* in the *Diagnostic and Statistical Manual of Mental Disorders* and AntiSocial Personality Disorder (ASPD), that currently *is* in the *DSM*. An understanding of the differences and relationships between psychopathy and ASPD should help us to better understand the clinical practice surrounding psychopathy as well as to critically evaluate claims that are made about the prevalence rate of psychopathy.

According to the *DSM-V* (2013: 659) (as well as the *DSM-IV-TR* (2000: 701)), the essential feature of ASPD is “a pervasive pattern of disregard for, and violation of, the rights of others that begins in childhood or early adolescence and continues into adulthood”. According to the *DSM-V*, for the diagnosis ASPD to be given, the individual must be at least 18 years old, and must have had a history of conduct disorder before age 15 (on conduct disorder see § 9.3). Additionally, it is required that the anti-social behaviour does not occur exclusively during the course of schizophrenia or bipolar disorder (DSM-V, 2013: 659).¹³ On top of what I have just mentioned, the person must present with at least three of the following criteria:

- (1) failure to conform to social norms with respect to lawful behaviors as indicated by repeatedly performing acts that are grounds for arrest

¹³ Instead of ‘bipolar disorder’, the DSM-IV-TR refers to ‘a manic episode’ here.

- (2) deceitfulness, as indicated by repeated lying, use of aliases, or conning others for personal profit or pleasure
- (3) impulsivity or failure to plan ahead
- (4) irritability and aggressiveness, as indicated by repeated physical fights or assaults
- (5) reckless disregard for safety of self or others
- (6) consistent irresponsibility, as indicated by repeated failure to sustain consistent work behavior or honor financial obligations
- (7) lack of remorse, as indicated by being indifferent to or rationalizing having hurt, mistreated, or stolen from another¹⁴ (DSM-V, 2013: 659)

Let us recall that *in contrast to* ASPD, psychopathy is not now in the *DSM- V* (nor was it in any of the *DSM-IV* versions) as a separate diagnostic category and let us recall this while noting that when it comes to the list of criteria above belonging to an ASPD-diagnosis, there is some clear overlap with the PCL-R criteria for diagnosis of psychopathy. In what follows, I explain the reasons for the overlap in criteria. The same story will also offer reasons for why ASPD is in the *DSM* now, while psychopathy is not.

Traditionally, affective and interpersonal traits such as egocentricity; deceit; shallow affect; manipulateness; selfishness, a lack of empathy; and a lack of guilt and remorse have played a central role in the conceptualization of psychopathy. In 1980 however, this tradition was broken with the publication of the *DSM III*. In that now outdated version of the *DSM* psychopathy was renamed AntiSocial Personality Disorder and was now defined by persistent violations of social norms, including lying, stealing, truancy, inconsistent work behavior and traffic arrests (Hare, 1996). Amongst the reasons given for this shift away from the use of clinical inferences were that personality traits are difficult to measure reliably, and that –on the side of behaviour- it is relatively easy to agree on the behaviours that typify a disorder. This focus on behaviour rather than traits is a fundamental cause of the current fact

¹⁴ These criteria are the same for the DSM-V and DSM-IV-TR.

that currently the clinical diagnostic label ASPD is applied to a very diverse class of people, in terms of personalities (and the etiology of the syndrome) rather than behaviour.

A further cause of and no help at all for getting rid of the heterogeneity is the fact that in the *DSM-IV* and the *DSM-V* the description of ASPD I have already presented, i.e. “a pervasive pattern of disregard for, and violation of, the rights of others that begins in childhood or early adolescence and continues into adulthood”, in the *DSM-IV* and the *DSM-V* is directly followed by the following sentence: “This pattern has also been referred to as psychopathy, sociopathy or dissocial personality disorder” (DSM V, 2013: 659; DSM-IV-TR, 2000: 701-2). Things become even more nasty and confusing given that a particular section of the *DSM* (the ‘associated features supporting diagnosis-section’) that succeeds the ASPD description in the *DSM-V*, states the following:

Individuals with antisocial personality disorder frequently *lack empathy and tend to be callous, cynical, and contemptuous of the feelings, rights, and sufferings of others. They may have an inflated and arrogant self-appraisal* (e.g., feel that ordinary work is beneath them or lack a realistic concern about their current problems or their future) and may be excessively opinionated, self-assured, or cocky. They may display a *glib, superficial charm and can be quite voluble and verbally facile* (e.g., using technical terms or jargon that might impress someone who is unfamiliar with the topic). *Lack of empathy, inflated self-appraisal, and superficial charm are features that have been commonly included in traditional conceptions of psychopathy that may be particularly distinguishing of the disorder and more predictive of recidivism in prison or forensic settings where criminal, delinquent, or aggressive acts are likely to be non-specific. These individuals may also be irresponsible and exploitative in their sexual relationships.* (DSM-V: 660; see also the identical passage in the DSM-IV-TR, 2000: 703, my italics; cf Hare, 1996)

In the passage above one sees a clear reference to traditional features of psychopathy (as represented by Robert Hare’s construct and measurement tool, but even down to Cleckley). All that is italicized represent traditional features of psychopathy.

Overall, it seems sensible to conclude that there is significant intermingling between the term and concept 'psychopathy' and the concept 'ASPD'. The seriousness of this should not be underestimated. It means that some significant number of people have been and continue to be mistakenly referred to as psychopaths on the basis of *DSM-III*, *DSM-III-R* (revision), and *DSM-IV* and *DSM-V* criteria for ASPD (adapted from Hare, 1996). In worst case scenarios, this can lead to a mistaken judgement that someone should be given the death penalty, a view that is probably indirectly expressed by this quotation from Robert Hare:

We don't know how many ...inhabitants of death row actually exhibit the personality structure of the psychopath, or how many individuals meet the criteria for ASPD, a disorder that applies to the majority of criminals...If a diagnosis of psychopathy has consequences for the death penalty- or for any other severe disposition, such as an indeterminate sentence or a civil commitment- clinicians making the diagnosis should make certain they do not confuse ASPD with psychopathy. (Hare 1996)

Note that when it comes to the relationship between psychopathy and the death penalty, as an outsider or lawyer, one sometimes cannot easily predict in advance whether a psychopathy diagnosis of an expert will *prevent* an offender from receiving the death penalty (rationale: he should not be given one because there is something wrong in his brain that is not under his control) or whether it makes him more susceptible to one (as a psychopath he clearly wasn't mad/suffering from psychoses and the like, when he committed this crime).

6.2. The general prevalence rate of psychopathy

There is a lot one can say about the prevalence rate of psychopathy. A leading scientific opinion is that the general prevalence rate of psychopathy is around 1% (see primarily Hare, 1993: 193; 2003). As it is general, it does not say anything specific about particular populations, but when one looks at the rate one needs to be aware of the fact that it heavily relies forensic populations for the very simple reason that most research has been done with this population. And that is the case because the base rate of psychopathy in forensic populations is high and the information needed for reliable assessments is readily available (see Babiak, Neumann & Hare, 2010).

Wondering what type of evidence there is for this general prevalence rate of 1%, I have discussed this with a forensic psychologist. This discussion has led me to the tentative conclusion that in fact there is no hard scientific evidence for the prevalence rate of 1%. The expert with whom I have discussed matters tentatively suggests that the thought that leads Hare to the conclusion that the prevalence rate is 1% might be the following one. In the 1980's, a study with no fewer than 20000 adults was carried out in the United States: the Epidemiological Catchment Area Study. This study found a general prevalence for ASPD of around 3-4%. In prisons in the United States ASPD is diagnosed in around 75% of prisoners and the rate of PCL-R psychopathy in prisons is around 20-25%. The expert suggests that the combination of all of these findings might have led Hare and others to the conclusion that the general prevalence rate of psychopathy is around 1%, since the 1% amounts to about a third of the ASPD prevalence rate found for the general population in the Epidemiological Catchment Area Study.

6.3 Psychopathy in the forensic population

According to Blair, Mitchell & Blair (2005), studies reveal that 12-25 % of inmates in the United States meet criteria for psychopathy as laid down by the PCL-R. Hare (1996) claims a prevalence rate between 15 and 20% in offenders (no claim about countries included). Vitale & Newman (2001) claim 15-30% for the offender population (no claim about countries included). Hobson and Shine (1998) found that 26% of their sample of UK prisoners qualified as a psychopath (based on the PCL-R). This is considerably higher than what previous research has shown about *UK* samples (Hobson & Shine, 1998; my italics). And it is also considerable higher than what Coid, Yang, Ullrich, Roberts, Moran and Bebbington found in their (2009) study that looked at 500 prisoners in different prisons in England and Wales. Using the standard cut-off score for psychopathy of 30, the researchers found that the prevalence rate was 7, 7% for men and 1, 9% for women. More about psychopathy in females in § 8.

6.4. Psychopathy in the household population

Research on psychopathy in the household population is very rare. It is almost limited to the following (2009) study in Great Britain. As a measurement tool this study used the PCL: SV (screening version; recall footnote 5). That is, the shorter variant of the PCL-R.

In the (2009) study of Coid, Yang, Ullrich, Roberts and Hare, a two phase-survey was conducted with 8886 participants in the first phase and 638 in the second. In the first phase participants had to fill in a questionnaire on basis of which a smaller second sample was selected. Those included in the second sample were considered as serious candidates for people having psychological/psychiatric disorders. On basis of the information the latter group provided, i.e. in terms of outcomes of clinical questionnaires and further collateral

information as part of the PCL-R: SV, the researchers concluded that psychopathy is a rare condition in the general population. Only a very small minority of individuals met common criteria for psychopathy or demonstrated elevated levels of psychopathic traits. The researchers concluded that “the weighed prevalence [for psychopathy in the household population] is 0.6” (Coid, Yang, Ullrich, Roberts, Hare, 2009).¹⁵

The above study suggests that psychopathy prevalence rates in the normal population are low. This research claim on basis of empirical data may reflect the fact that there indeed are hardly any psychopaths in the community population. It theoretically could also mean however that this study has not managed to catch the psychopaths living in the community population, psychopaths who manage to escape prisons and perhaps even contact with the criminal justice system. If there are any psychopaths in the household population, the explanation why they have not contributed to the psychopathy rate found by researchers is readily available. Successful psychopaths will not be interested in participating in academic research, unless perhaps, on rare occasions, for reasons that can fulfill their extreme need for self-gratification, whatever those reasons might be.

6.5. The prevalence of psychopathy in the business world

According to Hare “psychopaths have relatively little difficulty infiltrating the domains of business, politics, law enforcement, government, academia and other social structures” (Hare, 1996). Of these domains, the one investigated most is business. In this section, I will explain

¹⁵ At face value this conclusion means that this study shows that 1 out of 160 individuals presents with psychopathy. However, if one has a somewhat closer look at the results of the study, it turns out that a different cut-off point for the PCL-R: SV was used. Instead of the typical cut off score of ‘18’, cut off points of 13 and 11 were used. The researchers offer some reasons for that that have to do with cut-off points used in other studies (see Coid et al., 2009 and Steadman et al., 2000). Be that as it is, using a lower cut-off score may well mean that one’s group of ‘psychopaths’ is (far too) over-inclusive.

a bit more about this phenomenon of corporate psychopathy that according to some researchers is clearly there. I will start off with an opening quotation of a recent article:

“Not all psychopaths are in prison. Some are in the Boardroom”

This assertion came in the guise of a rather casual response from Robert Hare to a question asked at the end of the 2002 Canadian Police Association meeting (Babiak, Neumann, Hare, 2010). The authors subsequently comment:

The questioner turned out to be a journalist, and over the next few days the international media picked up his newspaper article, treating the statement as somewhat of a revelation. The media reports clearly reflected both the popular view that psychopathy equates to criminality and violence, and the public and media fascination with murder and mayhem, typically attributed to ‘psychopaths’ or ‘sociopaths’. Unfortunately, media headlines and popular television crime shows are often the only exposure the public gets to the concept of psychopathy, resulting in considerable misinformation and misunderstanding. This is also the case with business professionals, who see even less relevance of such portrayals to their daily interactions with co-workers. The problem is exacerbated by the paucity of research on the prevalence and implications of psychopathy for society in general, and by the heavy emphasis on research with offender and forensic psychiatric populations (where the base-rate for psychopathy is high and the information needed for reliable assessments is readily available). (Babiak, Neumann & Hare, 2010)

What this quotation reflects is the view that there are individuals who on basis of their personality characteristics and behaviour can be described as psychopaths, but despite that, generally are not violent in the sense of physically, or even, mentally, aggressive. Their psychopathic nature reveals itself in the fact that that they bring about financial and social-emotional disasters for their work environment and the community in a wider sense by

acting as an intra-species predator¹⁶: by exploiting other human beings for the fulfillment of their desires (paradigmatically for power and money).

As indicated by the above quotation, in contrast to psychopathy in forensic populations, there is relatively little known about ‘corporate psychopathy’ (or ‘industrial or ‘while-collar’ psychopathy). There has been conducted far more research on the criminal populations than on the psychopaths or more careful, candidate psychopaths in the business world. As indicated by Babiak et al, an important reason for the rather huge quantitative difference in research is that for offender and criminal settings it is the case that the (expected) base-rate of psychopathy is high and the information needed for reliable assessment is readily available. A further reason that explains the quantitative difference in research is that it is very hard for scientists to obtain the active cooperation of business organizations and their personnel for research purposes. Hardly any employer is eager on letting researchers investigate their company just for the reason that it benefits science. Having said that, they may be sufficiently or even enormously happy to let empirical psychologists research their organizations in two cases: either when there is interpersonal trouble in a company around a particular individual or when there is another interest for the company in getting insight into the nature of its employees.

So what does research indicate about prevalence rates of psychopathy in the business population? In their (2010) study, Babiak et al. found that of the 203 participants that consisted of managers and executives, the vast majority of PCL-R scores (80%) were between 0-3. Having said that, at the other end of the spectrum, 8 participants (3.9%) had a score of 30 or higher; two had a score of 33 and one had a score of 34. So, if the PCL-R is taken as a criterion for a rightful diagnosis of psychopathy, then this study suggests that there are psychopaths in the business world.

¹⁶ I borrow this term from Meloy, J. R. & Meloy, M.J. (2002). ‘Autonomic arousal in the presence of psychopathy: a survey of mental health and criminal justice professionals’. *Journal of Threat Assessment* 2, pp. 21–33; referred to by Meloy and Reavis (2007).

6.6. United States versus United Kingdom

It looks like we can add something to the list of things that are bigger, higher, more extreme in the US: psychopathy scores. On basis of the data of their study, Cooke, Michie, Hart and Clark conclude that (generally speaking) PCL-R rates are higher for North American than for European people (PCL-R total scores, factor scores as well as individual item scores (Cooke, Michie, Hart and Clark, 2005b).

I have addressed the topic of factor analysis in § 5. As mentioned there, there exist 2-factor, 3-factor and 4-factor models of psychopathy. The research I discuss below bases itself on the three-factor model. Recall that the three factor model distinguishes between the factor *arrogant and deceitful interpersonal items* (items: glibness/superficial charm; grandiose sense of self-worth; pathological lying; conning/ manipulative); *deficient affective experience* (items: callous/lack of empathy; shallow affect; lack of remorse or guilt; failure to accept responsibility for own actions), and *impulsive and irresponsible items* (items: need for stimulation/proneness to boredom; parasitic lifestyle; lack of realistic long-term goals, impulsivity; irresponsibility).

Starting off from their general discovery that North-Americans generally score higher on the PCL-R than European people do, Cooke et al discovered something more specific in their study. First, Cooke and colleagues found that for the first factor ‘arrogant and deceitful interpersonal items’, there are substantial differences, “particularly at the high end of the trait” (Cooke et al. 2005a). That is supposed to mean that on the upper end of the PCL-score dimension, North-American people score higher than people in the United Kingdom. Another finding that comes out of their research is that the difference between North-Americans and European people was largest for factor scores 1 and 3: ‘*arrogant and deceitful interpersonal items*’ and ‘*life style*’, respectively. As Cooke and Michie (2005b) suggest, this may well mean that character features related to items that fall under the second factor

'deficient affective experience' form the core element of psychopathy. Note that the cultural difference observed was similar to that reported in previous research (Cooke & Michie, 1999), although somewhat smaller (Cooke & Michie, 2005a).

6.7. Psychopathy in primitive cultures

There is evidence for the conclusion that psychopathy is not just some kind of Western phenomenon. Research points to the tentative conclusion that it may well be the case that both amongst Eskimos in Alaska and amongst a primitive group called 'Yorubas' in the jungle of Africa there exist such individuals.

Anthropologist Jane Murphy spotted that the Eskimos of North-West Alaska have a word, *kungaleta*, which means 'his mind knows what to do, but he does not do it'. Murphy states:

[This word] might be applied to a man who, for example, repeatedly lies and cheats and steals things and does not go hunting and, when the other men are out of the village, takes sexual advantage of many women-someone who does not pay attention to reprimands and who is always being brought to the elders for punishment. (Murphy, 1976: 1026)

At the time of research, one Eskimo among the 499 on their island was given the qualification '*kunlangeta*'. When the Eskimos were asked by Murphy what would have happened to such a person traditionally, one Eskimo said that probably "somebody would have pushed him off the ice when nobody else was looking" (Murphy, 1976: 1026).

Murphy spotted something similar in the jungle of Africa. As noted by her, a group called the *Yorubas* use a word in their communication the meaning of which is quite close to the word '*kunlangeta*' used by the Eskimos. The Yorubas' word '*arankan*' refers to a similarly egocentric creature; more specifically someone who always goes his own way regardless of

others, who is uncooperative, full of malice, and bull-headed. Neither *kunlangeta* nor *arankan* were thought to be curable by native healers (Murphy, 1976: 1026).

7. More scientific approaches to psychopathy

7.1 Five further conceptions

As we have seen psychopathy is most frequently considered as a distinct cluster of personality traits and behaviours. There is more to be said however in terms of what scientists think about psychopathy. Looking at the scientific literature, it seems fair to say that one can find several more ideas about it. Below I explain five further conceptions. While I aim at presenting a representative sketch of the field, I may overlook one or the other conception. Furthermore, it needs to be said that while I distinguish five conceptions of -or perhaps better- 'approaches to' psychopathy, this is not to say that a particular presented idea is incompatible with the others; I think it is better to consider them as accounts that have one or the other specific idea about the nature of psychopathy, an idea that however does not rule out a compatibility with other accounts. The five conceptions of/approaches to psychopathy I discuss below are: (i) psychopathy as a personality disorder, (ii) psychopathy as an extreme variation of normal personality traits, (iii) psychopathy as a neurological deficit, (iv) psychopathy as a genetic deficit, (v) psychopathy as an inhibition problem, (vi) psychopathy as an adaptive reproductive strategy.

7.2. Psychopathy as an extreme variation of normal personality traits

It is possible to regard psychopathy as an extreme variation of normal personality traits. One could say that there are two variants of this idea in scientific literature. The first one starts

off from what in psychology is known as ‘The Big Five’: five umbrella core features that capture our personality traits. The second one relies heavily on research tied to the conception of psychopathy presented already above: the one that considers psychopathy as a distinct cluster of personality traits and behaviour.

At the end of the 20th century, McCrae and John (1999; *cf* McCrae et al. 2000) argued that personality traits come down to an umbrella set of five. The following are the five core traits identified:

1. openness to experience
[imaginative versus down to earth; variety versus routine; independent versus conforming]
2. conscientiousness
[organized versus disorganized; careful versus careless; self-disciplined versus weak-willed]
3. extraversion
[social versus retiring; fun-loving versus sober; affectionate versus reserved]
4. agreeableness
[soft hearted versus ruthless; trusting versus suspicious; helpful versus uncooperative]
5. neuroticism
[worried versus calm; insecure versus secure; self-pitying versus self-satisfied]

From the top downwards, the first letters of the five traits lead to the acronym OCEAN. (Gazzaniga & Heatherton, 2002: 485, 494-495).

There is a substantial amount of research on the relationship between ‘The Big Five’ and different types of ‘socially malevolent personalities’. (See e.g. Paulhus & Williams (2002); Vernon, Villani, Vickers, Aitken, Harris, 2007.) Typically, researchers who investigate the relationship between ‘The Big Five’ and socially malevolent personalities separate three types of socially malevolent character: Machiavellianism, psychopathy and narcissism (together they are often called ‘the Dark Triad’). Machiavellianism is in particular associated with manipulative behaviour. Besides investigating what one can say about the presence of the Big Five character traits in these different sorts of socially malevolent character, researchers also investigate how these three types of socially malevolent character relate in terms of the degree in which they incorporate the character traits of the Big Five. Here, from an empirically informed conceptual perspective it is assumed that Machiavellianism, psychopathy and narcissism have something in common in that they refer to a socially malevolent character with behavioural tendencies toward self-promotion, emotional coldness, duplicity, and aggressiveness (see Paulhus & Williams, 2002). Researchers then try to find out how this assumed similarity expresses itself in similarities and differences in terms of the degree to which psychopaths, Machiavellianists and narcissists are open, conscientious, extrovert, agreeable and neurotic.¹⁷

There is a second variant of the ‘extreme variation of normal personality’-conception. Rather than being based on/in *The Big 5*, it relies heavily on the traits that are ascribed to psychopaths on the conception of psychopathy as advocated by Hare. On this second ‘extreme variation of normal personality view’, the idea is that psychopathic features of an ‘affective’, ‘interpersonal’, ‘lifestyle’ and ‘anti-social’ kind range from (near) nothing to abnormally high (see Babiak & Hare, 2006: 29). Those at the upper end are referred to as psychopaths. As mentioned by Babiak & Hare, clearly most people fall between the extremes.

¹⁷ When it comes to research on the Dark Triad (including ‘psychopathy’) and the Big Five, it must be mentioned that those who are referred to as psychopaths do not meet the standards of the PCL-R. Also, it is often self-report scales for ‘psychopathy’ rather than the PCL-R that are used to investigate the relationship between ‘psychopathy’ and the PCL-R.

If we interpret this by means of the PCL-R that means that there is a psychopathy dimension, ranging from 0 to 40 on which -by far- most people score in between; and -noteworthy-most do so towards the lower extreme (Hare & Babiak, 2006: 29).¹⁸ Those in the middle have a significant number of psychopathic features, but they are not psychopaths in the scientific strict PCL-R sense of the term (see Hare & Babiak, 2006: 29). These people clearly are not model citizens as most people would understand them, but lack any clinical form of psychopathy; sometimes scientists refer to them as ‘subclinical psychopaths’ (Hare & Babiak, 2006: 29).

Now, Babiak & Hare mention:

Whatever cut score is used [30 for the PCL-R], individuals who meet or exceed the score are different from those with lower scores. (Babiak & Hare, 2006: 28)

This comment is in clear need of some critical questions, partly provided by the authors of the assertion themselves. The initial questions about their remark seem to be: (a) is their claim correct and (b) in what way are individuals scoring minimally the cut-off score different from those who score below the cut-off point?

Let us consider the following assumptions:

- a) there is a clear difference between people who meet or exceed the cut-off score of the PCL-R and those who do not
- b) whether this is a difference in kind or rather degree is yet to be firmly established, (Babiak & Hare, 2006: 28).

Both assumptions seem compatible with an account of psychopathy as an extreme variation of normal personality traits, even though at face value the second assumption is not, for –the

¹⁸ Note that according to Hare (1996), the PCL-R mean score for non-criminals is 5.

idea would be- how can there be anything else than a difference in degree if one thinks of psychopathy as an extreme variation of normal personality traits? Well, there can be, I think. What one would have to do is thinking of psychopathy as a syndrome, i.e. a cluster of personality traits, (e.g. 'being manipulative', 'lacking empathy', 'having a sense of grandiosity', 'lacking guilt and remorse'), while granting that these traits can exist in different degrees in an individual as milder or more severe variants of normal personality traits. Next one would have to conclude that when these traits are *at the upper end* of a continuum of normal personality traits, the *sum total* of these traits when they exist together in the psychology of an individual - dependent on what exactly the necessary and sufficient traits are for psychopathy - could reflect a (pathological) syndrome that is qualitatively different from the case where these traits are found in the psychology of an individual in a mild variant. This view would consist in the idea that it is a *falsehood* that because certain particular milder variants of psychopathic traits ('being manipulative', 'lacking empathy', 'having a sense of grandiosity', 'lacking guilt and remorse') are milder variants of something more severe (the same traits, but more severe), that therefore the bunch of milder variants is a 'milder variant' or 'subclinical form' of the severe something that is legitimately referred to as psychopathy.

7.3. Psychopathy as a neurological deficit

7.3.1 On differences and defects

In due course I will come to talk about the fact that some scientists think of or at least *talk* about psychopaths' brain states as being 'defective' or 'abnormal', while other scientists prefer not to refer to psychopaths' brains by means of these terms and rather talk about them as just being 'different'. Before doing so, let me point out that issues about normalcy and abnormalcy look extremely conceptually difficult. The following list of descriptions from a

common dictionary shows clearly why it is one should not take conceptual talk about normalcy lightly. The dictionary gives the following six meanings of the term ‘normal’.

1. conforming to the standard or the common type; usual; regular; natural;
2. serving to *fix* a standard;
3. of natural occurrence;
4. approximately average in any psychological trait, as intelligence, personality, or emotional adjustment
5. free from any mental disorder; sane;
6. free from disease or malformation (Costello, 1991¹⁹, cited in Dupré, 1998: 225)

A distinction one implicitly finds in this list is the one between statistical (nr 1 and 4) and normative notions of normalcy (nr 5 and 6). As mentioned, some scientists think of or at least *talk* about psychopaths’ brains as suffering from a normative defect, while other scientists prefer not to do so and rather talk about psychopaths’ brain states as being ‘different’ compared to some other class(es) of brains. It seems to be the case that the latter can be explained by neuroscientific uncertainty and unclarity as to what type(s) of brain states count as *normal* brain states; something from which uncertainty and unclarity as to what count as *abnormal* brain states follows organically and quickly. When it comes to researchers’ unwillingness to describe psychopaths’ brains as abnormal due to uncertainty and unclarity about what count as normal brains, the uncertainty and unclarity seems to apply to both statistical and normative notions of normalcy. This then seems to be due to the uncontroversial fact that there is an absence of large-scale empirical studies on ‘candidate normal brains’.²⁰ Without that it clearly cannot be settled what counts as *statistically* common neural make-up in certain (experimental) circumstances. And without that, on its

¹⁹ Costello, R.B. (ed.) (1991). *Random House Webster’s College Dictionary*. New York: Random House.

²⁰ This view is grounded in remarks made during the workshop ‘Psychopathy: from laboratory to court room’ which I attended (University of Kent, Canterbury, June, 22, 2010, led by Robert Hare; pre-conference workshop to the Division of Forensic Psychology Annual Conference 2010).

turn, it seems hard, if not utterly impossible, to determine what counts as a *normatively* abnormal brain.

It must be noted that being a member of a brain population reflective of statistical normalcy might not be a necessary condition for being a member of a class of brains reflective of normative normalcy, i.e. at least when it comes to the issue whether *psychopaths'* brains are normatively abnormal. That is, it could be the case that both (a rare type) psychopathic brain and a common type of brain of other human beings can rightfully be classified as normatively normal. A plausible case for this verdict could be one in which both the brains of psychopaths and the brains of a large class of other human beings reflect neural correlates of evolutionary determined different genetic recipes for brain states, personality and behaviour. The brains of 'others' could reflect some advantage coming from pro-social traits and behaviours, while the brains of psychopaths could be reflective of individuals that were genetically pre-supposed to certain exploitative behaviour and through evolutionary history have found and made use of some adaptive ecological niche for this behaviour (more on this in § 7.6).

7.3.2. Some cautionary notes

Let us now turn to some research on psychopaths' brains. Neuroscientists have different ideas about differences/abnormalities in the brains of psychopaths. I will mention a few. Note that the neuroscience of psychopathy is a quickly growing and extensive field of research. It is beyond the purposes of this chapter to offer more than the introductory sketch below.

When it comes to neuroscientific research on psychopathy, there is both research on *structural* (anatomical) and *functional* anomalies (how does a brain process certain information) in psychopaths. Having said that, one has to be careful with interpretation of the literature, especially when it comes to structural anomalies. Weber, Habel, Amunst and

Schneider (2008) provide a literature review on structural brain anomalies in psychopaths. They argue that:

Data in the literature report a reduction in prefrontal gray matter volume, gray matter loss in the right superior temporal gyrus, amygdala volume loss, a decrease in posterior hippocampal volume, an exaggerated structural hippocampal asymmetry, and an increase in callosal white matter volume in psychopathic individuals. These findings suggest that psychopathy is associated with brain abnormalities in a prefrontal–temporo-limbic circuit—i.e. regions that are involved, among others, in emotional and learning processes.

In their review article, Weber and colleagues furthermore conclude that “data indicate that psychopathic individuals cannot be seen as a homogeneous group”. Let me comment as follows. It could very well be correct that those human beings who (for one reason or another) can legitimately be called ‘psychopaths’ are in some significant way not a completely homogeneous group, possibly also in terms of their brain architecture. We can think here for example of the different psychopathic styles advocated by Hare & Babiak (2006) that might come with somewhat different brain structures (*cf* Karpman, 1948 who separates out two real types of psychopaths amongst six ones that clinicians might want to call psychopaths). However, we should not automatically conclude that psychopaths are not a homogenous group on basis of the existing variety of research conclusions about the brains of human beings described by the authors of the studies as ‘psychopaths’. We should not do so when there is an absence and given the absence of homogeneity in the neuroscientific literature on psychopathy as to what the construct of psychopathy amounts to. Sometimes, in neuroscientific literature, ‘psychopathy’ is defined loosely and as such incorporates or is identified with some sort of rather severe ‘anti-sociality’ more generally. An identification of psychopathy with some sort of broad construct of severe anti-sociality invites for unhelpful but accurate conclusions about heterogeneous brain structures (that would be the physical foundation corresponding to the anti-sociality reflecting (highly) significantly different

syndromes/cognitive anomalies). This problem of loose definitions one finds in the neuroscientific literature on 'psychopathy' is a methodological problem that is different from another methodological problem any reader of neuroscientific literature on psychopathy should be alert on: the problem of researchers using a plausible and strict definition of psychopathy in their study, while working with research subjects (the alleged psychopaths) that fail to match the theoretical construct of psychopathy as advocated by the authors and/or fail to meet the strict demands of the (often and hopefully) very valid and reliable measurement instrument the neuroscientific researchers choose to identify a group of psychopaths whose brain they then subsequently investigate. Here is an example of this problem. Suppose, as an author of a neuroscientific study, one intends to identify one's research objects by means of the PCL-R because (reasonably) to one that seems to be the best measurement instrument to identify psychopathy. As we have seen, the PCL-R demands a cut-off score of 30 (out of the possible 40). Unfortunately, occasionally, it does happen that researchers take this cut-off score somewhat loose and take psychopaths to be those research participants that score above 25 rather than 30 (or, incidentally, even lower than 25). This practice is helpful to/convenient for a researcher in the way that otherwise, as a researcher, one may end up with a (far) too low number of real PCL-R psychopaths to draw significant research conclusions. Be that as it is, researchers should not lower the cut-off score. This namely runs the risk of not investigating the nature of psychopathy, but rather some other severe anti-sociality syndrome. A difference of five points in terms of cut-off score may mean that one investigates too many other severally socially malevolent beings besides some real psychopaths.

There is a further important methodological point that needs to be made in regard to neuroscientific research on psychopathy. Let us suppose that a particular neuroscientific study is conceptually reasonable (i.e. it works with a sufficiently conceptually plausible –e.g. Cleckley or Hare- psychopathy construct that as such is not a loosely –broadly 'anti-sociality'-construct) and reflects a match between the conceptualization of psychopathy and

the research participants. Suppose next that all research participants of this study are a sufficiently homogenous group for scientific purposes. That is, research participants are similar in intelligence, socio-economic background and some other features that could possibly lead to brain differences. Suppose furthermore, that on basis of research with their participants (the) researcher(s) of the study conclude that there is a widely shared or universal brain anomaly in their group of research subjects (compared to some control group of more statistically (and normatively) normal human agents). In this situation, the researcher(s) who conduct(s) the study (and their colleagues who read their published papers) still need to be alert on the possibility that a particular substantive research conclusion about a brain anomaly in psychopaths (say, the size of a particular brain bit X is really different from the size of this particular brain part in typical human beings) reflects a factor that is widely shared by one's research participants, while it does not have anything whatsoever to do with the nature of the personality syndrome 'psychopathy'. One thing one could think of e.g. is substance abuse of research participants. What could happen is that the brains of real psychopaths lead researchers to wrong conclusions about the neural correlates of psychopathy due to the fact that the homogeneity they find in their research populations reflects a shared factor not having anything to do with psychopathy's brain essence, but rather reflects something else, most probably something that often goes with criminal behaviour in general, say excessive use of alcohol or drugs. So, the brains of psychopaths could lead to conclusions about structural anomalies in psychopaths not because of some structural brain anomaly that is characteristic of the syndrome psychopathy, but rather as a result of –in this case- the brain upshot of substance abuse. I have pointed out the just mentioned issue in order to further support a general point I would like to make: the point that one needs to be careful with interpreting neuro-scientific research on anti-sociality. I have tried to support this point by means of what I observe in what at places looks like a somewhat messy neuroscientific field of 'psychopathy'; one with loose/inadequate definitions when it comes to anomalies in 'psychopaths' brains'. Having said that, there clearly are really

good neuroscientific studies on psychopathy and –as far as I can see- these studies that are conceptually sufficiently sound (and have their research participants matched accordingly) do not have research participants with a history of substance abuse which - on its turn- is in line with dominant ideas about the psychopathy construct. Typically, alcohol and drugs use are considered as unessential to the nature of psychopathy. Indeed, researchers typically do not even want to consider alcohol and drug use as some sort of peripheral feature of psychopathy. Alcohol and drugs use seem utterly irrelevant to the psychopathy construct.²¹

In this section I have wanted to make the general point that one should be careful in terms of interpretation of neuroscientific research on anti-sociality syndromes, including psychopathy. I have done so by giving two examples of methodological trouble on may come across when having a look at the neuroscientific literature. I commented first on Weber et al.'s claim in their review article that psychopaths form a heterogeneous population. In short, my response has been that if study (a) on 'psychopathy' works with anti-social research participants that are very different from the participants in study (b) and (c) (either by definition or otherwise), then it should not come as a real surprise that 'psychopaths' form a heterogeneous population (NB: (mistaken) conclusions about psychopaths being a heterogeneous population could also arise from some sort of in-group variation amongst the participants of one single study.) Secondly, I made the point that when interpreting the literature on neuroscience of psychopathy, one also needs to be careful that the authors' conclusions about characteristic features of 'psychopaths' brains, conclusions grounded in a particular spotted similarity in the brains of research participants that is markedly differently

²¹ Granting for the sake of the argument and apparent reality, that the substance abuse problem I have laid out in the main text *does not apply to the better neuroscientific studies on psychopathy*, let me note three things, the first one being having been mentioned already. First, the neuroscientific literature on anti-sociality suggests that substance abuse is a clear candidate for misleading conclusions about the neural correlates of anti-sociality syndromes. Secondly, being informed by neuroscientific studies, I cannot think of any other factors that are such a very clear candidate for misleading conclusions about brain anomalies as substance abuse seems to be. Thirdly, having said that, it will not harm, indeed benefit the neuroscientific field of psychopathy and other anti-sociality syndromes (disorders) if, by default, scientists and others suppose that there may be other factors besides substance abuse that relatively homogenously characterize the brains of the research participants, while the neural correlates of these factors that lead researchers to the conclusion that there is a brain anomaly, are unreflective of the syndrome (disorder) the researchers want to investigate in their study.

from what is found in the brains of a control group, do not reflect something else that has nothing to do with the nature of the syndrome.

7.3.3. Studies on psychopaths' brains: some examples

Neuroscientific literature on psychopathy shows that it is a widespread thought that *amygdala* dysfunction is one of the core problems that characterizes psychopathy. Note that the amygdala is a little brain part –almond shaped–that as a core component belongs to the *limbic system*. The limbic system is the part of the brain that is related to drives/motivations/inclinations and emotions (see also footnote 30). One conclusion that seems to hold in terms of *structural* anomalies comes from research from Tiihonen et al. (2000²², cited in Blair 2003). They conclude that high levels of psychopathy were associated with *reduced volume* of the amygdala.

In terms of *functional* amygdala anomalies, research has demonstrated that psychopathic individuals experience emotions differently, showing qualitative and/or quantitative differences in their ability to experience emotion (Hare, 1993, 1998²³; Steuerwald & Kosson, 2000²⁴) and process affective language (Gillstrom & Hare, 1988²⁵; Hare & McPherson, 1984²⁶; Williamson, Harpur, & Hare, 1991²⁷). Significant abnormalities have

²² Tiihonen, J., Hodgins, S., Vaurio, O., et al. (2000) 'Amygdaloid volume loss in psychopathy'. *Society for Neuroscience Abstracts*, 2017.

²³ Hare, R.D. (1998). 'Psychopaths and their nature: implications for the mental health and criminal justice Systems'. In Millon, T., Simonsen, E., Birket-Smith, M., Davis, R. (eds.), *Psychopathy: Antisocial, Criminal, and Violent Behaviour* (New York: Guilford Press), pp. 188-212.

²⁴ Steuerwald, B.L. & Kosson, D.S. (2000) 'Emotional experiences of the psychopath'. In Gacono, C.B. (ed.). *The Clinical and Forensic Assessment of Psychopathy* (Mahwah, New Jersey: Lawrence Erlbaum Associates), pp. 111-135.

²⁵ Gillstrom, B.J. & Hare, R.D. (1988). 'Language-related hand gestures in psychopaths'. *Journal of Personality Disorders*. 2(1), pp. 21–27.

²⁶ Hare, R.D. & McPherson, L.M. (1984). 'Psychopathy and perceptual asymmetry during verbal dichotic

also been found in physiological responses to affective material (Hare, 1978²⁸; Patrick, Bradley, & Lang, 1993²⁹) and memory for emotional events (Christianson et al., 1996³⁰) (all cited in Hastings et al., 2008).

One further study comes from Kiehl et al. (2001). Below I give a brief description of this study as an example of studies investigating functional anomalies in psychopaths. In their (2001) study, Kiehl et al. exposed three groups of people to an emotional memory word task, where the participants processed words of neutral and negative valence. One of those groups consisted of people scoring high and the other two groups of participants scoring considerable lower on the *PCL-R*. One of those groups consisted of criminal non-psychopaths and the other of non-criminal control participants. Kiehl et al. found a reduced amygdala response in the high-scoring group during the processing of words of *negative* valence (e.g. 'death', 'rape'). Their study resulted in the conclusion that (compared to criminal non-psychopaths and non-criminal control participants), criminal psychopaths showed significantly less affect-related activity in the *limbic and paralimbic systems*³¹ (NB: not just in the amygdala; also in the hippocampus, parahippocampal gyrus, ventral striatum

listening'. *Journal of Abnormal Psychology*. 93 (2), pp. 141–149.

²⁷ Williamson, S., Harpur, T.J., Hare, R.D. (1991). 'Abnormal processing of affective words by psychopaths'. *Psychophysiology*, 28, pp. 260–273.

²⁸ Hare, R.D. (1978). 'Electrodermal and cardiovascular correlates of psychopathy'. In Hare, R.D. & Schalling, D. (eds). *Psychopathic behavior: Approaches to research* (Chichester, England: Wiley), pp. 107-144.

²⁹ Patrick, C.J., Bradley, M.M. & Lang, P.J. (1993). 'Emotion in the criminal psychopath: startle reflex modulation'. *Journal of Abnormal Psychology*, 102 (1), pp. 82–92.

³⁰ Christianson, S., Forth, A.E., Hare, R.D., Strachan, C., Lidberg, L. & Thorell, L. (1996) 'Remembering details of emotional events: A comparison between psychopathic and nonpsychopathic offenders'. *Personality and Individual Differences*, 20 (4), pp. 437–443.

³¹ The 'limbic' and 'paralimbic' areas of the brain are located below the corpus callosum (the primary route of communication between the two cerebral hemispheres); the limbic system is directly beneath the corpus callosum; the paralimbic system more or less embraces the limbic system spatially. Both the limbic and paralimbic system are associated with emotions and motivations.

and the anterior and posterior cingulate gyri.)³² Kiehl et al. drew a few other conclusions on basis of the data of their study, amongst which the following one: while there is clear underactivation in the limbic and paralimbic system of psychopaths while processing words like ‘death’, that is: emotionally loaded words of negative valence, there is extensive activity elsewhere in the brains of psychopathic individuals while processing these words. What Kiehl and colleagues found is that psychopaths show greater activation for affective than for neutral stimuli in a number of brain regions, that are associated with semantics and decision-making (Kiehl et al. 2001). These results are consistent with the hypothesis that criminal psychopaths employ nonlimbic cognitive strategies to process affective material (Williamson et al 1991³³, cited in Kiehl et al. 2001). This may mean that the absence of appropriate limbic input regarding the affective characteristics of stimuli forces psychopathic individuals to use alternative cognitive operations and/or strategies to process affective material (see Kiehl et al, 2001).

There is a fairly large number of other studies on psychopaths and affective processing of language (see Blair, Mitchell & Blair, 2005: chapter for a description of some studies). The objects of those studies sometimes differ a little, but quite a few of them offer support for the conclusion that while for most typical human agents language has the capacity to elicit powerful emotional feelings, for a psychopath a word is just a rather dry word. While in typical human agents, the word *cancer* for example evokes not only a clinical description of a disease and its symptoms but a sense of fear, apprehension, or concern, and perhaps disturbing mental images of what it might be like to have it, for psychopaths it has no more than a dictionary meaning (see Hare, 1993: 131).

³² As to paralimbic system dysfunction, compare Kiehl (2006), although see Blair (2005) for some critical remarks as to paralimbic system dysfunction.

³³ Williamson S, Harpur T.J., Hare R.D. (1991). ‘Abnormal processing of affective words by psychopaths’. *Psychophysiology*, 28, pp. 260–273

7.4. Psychopathy as a genetic deficit

Many researchers and clinicians think of psychopathy as a disorder that has a genetic basis (see e.g. Blair, Mitchell and Blair, 2005; *cf* Hare, 1993). Researchers in behavioural genetics are increasingly using the tools of molecular genetics to extend upon discoveries from twin, family, and adoption studies when it comes to the genetic foundation of antisocial spectrum disorders and psychopathy (see Gunther, Vaughn, Philibert, 2010). Perhaps a few words about molecular genetics. Molecular genetics as a scientific domain studies the structure and function of genes at a molecular level. Their scientific results help to understand psychopathy and do so on top of twin, family and adoption studies that provide insight into whether certain character traits or behaviour are (primarily) caused by a genetic blueprint or by some environmental influence.

There is one very important remark that needs to be made about any influence of genes on psychopathy –or broader- anti-social spectrum disorders: it is extremely unlikely that there is a direct genetic contribution to *specific* anti-social *behaviours* (Ellis & Walsh 1997; Blair, Mitchell and Blair, 2005: 29; Blair, Peschardt, Budhani, Mitchell and Pine, 2006). It is naïve to believe this. Nevertheless, there could still be numerous genes that influence how the brain works in ways that increase (or decrease) the probability of criminal behaviour (Ellis & Walsh, 1997). Where genetics are likely to play a role is in determining the probability that the individual will learn an anti-social strategy to achieve her goal (Blair, Mitchell and Blair 2005: 29; Blair, Peschardt, Budhani, Mitchell and Pine, 2006; Ellis & Walsh, 1997). Many have argued that the particular emotional experience of psychopaths make them more likely to learn anti-social strategies to reach their goals (Blair, 1995; Eysenck, 1964³⁴; Trasler, 1973³⁵; Lykken 1995³⁶; cited in Blair et al, 2006).

³⁴ Eysenck, H.J. (1964). *Crime and Personality*. London: Routledge & Kegan Paul.

³⁵ Trasler, G.B. (1973). 'Criminal behaviour'. In Eysenck (ed.), *Handbook of abnormal psychology* (London: Pitman), pp. 67–96.

Having this in mind, there is valuable information coming from molecular genetics on anti-social behaviour. Molecular geneticists have identified several genes relevant to anti-social behaviour and (even) psychopathy in particular. As explained by Gunter, Vaughn and Philibert (2010):

Of particular interest to researchers examining antisocial spectrum disorders [that covers what we can call psychopathy from a perspective of clinical practice] have been genes related to serotonergic and dopaminergic activity. Examples of genes related to metabolism include catechol-O-methyl transferase (COMT), monoamine oxidase (MAO-A), and dopamine beta hydroxylase (DBH); those related to receptor morphology include dopamine receptor D2 (DRD2), dopamine receptor D4 (DRD4), serotonin receptor 1B (5HTR1B), and serotonin receptor 2A (5HTR2A); and those related to transporter activity include the serotonin transporter (in particular a polymorphism in the promoter region of the gene referred to as the serotonin transporter linked polymorphic regions or 5HTTLPR) and dopamine transporter (DAT).

As mentioned by Gunter et al. current scientific evidence seems to point to MAO-A and 5HTT as the most promising sites for additional inquiry for antisocial spectrum disorders and psychopathy” (Gunter et al., 2010).³⁷ It needs to be said that molecular genetics on psychopathy is still in its infancy. It is far from clear yet how to understand the development of psychopathy on the level of molecular genetics. That is, if psychopathy has indeed a genetic basis. Something which many researchers believe. More on this in § 9.

³⁶ Lykken, D.T. (1995). *The antisocial personalities*. Hillsdale, NJ: Erlbaum.

³⁷ The MAO-A (Mono-amine oxidase A) gene is responsible for the production of the MAO-A enzyme. The MAO-A enzyme regulates the chemical breakdown of various neurotransmitters which are a kind of chemical messengers that enable neurons to communicate with each other. MAO- A is involved in the regulation of many neurotransmitters, but most notably the neurotransmitters serotonin and dopamine (Zuckerman 1994: 295; cited in Ellis & Walsh, 1997). 5HTTLPR is a specific part of a gene that codes for the transportation of serotonin. It looks like serotonin acts as an inhibitor of anti-social impulses (Barr & Quinsey, 2008:276).

7.5. Psychopathy as an inhibition problem

Impulsivity...relates to 'bottom-up' kicking of the brain into emotional gear by the limbic system. But it also relates to the ability to exert 'top-down control over those emotions once they have been kicked into gear. (Oakley, 2007: 199)

According to some researchers psychopathy should be understood as an inhibition problem rather than anything else. What these researchers propose is that psychopaths are not limited by the inhibitory mechanisms that restrain the behaviour of most people. This account is perfectly compatible with the other mentioned conceptions, but it clearly puts the emphasis on the inhibition peculiarity. This inhibition peculiarity/problem is supposed to cause psychopaths to behave in an impulsive, irresponsible and reckless manner (see Barr & Quinsey, 2008). Before deepening this conception of psychopathy a little more, let me make the following two remarks.

First, researchers advocating the Hare conception of psychopathy seem to think that the feature 'impulsivity' is modestly relevant to the construct of psychopathy. (Note that impulsivity is an item of the PCL-R.) I take the idea to be that this feature 'impulsivity' is relevant, though not part of the core of the psychopathy construct. As we have seen in § 6.6, typically emotional anomalies are considered as representative of the core of psychopathy. Many Hare advocates may want to see emotional anomalies (either one emotional anomaly or all of them) as fully exhausting psychopathy's core, while others may want to add other features, e.g. manipulative behaviour and a grandiose sense of self to the core of the construct.

Here is the second remark. A distinction of relevance when thinking about psychopathy as an impulsivity disorder is the distinction between *reactive aggression* and *instrumental aggression*. In reactive aggression (also referred to as impulsive or affective aggression), a frustrating or threatening event triggers the aggressive act. Importantly, the aggression is initiated without regard for any potential goal (for example, gaining the victim's

possessions or increasing status within the hierarchy). Reactive aggression is seen as the ultimate mammalian response to threat. The idea is this. When perceiving a threat, dependent on the distance of the threat, mammals show different types of responses. When the threat is at great distance mammals typically freeze. When the threat is somewhat closer they show a flight response. And when the threat is very near, mammals show reactive aggression (Blanchard et al., 1977; cited in Blair et al. 2005: 96). When it comes to human beings, reactive aggression shows up in many different types of anti-sociality. Reactive aggression is also seen in individuals exemplifying PCL-R- psychopathy.

Instrumental aggression is different from reactive aggression in that it is not an immediate response to a frustrating or threatening event, but rather a form of aggression that is used as a means to achieve a specific desired goal. Instrumental aggression, unlike reactive aggression, is purposeful and goal-directed (Berckowitz, 1993³⁸; cited in Blair, Michel and Blair, 2005). Bullying is considered to be an example of instrumental aggression (Blair et al., 2005). It seems to be instrumental rather reactive aggression that is a marked feature of psychopathy and it also seems to be instrumental aggression that separates psychopaths from a rather wide variety and large number of other severely anti-social human beings, many of whom present with reactive aggression.

Let us now return to Oakley's claim about impulsivity. We have seen that she argues that:

Impulsivity...relates to 'bottom-up' kicking of the brain into emotional gear by the limbic system. But it also relates to the ability to exert 'top-down control over those emotions once they have been kicked into gear. (Oakley, 2007: 199)

What would this mean in the case of psychopaths? Applied to an assumed deficiency in self-regulation this idea means that we should conceptually separate strong inclinations

³⁸ Berckowitz, L. (1993). *Aggression: its causes, consequences, and control*. Philadelphia, PA: Temple University Press.

psychopaths have and a deficiency in their mind/brain that makes it exceptionally hard (up to impossible) to control them. We would think of impulsivity and self-regulation in psychopaths in terms of the dominance hierarchy of, on the one hand, a properly functioning or dysfunctional prefrontal cortex (necessary to exercise control) and on the other hand, activity in brain parts responsible for conation. Strong bottom-up input would require a psychopath with a well-functioning prefrontal cortex to prevent impulsive behaviour.

It seems reasonable to think that in psychopaths the sort of (dominance) relationship between certain brain parts responsible for bottom-up motivational input and the prefrontal cortex involved in top-down control can make a real difference in terms of being a successful psychopath who achieves his highly egoistic goals and one who fails to do so because of some sort of impulsivity. This hierarchy could also make a crucial difference when it comes to being a successful or unsuccessful psychopath in terms of escaping the criminal justice system.

7.6. Psychopathy as an adaptive reproductive strategy

What we can add to the current list of psychopathy conceptions - as a final idea - is the thought that psychopathy is the outcome of a 'genetic adaptive reproductive strategy'.

There are several closely related theories in the area of gene-based evolutionary theories. Such theories assume, centrally, that genetic factors predispose people to varying degrees to criminal/antisocial behaviour and that natural selection has operated on human populations and subpopulations to favour varying tendencies towards such behaviour. Below I discuss the 'Cheater'-theory and a little more broadly the 'Cad versus dad theory'.

In evolutionary theories, claims have been made that behaviours as deception and cheating can often be advantageous to reproduction. The 'Cad versus Dad theory' assumes

that males have been naturally selected to make lower parental investment in their offspring than women. The theory assumes as well that women have been naturally selected for choosing mates who will make high parental investments. The theory subsequently supposes that one result of these competing natural selection forces has been to split males into two sub populations. One sub population more or less complies with female preferences for males who make high parental investments, (although still not as high as most women would prefer). These are the 'dads'. The others, the 'cads' or cheaters merely mimic high-investing males and use devious tactics to opportunistically secure numerous sex partners. According to the cheater theory, these devious tactics often include the use of violence, chronic deception, and 'get rich quick' tactics, all actions that can be very hurtful to others (Ellis & Walsh, 1997).

The Cad versus Dad theory with its room for creatures who deceive provides an explanation of cheating behaviour by assuming that it fits well with the dynamics of a population in terms of reciprocity and cooperation. Ellis and Walsh (1997) argue:

If genes promoting altruistic and cooperative behaviour can evolve by natural selection (Rushton et al., 1986³⁹), genes could also evolve that predispose organisms to take advantage of the altruistic and cooperative behaviour of others without reciprocating (Badcock, 1986⁴⁰; Thompson, 1980⁴¹; Thornhill, 1979⁴²). Through this escalating set of social relationships may evolve reproductive strategies rooted in deception and cheating (Cosmides and Tooby, 1992⁴³; Ellis, 1990['b'

³⁹ Rushton, J. P., Fulker, D., Neale, M., Nias, D., Eysenck, H.J. (1986). 'Altruism and aggression: the heritability of individual differences'. *Journal of Personality and Social Psychology*, 50, pp. 1192-1198.

⁴⁰ Badcock, C. R. (1986). *The Problem of Altruism*. Oxford: Basil Blackwell.

⁴¹ Thompson, P.R. (1980). 'And who is my neighbour? An answer from evolutionary genetics'. *Social Science Information*, 19, pp. 341-384.

⁴² Thornhill, R. (1979). 'Adaptive female-mimicking behavior in a scorpionfly'. *Science*. 205, pp. 412-414.

⁴³ Cosmides, L. & Tooby, J. (1992). 'Cognitive adaptations for social exchange'. In Barkow, Cosmides & Tooby (eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (New York: Oxford University Press), pp.163-228.

omitted]⁴⁴). ...In a number of social species an 'arms race' of deception/cheating, and attempts to detect and foil such tactics, may have given rise to the evolution of retaliatory tactics, which may exert pressures for the evolution of even more subtle forms of deception and cheating (Clutton-Brock & Parker, 1995⁴⁵). (Ellis and Walsh, 1997: 233)

Something similar is argued for by Linda Mealey (1995):

Those who use a deceptive strategy and defect after signaling cooperation are usually referred to as 'cheaters', and, as many authors have pointed out...⁴⁶, the presence of cheaters can lead to a coevolutionary arms race in which potential cooperators evolve fine-tuned sensitivities to likely evidence or cues of deception, while potential cheaters evolve equally fine-tuned abilities to hide those cues. (Mealey, 1995: 525)

In her (1995), Linda Mealey argues for two very different types of anti-social personalities. One is the outcome of frequency-dependent, genetically-based individual differences in use of a single (antisocial) strategy. To this type of personality she refers as 'primary sociopathy'. The other type she argues for is the outcome of individual differences in developmental response to the environment, resulting in the differential use of cooperative or deceptive social strategies. She refers to this type as 'secondary sociopathy'. About 'primary sociopathy' (a term which seems to match a concept that fits rather well with the dominant conception of psychopathy) she argues:

⁴⁴ Ellis, L. (1990). 'The evolution of collective counterstrategies to crime: from the primate control role to the criminal justice system'. In: Ellis & Hoffman (eds.) *Crime in Biological, Social, and Moral Contexts*. New York: Praeger,

⁴⁵ Clutton-Brock, T.H. & Parker, G.A. (1995). 'Punishment in animal societies'. *Nature*, 373, pp. 209-216.

⁴⁶ See Mealey (1995) for a list of authors.

To the extent that we understand it now, primary sociopaths come from one extreme of a polygenic genetic distribution and seem to have a genotype that disposes them "to acquire and be reinforced for displaying antisociality" (Rowe 1990: 122⁴⁷). That genotype results in a certain inborn temperament or personality coupled with a particular pattern of autonomic arousal which, together, seem to design the individual (1) to be selectively unresponsive to those environmental cues necessary for normal socialization and moral development and (2) to actively seek the more deviant and arousing stimuli within the environment. (Mealey, 1995: 530)

Mealey sees primary sociopaths as individuals that have a genotype that disposes them "to acquire and be reinforced for displaying antisociality"(Rowe 1990: 122⁴⁸; cited in Mealey, 1995: 536). Importantly, such individuals progress normally in terms of cognitive development and will acquire a theory of mind. Their theory of mind however, will be formulated purely in instrumental terms, without access to the empathic understanding that most of us rely on so much of the time. Without love to 'commit' them to cooperation, anxiety to prevent 'defection', or guilt to inspire repentance, these people remain free to continuously play for the short-term benefit in the Prisoner's Dilemma (Mealey, 1995: 536).

In relation to 'secondary sociopathy' Mealey argues that it is expressed by individuals who are not extreme on the genetic sociopathy spectrum, but who, because of their exposure to environmental risk factors, pursue a life strategy that involves *frequent, but not necessarily emotionless cheating*. Unlike primary sociopaths, secondary sociopaths will not necessarily exhibit chronic antisocial behaviour, because their strategy choices will be more closely tied to age, fluctuation in hormone levels, their competitive status within their referent group, and changing environmental contingencies (Mealey, 1995: 536-7).

⁴⁷ Rowe, D.C. (1990). 'Inherited dispositions toward learning delinquent and criminal behavior: new evidence', In Ellis & Hoffman (eds.) *Crime in biological, social, and moral contexts*. Westport: Praeger.

⁴⁸ Rowe, D.C. (1990). 'Inherited dispositions toward learning delinquent and criminal behavior: new evidence', In Ellis & Hoffman (eds.). *Crime in biological, social, and moral contexts*. Westport: Praeger.

What can be derived from the current section on psychopathy as an adaptive reproductive strategy is that according to some researchers the process of natural selection has led to a creation of a niche for individuals who are deceitful, manipulative, lacking in empathy and remorse, and who are willing to exploit any and all opportunities to gratify their own needs and desires. This is a radical departure from the idea that psychopathy is a personality disorder in the sense that gene-based evolutionary theories conceive of psychopathy as uncommon, but in no way normatively *abnormal* (see Barr & Quinsey, 2008: 275).⁴⁹

8. Psychopathy in women

Currently, literature on psychopathy in women is quite limited. The body of research is growing, but as a whole it is not a very mature field of research yet. Apart from the limitations in quantity, there are some limitations in quality, or at least there are research issues one should keep an eye on. Let me mention two important ones. First, different studies on psychopathy in females have used different populations of women (as research subjects). This makes it often impossible to legitimately generalize findings. Secondly, studies on female psychopathy have used different instruments for the measurement of psychopathy. This also makes it hard to generalize findings. Apart from the issue of different measurement tools being used in different studies on 'female psychopathy', there is another issue that needs to be addressed in regard to matters female psychopathy and measurement tools. An issue having to do with the 'validity' of measurement instruments.

A measurement instrument is valid, when, in case properly used (by a trained clinician and as the manual of the measurement instrument prescribes), it measures the thing

⁴⁹ Note that if psychopaths cannot be said to be naturally defect in some significant way (on natural defects see Foot 2001, Voorhoeve 2009, Dupré, 1998, Jacobs & Walter, 2011), while it is also implausible that they can be said to suffer from an environmental failure, then such people seem good candidates for limiting an account of moral objectivity.

it is supposed to measure and not something else. Unless an instrument has been properly validated it should not be used in clinical settings. When it comes to the PCL-R, Hare, in the manual of the first edition of the PCL-R, states that the use of this instrument in clinical settings should be limited to “those populations in which it has been fully validated. For the presence this means adult male forensic populations” (Hare, 1991⁵⁰: 5, cited in Vitale & Newman, 2001). Note that in the manual of the *second* edition of the PCL-R Hare argues that “standard PCL-R scores have much the same meaning, with respect to the construct of psychopathy, in several different groups [including female offenders] and settings” (2003: 75). Be that what Hare argues, some researchers have responded to this claim of Hare by arguing that “the grounds for this assumption have yet to be demonstrated” (Logan, 2009⁵¹: 3; cited in Kreis, 2009). These researchers propose that when it comes to measurement of psychopathy, PCL-R scores might not have the same meaning for different human *sub*populations, e.g. women. Perhaps a score of 25 or 30 means something different for males compared to females.

A related thought some researchers have is that it cannot be excluded that psychopathy manifests itself differently in males and females.⁵² On a broad abstract level, there are two ways in which psychopathy could manifest itself differently in males and females. First, it could be the case that both males and females share something highly characteristic that is the essence of the disorder (e.g. and plausibly a lack of empathy) while males and females differ on features that are somewhat up to significantly less important in terms of the intrinsic nature of the disorder. It could e.g. be the case that aggression and violence is more typical for/ in male psychopaths (this could have an evolutionary

⁵⁰ Hare, R.D. (1991). *Manual for the Hare Psychopathy Checklist-Revised*. Toronto, Ontario: Multi-Health Systems.

⁵¹ Logan, C. (2009). ‘Psychopathy in women: Conceptual issues, clinical presentation and management’. *Neuropsychiatry*, 23 (special issue), pp. 25-33.[Kreis gives page numbers 1-9, but this seems incorrect].

⁵² Note that theoretically it is possible that there are no female psychopaths, while there are males. That could e.g. be the case if a lack of empathy or predispositions for empathy is crucial to the psychopathy construct and when there are no females who have not a minimal sense of empathy or else no predispositions for the development of empathy. Some evolutionary theory might be able to account for a difference between males and females here.

explanation) while manipulation is more common in females (this could also have an evolutionary explanation). Secondly, it could be the case that a particular trait that is characteristic of both male and female psychopathy manifests itself differently at a behavioural level. Women e.g. might have different manipulation techniques compared to men.

We can helpfully understand the above two points about different manifestations of psychopathy in males and females plus the concerns about the PCL-R's validity for the female population by means of the following paragraph from Kreis & Cooke (2011).

If psychopathic women generally use more subtle and relational ways of exploiting and dominating other people, and if they primarily use aggression in a domestic setting against close intimates and dependents, their harmful acts likely go more unnoticed by the authorities. In contrast, greater and more public use of physical aggression, and more overt dominance and status-seeking, would make psychopathic men and their harmful acts more noticeable, and more likely to be officially recorded. Measures (e.g., the PCL-R) that rely strongly on officially recorded criminality and antisocial behavior, and of more male typical presentations of it, are clearly going to miss a great proportion of psychopathic women. (Kreis & Cooke, 2011)

Overall, the important take point to take home is that it could be the case that in the past, present and future clinicians (have) diagnose(d) some women illegitimately as presenting with psychopathy and/or failed to diagnose them as psychopaths, because the PCL-R in its current condition is insufficiently suitable for women (see also Kreis & Cooke, 2011).

9. The etiology of psychopathy

9.1. Introduction

In this section I will comment on the ultimate cause of psychopathy. Is it nature or nurture? In most cases, this is a false dichotomy. Nature and nurture operate together and, as well-formulated by Steve Cole, it is even the case that “social factors can play a significant role in regulating the activity of human genes. DNA encodes the potential for cellular behavior, but that potential is only realized if the gene is expressed - if its DNA is transcribed into RNA and translated into protein” (Cole, 2009). And it is social factors that can play a significant role in whether DNA gets to that point. But even leaving that more sophisticated level of thought aside many would agree that social and genetic factors operate together in the causal story of, say, a particular physical disease or say, introvert or extravert behaviour. But still, somehow there is this debate, and often a heated debate between nature and nurture advocates who know perfectly well that there is some truth in what the other camp advocates. As expressed by John Dupré: “Almost everyone agrees that there is something profoundly wrong with the dichotomy between nature and nurture, and yet it seems stubbornly unwilling to go away” (Dupré, 2003). I think he is right.

When it comes to psychopathy, there are both stronger suggestions and complexities as to its etiology. One complexity comes from the fact that the question as to what psychopathy is exactly, is open to conceptual discussion. Two other (possible) complexities come from the following two facts: (a) there may be core and more peripheral features to the disorder (a view that seems to be shared by many scientists) and (b) one feature may have different manifestations in traits and/or behaviour.

While there are the abovementioned complexities, it seems to be the case that at least the emotional anomalies that characterize psychopathy are strongly a result of nature rather than environment. Before I focus on scientific views that advocate this, I first focus on

environmental causes of crime in general. Many crimes, but probably not the ones committed by psychopaths, are primarily caused by environmental influences.

9.2. Environmental causes of crime

Criminals come to crime in a variety of ways. Here I identify three characteristic ways. First, some criminals learn to do crime by example because they are raised in families of social environments in which criminal behaviour, in one way or another and to one degree or another, is the accepted norm. Consider a criminal who had professional thieves as parents. One of Hare's research objects went already from an early age 'to work' with his father. Paradigmatic examples of these 'subcultural criminals' include the mafia families (Hare, 1993: 84).

Secondly, some criminals can be understood as largely the products of what is known as 'the cycle of violence'. Evidence is emerging to show that victims of early sexual, physical, or emotional abuse frequently become perpetrators of the same as adults. It is not uncommon, for example, to find that child molesters were themselves sexually abused, or for wife assaulters to have witnessed domestic violence at an early age (Hare, 1993: 84). In regard to this point. Blair et al. however argue that "we do not believe, on the basis of the available data, that physical/sexual abuse is a key factor in the genesis of psychopathy" (Blair et al. 2006: 264).

Thirdly, some criminals run afoul of the law because of a powerful need they have. One can think of drug addicts here or about people with hardly any skills or resources who turn to robbery out of desperation (see Hare, 1993: 84).

In short, for many criminals, negative social factors, i.e. e.g. poverty, family violence, child abuse, poor parenting, economic stress, alcohol and drug abuse, to name but a few – were contributors to, or even the cause of, their criminality. Indeed, had these factors *not*

been present, many of these criminals would not have turned to crime (Hare, 1993: 84). For psychopaths however crime seems to be less the result of adverse social conditions than of a character that has a strong genetic component.

9.3. The childhood development of psychopaths

Psychopathy does not suddenly spring, unannounced, into existence in adulthood. (Hare, 1993: 157). In this section, I will approach this claim from the perspective of probable childhood development of a psychopath (NB: the diagnosis psychopathy must not be given to anyone under the age of 18).

As Robert Hare argues “certain children remain stubbornly immune to socializing pressures. They are inexplicably ‘different’ from normal children-more difficult, wilful, aggressive, and deceitful; harder to ‘relate to’ or get close to; less susceptible to influence and instruction; and always testing the limits of social tolerance” (Hare, 1993: 157-8). Many of these children will be diagnosed as children with ‘conduct disorder’ about which more soon.

Most psychopaths begin to exhibit serious behavioural problems at an early age. These might include persistent lying, cheating, theft, fire setting, truancy, class disruption, substance abuse, vandalism, violence, bullying, running away, and precocious sexuality. Because many children exhibit some of these behaviours at one time or another, especially children raised in violent neighbourhoods or in disruptive or abusive families, it is important to emphasize that the psychopath’ history of such behaviours is more extensive and serious than that of most others, even when compared with those of siblings and friends raised in similar settings. An example of the psychopathic child is one who comes from an otherwise well-adjusted family and starts to steal, take drugs, cut school, and have sexual experiences by age ten or twelve.

Cruelty to animals is also something often seen in children with conduct disorder. When he was younger, psychopath Jeffrey Dahmer, for example, stunned classmates and neighbours by leaving a trail of grim clues to his preoccupations: the head of a dog impaled on a stick, frogs and cats staked to trees, and a group of animal skeletons kept as a collection (Hare, 1993: 66).

Below I give an example of a case – the case of John- where the child shows clear signs of the special group Hare refers to, signs that could be an indicator of adult psychopathy. For the sake of contrast, I will also present another case, the case of Bill. On basis of criteria for conduct disorder presented in the DSM-V (2013: 469-476) (as well as the DSM-IV, 2000: 93-99) one would assume that both children present with conduct disorder, while it is the case that they present with significantly different traits and behaviours. Because of the type of behaviour and traits John exhibits, John –unlike Bill- qualifies for a specifier for conduct disorder that is now present in the DSM-V (published in 2013), while not yet present in the DSM-IV. That is, John unlike Bill qualifies for the specifier ‘conduct disorder ‘with limited prosocial emotions’. This is the case of John:

John is an 11-year old boy from a middle-class family with two professional parents. He began to present with behavioural problems at an early age and was enrolled in a school for children with emotional and behavioural difficulties at the age of 5 years. John began running away from home and school at a young age. Now, he is frequently picked up by the police because he is roaming the streets of the local town late at night. He often spends time with local juvenile delinquents. He recently broke into a construction site and set fire to materials, causing \$ 15,000 worth of damage. John is often cruel to animals. He once dangled his pet hamster over a hot stove and threatened to drop it if his parents did not give him money. He is also frequently violent towards his parents, teachers, and peers. On several occasions he has threatened to hurt his mother, and stashed knives are often retrieved from his bedroom. On one occasion he threw a kitchen knife at his mother. John does not have any genuine friends at school. Teachers often express that they have difficulty treating him kindly as they feel that nice behaviors displayed by him are not at all

sincere. He is very boastful about his abilities generally, and has an inflated perception of his intelligence. John sometimes tricks people into thinking that he is simply misunderstood. (Blair, Mitchell and Blair, 2005: 1-2)

Then now for Bill.

Bill is an 11-year old boy from a troubled working class background. His mother and father are both in jail, his father for armed robbery and his mother for drug offenses. He is cared for by his older sister. Bill often presents with oppositional behavior at home and at school. He is rude to teachers, often refusing to complete assignments, and frequently truant. He has stolen merchandise from local shops. He often fights with classmates and has on occasion used a weapon (a brick) in these fights. However, he usually apologizes if he is genuinely to blame. He enjoys playing sports with his classmates. He also often expresses love towards his sister and is comforted when she is present. Bill's emotions can be turbulent. He is often self-deprecating. (Blair, Mitchell and Blair, 2005: 2)

As said, both children qualify for the diagnosis 'conduct disorder'. For this diagnosis to be given legitimately, the DSM V requires that the individual presents with at least three of the following 15 criteria. That is, in the past 12 months at the moment of diagnosis and with at least one criterion present in the past 6 months:

1. Often bullies, threatens or intimidates others.
2. Often initiates physical fights.
3. Has used a weapon that can cause serious physical harm (e.g., a bat, brick, broken bottle, knife, gun).
4. Has been physically cruel to people.
5. Has been physically cruel to animals.
6. Has stolen while confronting a victim (e.g., mugging, purse snatching, extortion, armed robbery).

7. Has forced someone into sexual activity.
8. Has deliberately engaged in fire setting with the intention of causing serious damage.
9. Has deliberately destroyed others' property (other than by fire setting).
10. Has broken into someone else's house, building, or car.
11. Often lies to obtain goods or favors or to avoid obligations (i.e., 'cons' others).
12. Has stolen items of nontrivial value without confronting a victim (e.g. shoplifting, but without breaking and entering; forgery).
13. Often stays out at night despite parental prohibitions, beginning before age 13 years.
14. Has run away from home overnight at least twice while living in the parental or parental surrogate home, or once without returning for a lengthy period.
15. Is often truant from school, beginning before age 13 years (DSM-V: 469-70).

Note that on top of the condition that an individual presents with three of the abovementioned characteristics, in order for the diagnosis conduct disorder to be given rightfully, it must be the case that the individual's disturbance in behaviour causes clinically significant impairment in social, academic or occupational setting. Finally, it must be the case that, if the individual is 18 years or older, criteria are not met for antisocial personality disorder (DSM-V: 470).

Let us return to the cases of John and Bill. Both John and Bill have engaged in at least three of the diagnostic criteria for conduct disorder. Both Bill and John often engage in fights; have on occasion used weapons and have truanted. However while John presents with psychopathic tendencies Bill does not (see Blair, Mitchell and Blair, 2005).

9.4. 'A strong genetic component'

While in the previous section I have focused on the childhood development of psychopaths, in this section I will look at the relationship between psychopathy and nature as an ultimate cause of it. Let us first look at the following remark from Robert Hare:

[P]erhaps the most popular generalization about psychopathy- [is] that it is the result of early psychological trauma or adverse experiences: poverty, emotional or physical deprivation or abuse, parental rejection, inconsistent disciplinary techniques, and so on. Unfortunately, the picture that emerges from clinical experience is far from clear on the matter. On balance, however, I can find no convincing evidence that psychopathy is the direct result of early social or environmental factors. (Hare, 1993: 170)

Let us also look at the following longer statement from Hare:

Evidence of the genetic and biological bases of temperament⁵³, the ability of some forms of brain damage to produce psychopathiclike symptoms, and the early appearance of psychopathic behaviours in children provide frameworks for several biological theories on the origins of psychopathy.

The position I favour is that psychopathy emerges from a complex-and poorly understood-interplay between biological factors and social factors. It is based on evidence that genetic factors contribute to the biological bases of brain function and to basic personality structure, which in turn influence the way the individual responds to, and interacts with, life experiences and the social environment. In effect, the elements needed for the development of psychopathy-including a profound inability to experience empathy and the complete range of emotions, including fear-are provided in part by nature and possibly by some unknown biological influences in the

⁵³ "Think about some relatives or friends you have known since childhood: the shy, inhibited girlfriend; the outgoing, gregarious brother; the fast-talking, sleazy cousin; the wild, hostile, aggressive neighbour. What were they like when they were ten years old?" (Hare, 1993: 97).

developing fetus and neonate. As a result, the capacity for developing internal controls and conscience and for making emotional ‘connections’ with others is greatly reduced.

This doesn’t mean that psychopaths are destined to develop along a fixed track, born to play a socially deviant role in life. But it does mean that their biological endowment—the raw material that environmental, social, and learning experiences fashion into a unique individual—provides a poor basis for socialization and conscience formation.....

Although psychopathy is not primarily the result of poor parenting or adverse childhood experiences, I think they play an important role in shaping what nature has provided. Social factors and parenting practices influence the way the disorder develops and is expressed in behaviour.

Thus, an individual with a mix of psychopathic personality traits who grows up in a stable family and has access to positive social and educational resources might become a con artist or white-collar criminal, or perhaps a somewhat shady entrepreneur, politician, or professional. Another individual, with much the same personality traits but from a deprived and disturbed background, might become a drifter, mercenary, or violent criminal.

In each case, social factors and parenting practices help to shape the behavioural expression of the disorder, but have less effect on the individual’s inability to feel empathy or to develop a conscience. No amount of social conditioning will by itself generate a capacity for caring about others or a powerful sense of right and wrong. ...[P]sychopathic ‘clay’ is much less malleable than is the clay society’s potters usually have to work with. (Hare, 1993: 173-4; *cf*178)

What becomes clear from this passage is that according to Hare nature has a strong influence on the development of psychopathy. It is not fully clear however how we should understand this. This is so because it very much looks like Hare (in the above passage) wants to say that biological influences are *vital to being able to perform social behaviour*. Support for this view comes from Hare’s claim that “[n]o amount of social conditioning will by itself generate a capacity for caring about others or a powerful sense of right and wrong.” If Hare indeed holds the view that biological influences are vital to being able to perform social behaviour,

then he probably would also hold that in case those biological influences are absent this means that someone does not have the potential to perform social behaviour. And that seems to mean that the only option that is open is anti-social behaviour. And it is exactly that view Hare seemingly does not want to commit himself to, at least not in the passage above. After having said,

[P]sychopathy emerges from a complex-and poorly understood-interplay between biological factors and social factors. It is based on evidence that genetic factors contribute to the biological bases of brain function and to basic personality structure, which in turn influence the way the individual responds to, and interacts with, life experiences and the social environment. In effect, the elements needed for the development of psychopathy-including a profound inability to experience empathy and the complete range of emotions, including fear- are provided in part by nature and possibly by some unknown biological influences in the developing fetus and neonate. As a result, the capacity for developing internal controls and conscience and for making emotional 'connections' with others is greatly reduced.

Hare argues: "This doesn't mean that psychopaths are destined to develop along a fixed track, born to play a socially deviant role in life".

There seems to be something to explain for Hare as to how exactly psychopaths are *contra-prepared for the performance of social behaviour* and *biologically prepared for the performance of anti-social behaviour*. Not just in terms of what sort of biological bits and pieces fundamentally contribute to the performance of social and anti-social behaviour, but also in terms of what this means in regard to how plastic the mind of psychopaths is in terms of becoming social while now being a psychopath. Interpreting Hare generally rather than just from the perspective of the passage just discussed, it seems fair to say that Hare is not optimistic in terms of psychopaths being able to change from an anti-social human being into a social one (recall also Hare's remarks in § 4.7 above about psychopaths being irresponsive to treatment) . What Hare might want to say is that for some or all psychopaths, there were

some early developmental possibilities to walk social pathways of life. If Hare would want to hold such a view, he would probably want to do that while also saying that despite the fact that there were some early developmental opportunities for psychopaths to live socially, it is still the case that they are contra-prepared for walking these pathways life.

Generally, there is still a lot of work to do for scientists to understand how psychopaths' constitution causes them to be who they are and to behave how they do. A promising account comes from James Blair (2006, see esp. 434-5; *cf* Blair, Mitchell and Blair 2006). Blair suggests that genetic anomalies reduce the salience of punishment information. Blair tentatively suggests that the reduced receptivity to punishment information might be a function of noradrenergic disturbance (caused by genetic anomalies). The suggestion Blair offers is that this noradrenergic disturbance has an effect on various aspects of amygdala function, most importantly the ability to form stimulus–punishment associations, including stimulus punishment associations related to punishment by a victim's distress. Blair assumes that children who socialize successfully form associations between representations of moral transgressions (acts which harm others) and the aversive 'punishment' caused by the victim's distress. This allows them to learn to avoid actions that will harm others. As individuals with psychopathy find the distress of the victim significantly less aversive, or so Blair's assumption I find plausible goes, they are less likely to learn to avoid actions that will harm others.

10. Looking ahead

We have seen that the classification of psychopathy, unlike the diagnoses of conduct disorder or antisocial personality disorder, identifies a relatively homogeneous population. There is a unitary disorder upon which a causal account can be developed. And on the basis of current research, there is reason to think that there is a genetic and not a social ultimate cause to the

core of the syndrome (*cf* Blair et al., 2006). Clearly there is a lot of unclarity as to what it means that there is a strong genetic component to psychopathy, but nevertheless there seems to be such a thing. Given that psychopaths seem to form a relatively homogeneous population and given that the core of the syndrome seems to have an innate basis, do we have to conclude from that that all psychopaths are people born with a ‘black heart’, say, biologically determined to think and act anti-socially’?

In the final chapter, I suggest that all psychopaths are constitutionally ‘contraprepared’⁵⁴ to take the interests of other people seriously in their practical judgements and actions. I furthermore suggests tentatively that a few and probably only a few psychopaths (as in line with the PCL-R diagnostic criteria) are ‘contraprepared’ to such an extent that their constitution provides them with a phenomenological space of possibilities that excludes the mental possibility to take the interests of others seriously in practical thought. These views will form the basis the second and third out of three criteria I present in this study that speak in favour of there being an anti-realist component to moral objectivity. The other, first, criterion revolves around Hume’s conception of reason. It is to Hume’s philosophy we now turn.

⁵⁴ I borrow this term from (Mealey, 1995: 533)

Chapter 2: Hume on reason and morality

1. Introduction

In this chapter I investigate a wide variety of aspects of Hume's philosophy. This will be essential to the identification of one criterion for there being an anti-realist component to moral objectivity. The chapter set up is as follows. In § 2, I explain Hume's overall philosophical project. I explain Hume's philosophical activity as a cognitive psychologist and focus on his distinction between impressions and ideas.

In § 3, I present a number of scholarly views on Hume's notion of 'reason'. In § 3.1, I focus on scholarly views having to do with a conception of 'reason' as a mental process/activity of '*reasoning*'. In § 3.2, I present, examine and evaluate scholarly views on 'reason' in relation to (i) the mental *act* and (ii) the mental *product* of '*judgement*'. In § 3.2, I discuss and evaluate the view that according to Hume 'reason' is a mental '*faculty*'.

In § 4, I focus on Hume's ideas about 'reason' in relation to Hume's ideas about mathematics and in § 5, I discuss Hume's ideas about 'matters of fact'. The discussion in § 4 is meant to help understand the relationship between Hume's conception of 'reason' and his (and other philosophers') thoughts about the *a priori*. The discussion in § 5 is meant to aid understanding of the relationship between 'reason' and 'impressions'. Both discussions should serve a good understanding and sensible evaluation of Hume's claim that 'reason cannot make *moral distinctions* without contributions of passions'.¹ A good understanding and

¹ Hume also makes the infamous claim that "reason alone can never be a motive to any action of the *will*" (T: 413; my italics). My focus in this chapter will mostly be on the claim pointed to in the main text however, because it is more relevant to my aim of identifying criteria for there being an anti-realist component to moral objectivity. I take it that the claim in the main text is also much harder to understand than Hume's claim about the limits of reason in terms of its capacity to influence the will. Once we understand the claim pointed to in the main text, the extra bit required to explain Hume's claim about reason and the will seems to consist in the rather uncontroversial Humean claim that 'passions' have an influence on the will. Be that as it easy as it seems to be, on a more detailed level there is a lot to say about Hume's claim that passions have an influence on the will. Also, there is a lot to say about how 'passions' operate together with 'reason' in influencing the will. I engage with the first matter (albeit often implicitly rather than explicitly) in § 7, 9 and 11, where I discuss Hume's account of passions; his account of 'sympathy'; and his views about our innate constitution, respectively. I focus on the second matter in § 8, where I discuss Hume's

evaluation of that particular claim then is crucial for determining the first criterion for there being an anti-realist component to moral objectivity.

In § 6, I discuss the relationship between ‘reason’ and ‘truth’ in Hume’s work. In § 7, I explain Hume’s idea that passions are ‘original existences’. I also comment on other aspects of Hume’s view on ‘passions’. In § 8, I explain how Hume’s ideas on ‘reason’, ‘passion’ (especially as ‘original existences’), ‘truth and falsehood’ come together in his moral philosophy.

In § 9, I explain some further ideas that are important to Hume’s moral philosophy. In § 9.1 I explain Hume’s thoughts on ‘sympathy’. In § 9.2, I discuss Hume’s idea of the ‘common’ or ‘general’ point of view. In § 9.3, I offer a brief discussion of Hume’s ideas on justice and the related notion of ‘utility’. In this section, I also briefly discuss Hume’s ideas on self-interest (largely by commenting on the relationship between Hume’s views on ‘self-interest’ and ‘utility’). In § 10, I discuss Hume’s account of ‘belief’. As we will come to see, Hume had an extremely rich account of belief and one that hardly shows any similarity with the dominant concept of belief operative in contemporary meta-ethics. Importantly, it is out of line with the idea that Hume held that beliefs are intrinsically motivationally inert. I focus on the relationship between Hume’s account and the meta-ethical account of belief in the next chapter. In § 11, I discuss Hume’s thoughts on human nature. I pay attention to his views about our natural constitution and to his views about education and culture.

account of morality. For the second matter see also §10 where I discuss Hume’s account of beliefs. Note that for Hume, ‘passions’ are one type of ‘impression’. Note also that for Hume the influence beliefs have on our mind resemble the force ‘impressions’ have on our mind.

2. Hume's philosophical project

2.1. A general introduction

Tis evident that all the sciences have a relation, greater or less, to human nature; and that however wide any of them may seem to run from it, they still return back by one passage or another. Even *Mathematics, Natural Philosophy, and Natural Religion*, are in some measure dependent on the science of Man; since they lie under the cognizance of men, and are judged of by their powers and faculties... If therefore the sciences of Mathematics, Natural Philosophy and Natural Religion have such a dependence on the knowledge of man, what may be expected in the other sciences, whose connection is more close and intimate? ... There is no question of importance, whose decision is not compriz'd in the science of man: and there is none, which can be decided with any certainty, before we become acquainted with that science. In pretending therefore to explain the principles of human nature, we in effect propose a compleat system of the human sciences, built on a foundation almost entirely new, and the only one upon which they can stand with any security. (T: xv-xvi)

As one might expect from this quotation, Hume's work is permeated by 'science of man'. In the above quotation Hume points out that this 'science of man' is supposed to explain the principles of human nature. Below I explain how exactly this relationship should be understood. In that explanation I include an elucidation of the term 'human nature' in the context of the mentioned relationship. Below I also comment on other meanings of 'human nature' in Hume's work.

Hume's philosophy of human nature is mostly 'cognitive psychology'. As a cognitive psychologist Hume aims at (sometimes detailed) *descriptions* of a variety of mental *states*. As a cognitive psychologist he also aims at typically detailed *causal explanations* of mental states in the form of *psychological mechanisms* that generate those mental states.

Hume's thoughts on human nature are mostly thoughts from a cognitive psychologist.² However, Hume also makes claims about the constitutional make-up (pre-dispositions) fundamentally underlying what Hume typically analyzes: actual psychology. Furthermore, Hume offers views about the *species-typicality* of actual operative mental states and pre-dispositions of these states. Occasionally, Hume also talks about those states in the context of issues having to do with freedom.

That much about the relationship between Hume's philosophical project and 'human nature'. What else is there to say about Hume's philosophical project? There are some important remarks that need to be made about his 'empiricist' program, which is a program with a strong link to the activity of doing 'science of man' in the form of cognitive psychology. It will not come as a surprise to many readers that Hume was an empiricist. But what does that label mean in regard to Hume? And how exactly does this 'empiricism' relate to Hume's wholeheartedly embraced activity of doing 'science of men'?

As a philosopher, Hume wants to deliver some *foundations* of 'knowledge claims' (I explain in a moment how to understand this). However Hume is also very much interested in the *limitations* of what we can know. Both of these aspects come under the form of a merely descriptive empirical science (see Michaud, 1987, esp. 375). The empirical science of the cognitive psychologist Hume that permeates his work reflects Hume's acknowledgement of both the true import and limitations of the empiricist program (see Michaud, 1987: 367). As to the limitations, Hume realized that we philosophers can never go beyond experience and observation when attempting to get philosophical wisdom. We must restrict ourselves to sensory data in our search for philosophical search for foundational knowledge. Now, *since* we can never go beyond experience and observation, there are clear limitations to the kind of conclusions philosophers can legitimately draw. We cannot make existence claims about what transcends experience. According to Hume any principle that transcends what is strictly given in experience must be recognized as such. In regard to such principles we can

² Note that the activity I ascribe to Hume the cognitive *psychologist* for Hume counts as doing *philosophy*.

only state and describe our experience of its reality (see Michaud, 1987: 373). That is the most we can do (*cf*T: 218).

Apart from the abovementioned methodologically limiting side of empiricism, there is a Humean interest in *foundations* of knowledge. While this is an accurate claim about Hume's project, it must be noted that Hume's knowledge claims represent less of a positive approach in terms of active epistemological knowledge construction than a negative strategy for distinguishing between fictions and sensory data (see Michaud, 1987: 372-3)³. Hume is interested in separating reality from fiction, but Hume's philosophical conclusions about reality *do not get beyond sense-data*; they are not about some extra-mental world. Strictly speaking for Hume impressions do not have intentional objects (see Landy, 2006: 137).

Here we encounter the problem as to what type of professional intellectual being is interested in separating reality from fiction. When it comes to Hume's interests in separating reality from fiction, these interests are certainly ones from the *cognitive psychologist* Hume. I take this to be rather uncontroversial. When it comes to him being an *epistemologist*, things are more controversial. In terms of normative views about knowledge, Hume certainly has views about what we *cannot know*, but it also looks like Hume has normative views about there *being* knowledge. The latter is controversial however.

It is important to point out that while facing the interpretative task of determining in what intellectual guise(s) Hume comes when writing about subject matter X, Y and Z, different conclusions seem to hold for (i) *different subject areas of his work* and possibly also for (ii) different published works about the *same subject matter*. As to (i), it seems reasonable to think that Hume held some sort of normative views in the area of mathematics. If he indeed did have such normative views in the area of mathematics, I read Hume as saying that in the area of mathematics *knowledge as scientia* is possible. Assuming for the sake of the argument that normative evaluations can be made in both the area of mathematics and in other areas of Hume's work, for Hume knowledge as *scientia* is *not* possible in other areas,

³ Note that for Hume as a scientist of men, fictions are as natural as sensory data are (see Michaud, 1987: 372-3).

including the intellectual domains of ‘extra-mental objects’, beauty and morality. Assuming *this* for the sake of the argument, this does not imply and in my reading of Hume does not imply that according to Hume *no normative claims can be made* in those areas where there is no such thing as knowledge as *scientia*.

Let me now connect (i), i.e. ‘we should be alert on the fact that different conclusions may have to be drawn for different subject areas’ in terms of Hume being an epistemologist and/or cognitive scientist and (ii), i.e. ‘different conclusions may hold for the different published works about roughly the same subject matter’. As I read Hume, when it comes to matters about *causality* Hume is primarily a cognitive scientist. He is also a cognitive scientist paradigmatically. If we compare and contrast Hume’s work on causality with everything else he ever wrote about, his intellectual guise as a cognitive psychologist shows up most prominently in the area of causality.

There is reason to think that this intellectual attitude of a paradigmatic cognitive scientist that characterizes Hume’s work about causality stands in significant opposition to what Hume writes in his *An Enquiry Concerning the Principles of Morals*. This particular book on *morality* however, *may* stand in significant contrast with book III of his *A Treatise of Human Nature*. Note that part three of the *Treatise* is about morality too and note furthermore that *EPM* and book III of the *Treatise* share much in terms of specific subject matters that relate to morality. This highlights an important and also difficult exegetical point. It *might* be the case that in *EPM* Hume takes a *stronger normative stance* than in the part of the *Treatise* that is about morality. In the *Treatise* he *may* take more strongly the guise of a cognitive scientist interested in (anatomically) dissecting our psychological mechanisms in regard to moral matters. This is a hard exegetical matter for the following reason. The later published book *EPM* is meant to be less anatomical than the *Treatise*. In *EPM* (as well as in *EHU* and some other parts of Hume’s work), Hume’s focus is on *practicality and/or readability*. The challenge is to understand how this difference in Humean philosophical/writing intentions between ‘describing and dissecting’ in the *Treatise*

and 'practicality and/or readability' in *EPM* stands in relation to Hume holding *more normative* views in the latter than in the former. There *may*, but *need not* be a difference between the two works on more or less the same subject matter. The differences in Hume's language *could* represent a difference in terms of him being more of a cognitive psychologist/normative philosopher in one work than in the other, but they might not do so. The opinion that Hume holds a more normative view in *EPM* (where Hume's language is smooth and practically orientated) is an *intelligible view* given that Hume was seemingly interested in the question how practical life should be lived. That Hume does *not* hold more normative views in *EPM* compared to the *Treatise* (assuming for the sake of the argument that in the *Treatise* he is primarily, if not exclusively, a cognitive psychologist) is a view that should be taken seriously, because one needs a justification for thinking that Hume revised or in a major way developed his earlier thoughts. One should not assume this too quickly on basis of the change in language between the two works, if only (but probably not only) because of the following: Hume was disappointed by and dissatisfied with how pretty badly his (fairly dry) anatomical *Treatise* was received. There is good reason, although not conclusive reason, to think that therefore he changed his language while writing about roughly speaking the same subject matter. Assuming that this holds, the question then is whether Hume's decision in regard to *EPM* to come up with something his readers can digest better than the anatomical language of the *Treatise* does or does not represent a *more normative* Humean approach in *EPM*. Note finally that if there is *no* significant difference in terms of Hume taking a normative approach in the two works about morality, this could also mean that book III of the *Treatise* is significantly/primarily/exclusively characterized by normative views. This does not seem to be a good first exegetical move, however, both because of the language used in book III and also because the *Treatise* generally is a strongly cognitive psychological work.

2.2. Impressions and ideas: the atoms of the mind

In the *Treatise*, Hume appeals frequently to two sorts of what one could call ‘mental items’. That is, large parts of Hume’s work are permeated by Humean talk about ‘impressions’ and ‘ideas’. Note that Hume calls all ‘mental items’, i.e. everything that could possibly (from a perspective of Humean psychological realism) pop up in the mind, ‘*perceptions*’. I point out that Hume had a *general* notion of ‘perceptions’ and held a *specific* notion of ‘ideas’. This in contrast to Locke *for whom all mental items were ‘ideas’ and who held a more specific notion of perceptions*. (Hume himself remarks on that on page 2 of the *Treatise*.⁴)

The more accurate version of the claim ‘large parts of Hume’s work are permeated by Humean talk about ‘impressions’ and ‘ideas’ is that for Hume each and every ‘perception of the mind’ resolves itself into one of the two mentioned types of ‘mental items’; every mental perception is either an ‘idea’ or an ‘impression’ (T: 1). For Hume, the difference between ‘impressions’ and ‘ideas’ lies in the degree of ‘force’ and ‘liveliness’, with which they “strike upon the mind and make their way into our thought and consciousness” (T:1). Those perceptions which enter with *most force and violence*, are called ‘impressions’. Under this name Hume comprehends all our “sensations, passions and emotions, as they make their first appearance in the soul” (T:1). In regard to ‘ideas’, Hume argues that these are the “faint images” of impressions “in thinking and reasoning” (T:1); ‘copies’ of impressions so to say.⁵

⁴ Having said that Hume had a ‘specific’ notion of what an ‘idea’ is, I do *not* want to deny that Hume could have been a bit more clear as to his philosophical conception of an ‘idea’. Generally and especially at specific places in his text it would have been helpful had Hume been a little more lucid as to the relationship between his conception of an ‘idea’ and some philosophical conception of ‘ideas’ as ‘mental items of thought’ and ideas as ‘objects of mental items of thought’. When I say that Hume had a ‘specific’ notion of an ‘idea’, I rather want to make clear that Hume’s notion of an ‘idea’ is not as all-encompassing as some other philosophers’ notions of ‘idea’ and – positively phrased – that Hume’s notion represents just one particular type of mental item.

⁵ After having made this clear about ‘impressions’ and ‘ideas’, in the beginning of this *Treatise*, Hume points out that there is another useful distinction to be drawn regarding our ‘perceptions of the mind’. That is, the distinction between what Hume calls ‘simple’ and ‘complex’ impressions and ideas. For Hume ‘simple’ impressions and ideas are “such as admit of no distinction and separation”. “The complex are the contrary to these, and may be distinguished into parts” (T:2). When it comes to impressions, another distinction Hume draws is the one between ‘impressions of sensation’ and ‘impressions of reflection’ (T:7, 84) or (conceptually similar) between ‘original’ and

What is implied by this little story about impressions and ideas is the conclusion that it is impressions rather than ideas that are the *most fundamental* ‘perceptions of the mind’. In other words, they are our ultimate source of mental perceptions. This on its turn regarding Hume’s two-side empiricism, implies two things. First, that impressions form the *limits* of all our claims about knowledge. Secondly, that the impressions form the *very evidential basis* of all our philosophical inquiry and conclusions about knowledge. So, for Hume all theoretical conclusions in some significant way can be *reduced* to what Hume calls ‘impressions’, while the same impressions are the foundation for separating reality from fiction. This, while all that philosophers can legitimately denote as ‘reality’ has its boundaries set by experience from our internal and external senses delivering impressions. Note that for Hume impressions coming from our internal sense seem to have a fundamental and partial foothold in our natural constitution (see § 11).

3.Hume’s views on ‘reason’

3.1 A variety of descriptions

I now make a start with an explanation of Hume’s concept of ‘reason’. There are a rather large number of definitions and descriptions of ‘reason’ Hume offers in his work. Hume’s claims about ‘reason’ include:

- reason is nothing but a wonderful and unintelligible instinct in our souls (T: 179)
- reason alone can never give rise to any original idea (T: 157)
- reason is the discovery of truth and falsehood (T: 458)
- reason alone can never be a motive to any action of the will....*secondly*...it can never oppose passion in the direction of the will (T: 413)

‘secondary impressions’ (T: 275). Impressions of sensation arise in the soul originally, while the second class of impressions result “in a great measure” from our ideas (see T: 7). More about this distinction in § 7.1.

- Thus it appears, that the principle which opposes our passion, cannot be the same with reason, and is only called so in an improper sense (T: 415)
- Reason as distinguish'd from experience, can never make us conclude, that a cause or productive quality is absolutely requisite to every beginning of existence" (T:157)
- We speak not strictly and philosophically when we talk of the combat of passion and reason T: 415)
- And as every quality, which is useful or agreeable to ourselves or others, is, in common life, allowed to be a part of personal merit; so no other will ever be received, where men judge of things by their natural, unprejudiced reason, without the delusive glosses of superstition and false religion (EPM: 270)

This list of rather different descriptions may make one wonder whether there is a single notion of '*reason*' in Hume's work. Whether or not Hume's various descriptions can finally be captured in one single concept is a difficult exegetical matter. The explanation and justificatory reasons for this claim will be revealed gradually as we proceed. I present my own answer to the mentioned problem in § 6.

Whether or not Hume's various descriptions can finally be captured in one single concept, there are anyhow a variety of aspects relevant to Hume's notion or notions of 'reason'. One essential concept and/or aspect of 'reason' in Hume's work is 'reason' as a mental process/activity of *reasoning*. This will be our object of attention now.

3.2. 'Reason' as *reasoning*

The understanding exerts itself after two different ways, as it judges from demonstration or probability: as it regards the abstract relations of our ideas, or those relations of objects, of which experience only gives us information. I believe it scarce will be asserted, that the first species of *reasoning* alone is never the cause of any action. As its proper province is the world of ideas, and as the will always places us in that of realities, demonstration and volition seem, upon that

account, to be totally remov'd from each other. ..'[T]is not of themselves [mathematics and arithmetic] have any influence. (Hume, T: 413; my italics)

In the abovementioned passage I have italicized the word 'reasoning'. Rachel Cohon starts her explanation of Hume's conception of reason by saying that "*It would not be amiss to substitute the gerund 'reasoning' almost everywhere Hume says 'reason'*" (Cohon, 2008: 66). According to Cohon, 'reason' for Hume is 'reasoning': a 'process' or 'mental activity' as applicable to what for Hume is 'demonstrative' and 'causal' reasoning. According to Cohon the former refers to "a process of comparing ideas and finding congruencies and incongruencies" (Cohon, 2010: 10). This, however, is not the only description Cohon gives of 'demonstrative reasoning'. Another description of that notion (or strictly speaking the notion of 'demonstration' (she seems to treat the two as identical) Cohon gives is:

In demonstration we proceed step by step in linking related ideas in sequence, preserving our assent—or the degree to which the ideas are evident—at each step, so that in the end we find evident (indeed, we are necessarily determined to grasp) a relation between our starting and resulting ideas. (*passim*, e.g. T: 95) (Cohon, 2008: 66)

'Causal *reasoning*' for Cohon seems to be identical with 'causal *inference*'. Cohon explicitly describes 'causal inference' as 'inferential discovery of the relation of cause and effect' (Cohon, 2008: 51).

We have seen Cohon's claim that "*It would not be amiss to substitute the gerund 'reasoning' almost everywhere Hume says 'reason'*" (Cohon, 2008:66). Related, we have seen that for Cohon Hume's concept of 'reason' is very much a process of discovering relations of ideas or the mental activity of comparing ideas. But we have also seen that she says that "In demonstration we proceed step by step in linking related ideas in sequence, *preserving our assent*—or the degree to which the ideas are evident—at each step, so that in the end we find evident (indeed, we are necessarily determined to grasp) a relation between our starting and

resulting ideas” (*passim*, e.g. T: 95; Cohon, 2008: 66). The phrase in italics ‘preserving our assent’, seems to ‘bend into the direction’ of a *judgemental outcome* rather than a mental *process/activity of comparing ideas*. But it is clear that Cohon herself is unsympathetic towards relating Hume’s notion of reason to judgemental outcomes rather than reasoning ‘processes’ or ‘mental activities’. To understand this, let us look at a passage from Hume’s *Treatise* cited in Cohon’s work:

[A]ll the actions of seeing, hearing, judging, loving, hating, and thinking fall under this denomination. The mind can never exert itself in any action, which we may not comprehend under the term of perception; and consequently that term is no less applicable to those *judgments*, by which we distinguish moral good and evil, than to every other operation of the mind. To approve one character, to condemn another, are only so many different perceptions. (T: 456; italics added by Cohon; cited in Cohon 2008: 80)

In this passage, Hume uses the term ‘judgement’. As Cohon states, the wording Hume uses here suggests an identification of reason with judgements. It looks like for Hume “reason in a strict sense just is a set of judgments, so that the very judgments themselves are identical with, or are parts of or instances of, reason. And some interpreters think that this is Hume's view” (Cohon, 2008: 69). In contrast to that, Cohon’s view is that in passages like these Hume does *not* refer to some kind of mental ‘product’-cum-‘outcome’. Rather, in these passages, by ‘judgments’ Hume means ‘judgings’—‘acts of judgment’, *not* ‘products’ that come out of a *judging process*.⁶ And those *acts of judgements*, according to Cohon, are ‘judgings’ in the sense that they are mental acts of reasoning that take place to discover relations between ideas. Just as when Hume says that “*Here then reasoning takes place to discover this relation...*”(Hume, T: 414). Let us look at what Cohon says when she argues for the view I have just ascribed to her:

⁶ While engaging with the passage mentioned above, Cohon argues that the string of gerunds one finds there points to the view advocated by her. This seems a plausible conclusion.

[For Hume] reasoning is the *discovering* of the truth or falsehood of our ideas. ‘Reason’ in a strict sense means not each and every belief about truth and falsehood, but *the judging* of truth and falsehood. ‘Judgement’ can mean *a* judgement, or the process of judgement, and here it means the latter. (Cohon, 2008:70)

While it is clear that for Cohon, judgements are ‘act of judgments’: ‘*judgings*’, there are a couple of remarks to be made. First, in order to render her interpretation of Hume’s notion of ‘reason’ in the guise of ‘demonstrative reasoning’ (assuming an identity relationship in Cohon’s account between ‘demonstrative reasoning’ and ‘demonstration’) fully consistent, she would have to make sure that *if* there is a notion of ‘judgement’ implied by the phrase ‘*preserving assent*’ in the quotation from her we have seen before and that I present again below, this notion is conceptually compatible with her notion of ‘reason’ as *reasoning* and her thought that when Hume uses the term ‘judgement’ he talks about *judgings-acts of judgements*. Here is the quotation again:

In demonstration we proceed step by step in linking related ideas in sequence, *preserving our assent*—or the degree to which the ideas are evident—at each step, so that in the end we find evident (indeed, we are necessarily determined to grasp) a relation between our starting and resulting ideas. (*passim*, e.g. T: 95; Cohon, 2008:66)

It seems plausible to assume that from a Cohon-independent perspective the concept ‘*preserving assent*’ somehow entails the notion of a ‘judgement’. It might also do for Cohon in the above passage (and similar ones). If it does for Cohon, that would mean the following, taking for granted Cohon’s view that for Hume ‘reason’ equates reasoning and the fact that for Cohon ‘demonstrative reasoning’ is a sub-form of ‘reason as reasoning’ more generally. It would mean that this concept of judgement for Cohon (a) needs to be one that in the psychology of a human being operates within a wider activity of ‘reasoning as the discovery

of relations between ideas’, while (b) this concept of ‘judgement’ is a mental *act* (not product), just as any other ‘judgement’ is for Cohon. Otherwise Cohon’s claim about the concept of ‘reason’ for Hume amounting to *reasoning*; her claims about what ‘demonstration’/‘demonstrative reasoning’ is, her claim that for Hume ‘judgments’ are necessarily mental *acts* and her idea that ‘preserving assent’ is something that happens in the process of demonstrative reasoning/demonstration would be inconsistent. I think it is possible for Cohon to meet (a). What Cohon apparently needs is an appeal to an *intuitive* form of *assent* -for example one from Locke. He indeed is a very plausible candidate for stepping in here, given that Hume shared some aspects of his notion of ‘reason’ with Locke. Indeed, Locke seems to *have influenced* some of Hume’s ideas about ‘reason’. Let us have a brief look at what David Owen says about this:

Hume inherited from Descartes and Locke a conception of reasoning that had nothing to do with formal validity [footnote 12 omitted]. [Descartes and Locke] had sufficiently argued against the formal, syllogistic account of reason, the precursor to our notion of deduction, and had put in its place *an account of reasoning, argument or inference in terms of a chain of ideas. An argument or inference is what gets you from one idea to another idea via a chain of intermediate ideas.* Hume took this notion on board.... (Owen, 1994: 200; *cf* Owen, 1992; my italics)

I focus on Hume’s notion of ‘deduction’ in due course. For now and in relation to the topic of *judgement* and Cohon’s view about Hume’s notion of ‘judgement’ let us recall Cohon’s claim that:

In demonstration we proceed step by step in linking related ideas in sequence, *preserving our assent*—or the degree to which the ideas are evident—at each step, so that in the end we find evident (indeed, we are necessarily determined to grasp) a relation between our starting and resulting ideas. (*passim*, e.g. T: 95; Cohon, 2008:66)

As mentioned, it seems to me that in order for Cohon to render her account consistent, *if* Cohon's notion of '*preserving assent*' reflects a judgement, it must be a notion that psychologically takes place *within a wider activity of 'reasoning as the discovery of relations between ideas'*. I have argued that this is possible and suggested that what she needs is an intuitive Lockean form of assent that in theory is compatible with a notion of 'reason' as 'demonstrative reasoning' on its turn *understood as* 'the discovery of relations between ideas'. Or rather, what she needs is an intuitive form of assent, *for example of a Lockean kind*, that *need not equate* the Cohonean sense of demonstrative reasoning as 'the discovery of relations between ideas', being a sub-form of reason as '*reasoning*', but is at least compatible with that. So how could Locke help out? Locke's account of knowledge and certainty that has a prominent place for 'reasoning' relies on 'intuition' in the sense that intuition is necessary in "all the Connexions of the intermediate *Ideas*, without which we cannot attain Knowledge and Certainty" (Locke, 4.2.1; cited in Owen, 1992:188). I cannot find any direct evidence in Hume's work that immediately ascribes this Lockean idea to Hume too (*cf* Meeker, 2007: 234). But, as far as I can see, there is *no reason to think* that this Lockean idea is *not* compatible with Hume's work. If I am right, what could help Cohon out in terms of rendering her account of demonstrative reasoning consistent is the idea of 'reasoning' as a chain of ideas needing intuitive judgemental assent at several places on the route of reasoning where 'assent' is understood as a mental *act*. Note that from the perspective of Hume exegesis (as well as an independent psychological angle) it seems plausible to conclude that this reasoning process typically at last results in a *judgemental outcome*. This judgemental outcome then as the inferential product of a reasoning process/activity Cohon will want to exclude from her conception of reason as reasoning.

In his (1988), A.T Nuyen uses two *terms* to describe something that refers to one and the same *concept* that shows clear similarities with Cohon's conceptions of 'demonstrative reasoning' and 'demonstration'. These two terms Nuyen regards as interchangeable. The terms Nuyen uses are 'argumentation' and 'ratiocination' (NB: these are terms that appear,

albeit rather occasionally, in Hume's work). When explaining the terms, Nuyen argues that they describe the process "in which we consider relations and comparisons of ideas, and *argue* (rather than infer) from causes to effects" (Nuyen, 1988: 380; emphasis original⁷). Nuyen claims that in this type of reasoning "we are guided by 'the power of abstraction of the mind' and we employ our 'intellectual faculties'" (Nuyen, 1988: 380). A second claim Nuyen makes is that 'argumentation' and 'ratiocination' give rise to judgements of pure relations (Nuyen, 1988: 380). A third claim Nuyen argues for is the claim that the modern equivalent of 'argumentation'/'ratiocination' "roughly speaking" is 'deductive reasoning' (Nuyen, 1988: 380). In due course, I comment on Nuyen's or rather *anyone's* conceptual association of '*arguing*' (rather than inferring) with the concepts Nuyen sees as belonging to the terms 'argumentation' and 'ratiocination', but first I would like to comment on Nuyen's remark that the modern equivalent of 'argumentation'/'ratiocination' "roughly speaking" is 'deductive reasoning'. This remark needs major attention. In what follows I rely on David Owen (1992, 1994) for an explanation of why this remark needs attention. However, while doing so and while agreeing with the major point Owen wants to make, I take a critical stance towards some (other) aspects of Owen's explanation.

As Owen (1994) points out, there is a conceptual discrepancy—indeed a major one—between our contemporary use of the term 'deduction' and Hume's use of it:

One seventeenth and eighteenth century sense of 'deduction', commonly used by Locke and occasionally by Hume [footnote 10 omitted], is as a synonym for argument and inference, and I would suggest that we should never use it any other way when speaking of Hume. 'Deductive' in our sense means formally valid. The only conception of formal validity available to Hume was syllogism. So if the deductive conception of reason was Hume's target in the arguments about reason, one might have expected him to mention 'syllogism' or its cognates. He never does in the *Treatise* [footnote 11 omitted], although there are occasional contemptuous references to

⁷ Although what I have italicized in the quotation is underlined in the original article.

‘scholastic headpieces and logicians’ (T 175). Scholastic logic, and its attendant conception of reason, was beneath contempt to Hume, and not the target of his arguments.

Hume inherited from Descartes and Locke a conception of reasoning that had nothing to do with formal validity [footnote 12 omitted]. They had sufficiently argued against the formal, syllogistic account of reason, the precursor to our notion of deduction, and had put in its place *an account of reasoning, argument or inference in terms of a chain of ideas. An argument or inference is what gets you from one idea to another idea via a chain of intermediate ideas.* Hume took this notion on board ... (Owen, 1994: 200; cf Owen, 1992; my italics)

I agree with Owen in that Hume’s notion of deduction is *not* a formal one. Having said that, I wish I had a much better idea as to how Owen conceives of the relationship between Hume’s non-formal notion of ‘deduction’ and Hume’s concept of ‘reason’. First, if the non-formal notion of deduction Owen intends to point to as embraced by Hume is a *mental process of reasoning*, does it include *some sort of conclusion/a judgemental outcome*? A **formal conclusion** is part of our contemporary notion of deduction. Be that as it is, what are Owen’s thoughts on Hume’s ideas about some sort of final product being conceptually part of Hume’s notion of ‘deduction’ given that it is largely, if not exclusively a concept of *reasoning*? Secondly, Owen argues for a relationship between Hume’s notion of ‘deduction’ and ‘argument’. What does ‘argument’ mean here and what is the conceptual relationship with proving and/or proof? What I, thus far, have said about ‘demonstrative reason(ing)’ probably has not made this clear yet, but there is a lot of textual evidence in Hume’s work that speaks in favour of there being some sort of relationship(s) between ‘demonstrative *reasoning*’ (that will have some relationship with Hume’s concept of ‘deduction’ as Owen conceives of it) and ‘proving’.

Thirdly, whatever exactly Owen’s conception of ‘deduction’ as some form of ‘reason’ (on its turn) having an essential relationship with ‘*reasoning*’ is, according to Owen, does this reasoning-related conception of deduction *exhaust* Hume’s account of ‘reason’? Below I

will explain why the third and second question raised are relevant ones. I start with the third question which will organically lead us to the second one.

As mentioned, Owen argues that for Hume ‘deduction’ is just a *generic term* for ‘argument’ (Owen, 1992: 194). Given Owen’s intention and given Hume’s primary work, I understand Owen’s willingness to opt for something *generic*, but there is a problem with Owen’s claim that ‘deduction is just a generic term for argument’. The concept ‘argument’ seems to imply a process and/or outcome having something essential to do with *proving and/or proof*. But, while the concept of proof plays a major role in Hume’s work about mathematical reasoning (it might even be a dominant necessary condition of his concept of ‘demonstrative reasoning/demonstration’) –I would say- that in Hume’s work this aspect of proof is not a *not* a necessary condition of some sort of (maximally) generic form reasoning we find in Hume’s work defined by ‘ideas being mentally connected’. To understand this more deeply, let us look at a thought from Christopher Belshaw.

In his (1989: 156) Belshaw argues that: “it is thought that a proof is involved when something previously unknown is shown to be the case, and is thereafter incontestable” (Belshaw 1989:156).⁸ I take it that there is something highly important in Belshaw’s idea about the connection between ‘demonstrative reasoning’ and/or ‘demonstration’ and ‘proof’ which as an idea is essentially a claim about the relationship between ‘demonstration’ and/or ‘demonstration’ and what one could call ‘knowledge-extension’. I take it that textual support for there being such a relationship comes from a claim Hume makes in the first book of his *Treatise*. There Hume seems to argue for the following:

1. “In all demonstrative sciences [exhausting or at least paradigmatically including the area of mathematics] the rules are certain and infallible”. (T: 180)

⁸ Note that this paper from Belshaw is an article on Hume and one that intends to elucidate Hume’s account of demonstrative knowledge. Note however as well that the mentioned claim seems to be one unconnected to Hume exegesis that however is then used to elucidate Hume’s account.

2. What comes with that for us as human beings is the possibility of *knowledge* (see (T: 180), a concept Hume wants to distinguish from probability.⁹
3. Be that as it is, “knowledge degenerates into probability”, because of disturbing factors, amongst other things “inconstancy of our mental powers”. (T: 180)
4. The more a mathematician and his colleagues run over a particular *proof*, the more (constitutive? Non-constitutive?) evidence there is for the certain and infallible truths there are (seemingly, according to Hume) in mathematics. (see T: 180)¹⁰
5. The more evidence there is for certain and infallible truths, the higher the probability that can be assigned to a purported knowledge claim.
6. It is by an increase in probability that we come closer to the maximal (constitutive? non-constitutive?) evidence for the certain and infallible truth.
7. By increasing the evidence and thereby increasing the probability that can be assigned to a purported knowledge claim, knowledge understood in a general way can be locally extended when (A) the subjective probability *reflects 1*. If not (A), the better reading will be (B): it is sufficient that the subjective probability ‘*closely approaches*’¹ (whatever that exactly might have to mean), where ‘1’ is reflective of absolute *certainty* and *objective truth* (in whatever realist or constructivist way that needs to be understood¹¹).

Let me now –by means of an example– explain my claim that we do not speak accurately when we say that all forms of reasoning present in Hume’s work and characterized by connecting ideas conceptually include something having to do with to proof/proving/to demonstrate/demonstration. What my example should make clear is that there are types of reasoning in Hume’s work in which (a) there is *no* such thing at issue as *knowledge*

⁹ As to the distinction, see e.g. T: 181. Hume argues: “But knowledge and probability are of such contrary and disagreeing natures. . .”. Let me point out however that on the same page, in a sentence directly preceding the one cited, Hume also makes a remark that points to anything but a sharp distinction between knowledge and probability. Hume argues: “[W]e shall find it impracticable to shew the precise limits of knowledge and probability, or discover that particular number, at which the one ends and the other begins”.

¹⁰ As to point 4, on basis of a wider textual reading, I am interpreting Hume more from an epistemological angle, than the descriptive psychological talk on pages 180-1 of the Treatise allows for.

¹¹ This is a very difficult exegetical problem. I say something more about it in § 4.

extension’ and in which (b) an agent does not *intentionally (nor unintentionally) prove something* (if unintentionally proving something is possible). If -I suggest- (a) and (b) are not *individually* sufficient for theoretically rejecting an identification of a generic form of reasoning with ‘to demonstrate’/demonstration (assuming that these have a conceptual relationship with something related to ‘proof’), then they might be jointly sufficient.

Suppose you believe, in the form of an expectation, that this year between three and seven undergraduate students will fail their first assignment of the introductory course in philosophy you are teaching. As far as I can see, Hume would want to say that your reasoning that leads you to such a belief is (1) based on past experience; (2) need not have anything to do with ‘knowledge extension’; and (3) need not imply an *intention* in you *to prove* something. In my view, in the ‘three to seven students will fail’ case Hume would want to say the following about your reasoning, i.e. in terms of a descriptive and causal story.

First, earlier in your career, i.e. when you were quite a bit younger and when (in similar educational circumstances) you held the same belief in regard to the number of fails in the upcoming course, your belief was the result of a conscious effortful mental process of reasoning based on significant, although limited, past (teaching) experience. I assume that Hume wants to tell a different descriptive and causal story about your belief now. That is, while in the past your belief came about as a result of a conscious effortful mental process of reasoning, in your current or future stages of your life as a teacher – having been subject to more experiential evidence – your belief has/will become a non-effortful habitual belief.

As far as I can see, in the abovementioned example your judgement about the number of fails does *not* seem to have any relationship with ‘proving’. Recall that Belshaw argues that, “it is thought that a proof is involved when something (a) previously unknown (b) is shown (c) to be the case, and (d) is thereafter incontestable” (Belshaw 1989:156). Note as well that for our purposes having to do with the relationship between Owen’s reasoning-related notion of deduction and ‘proof’ we only need the former variant of the case present, since the latter case is not well understood in terms of the reasoning Owen talks about.

Aspect (b) –as I have dissected it- of Belshaw’s statement, concerns the concept of ‘something *being shown* to be the case’. In the mentioned example your early (and later) self seemingly did not have an *intention* to show something (nor, is there a deliverance of proof in an unintentional way).¹² You could have had such an intention however. Consider the following. When having a disagreement with your Head of Department about how your students were doing last year, one can think of situations in which you have an *intention* to prove that his belief is wrong. Suppose your Head of Department does *not* believe that you did not have more than seven fails last year, while you are sure that only five students failed their essay last year. You may collect all the data in order to systematically be able to point out to him that, contrary to what your Head of Section believes, you indeed do not have more than seven fails. He is wrong.

This case about the disagreement between you and your Head of Department, departs from Belshaw’s description in the sense that there seems to be an objective empirical evidential fact that there are ‘no more than seven fails’. And thereby there seems to be something Hume would describe as ‘known’. In the example, this fact is just not known by your Head of Department. That it was *already known* then, considering aspect (a) of Belshaw’s description, might well be a sufficient condition to say that the example does not fit with Hume’s concept of ‘demonstration’. This despite the fact that you as the teacher in the example, make an attempt to *show* something (see aspect b as I have dissected it in Belshaw’s description), something that seems to be a necessary condition of Hume’s notion of ‘to demonstrate’. Assuming that there is an objective fact of the matter that there are no more than seven fails, there would also be a compatibility between Belshaw’s definition and the teaching case in that after your efforts, the evidence is incontestable (aspect d). On basis of textual evidence, one can doubt however, that even for mathematical cases, where Hume’s

¹² For Hume himself there seems to be an ‘active attempt’-component to his notion of demonstrative reasoning/demonstration, meaning that for Hume it does *not* seem to be possible to ‘unintentionally prove something’. From a Hume-independent perspective (rather than for Hume himself) it could be the case that it is possible for practical agents (and/or for inanimate objects) to ‘unintentionally prove something’, The option that it is not possible for practical agents (and/or inanimate objects) to prove something in an unintentional way seems also intelligible however.

concept paradigmatically if not exclusively applies, Hume wants to work with the notion of ‘incontestability’. While Hume often linguistically relates mathematical judgements to knowledge as *scientia*, knowledge as certainty, and indeed seems to think that some mathematical judgements can reflect *certain* knowledge, there is textual evidence that speaks against Hume thinking that mathematical judgements are ‘incontestable’. This I discuss in § 12.4. Briefly, this seems to be a result of the apparent fact that Hume –like the ancient Greeks- not just had an inclination to be sceptical about *knowledge*, but also about *justified belief*. As a result, what we seem to get in Hume’s work is this. Even with regard to our most accurate judgements that for Hume represent certain knowledge, Hume forces the clever mathematician to consider the probable fact that he has reasoned incorrectly in the past. Therefore his seemingly correct judgement might be wrong on this very occasion where it seems to reflect certain knowledge. And at certain points in his work, this then leads Hume into surprising and difficult arguments that ask mathematicians to lower the probability attached to their most certain mathematical claims. While Hume’s thoughts on contestability of the most accurate mathematical judgements deserve a real place in his account of mathematics, I also think that they represent a remarkable outlier in Hume’s primary work on mathematics that seems to claim that some mathematical judgements can represent certain knowledge.

3.3. Reason as a ‘faculty’

We have seen that when it comes to ‘judgments’ Cohon argues that judgements are exclusively mental *acts*. I am not so sure about this. Hume every now and then seems to be talking about judgements as *products*. The trouble is that Hume does not have a very elaborate account of judgement.

The discussion about what kind of things ‘judgements’ are –we can restrict it here to the problem of whether they are mental acts and/or mental products- matters quite a bit for the challenge of pinning down Hume’s conception or conceptions of ‘reason’. Namely, given Hume’s work about mathematics, it seems plausible to read Hume as saying that besides ‘reason’ being ‘reasoning’, ‘reason’ is also ‘*a priori judgement*’ where ‘judgement’ here is partially or even exclusively a mental *product* rather than an *act*. I make an attempt to explain and justify this claim in § 4.

Not only does there seem to be a ‘judgement as product view’ we have to take seriously as part of Hume’s conception(s) of ‘reason’, there seems to be an ‘entity-cum-faculty-view’ as well that should be taken seriously as one (aspect of) Hume’s conception(s) of ‘reason’. In order to understand this, let us first look at a passage in Cohon’s work that looks rather confusing to me. In her book Cohon says:

Admittedly, Hume does not talk without exception of reason or the understanding as an activity rather than an entity. Nor is his notion of moral discrimination always sharply in view; in places he quite understandably slides into treating moral distinctions as outcomes rather than distinguishings. (Cohon, 2008: 80)

Cohon thinks that what she says here “complicat[es] (though not ultimately undermin[es]) [Hume’s] logic” (Cohon, 2008:80). By this she just means that when Hume talks about judgements we should interpret them as referring to *judgings*, as we have discussed above.

What seems to me slightly confusing in the above-mentioned passage is the lack of an explicit distinction between the ‘judgements-as-outcome-view’ and the ‘entity-cum-faculty-view’ of ‘reason’. According to the former, the concept of ‘reason’ either fully consists in or else at least *entails* ‘judgements as outcomes’. According to the latter there would be a ‘faculty of reason’ that manifests itself in one way or another. That could be in terms of the *judgings* Cohon advocates or else in the formation of true-or false judgements. Or both. Or perhaps something else. To me it looks like what is implied by Cohon’s view of ‘reason’ as

reasoning (being a process or activity) is not only a 'non-outcome-view', but also a non-entity-cum-faculty view. So, she seemingly not only wants to deny that 'reason' consists in 'judgemental *outcomes*', but also that 'reason' consists in some kind of entity-cum-faculty that manifests itself in something, be it the reasoning activity or processes she advocates or something else, most probably *judgements as outcomes*. What points to the fact that she also denies the faculty-cum-entity-view is her claim that: "If reason were an organ of the mind that could be identified apart from what happens when reasoning activity is going on, then it would be open to speculation, and to observation, what other things it might do or what its products might be. But as a thoroughgoing empiricist Hume cannot use such a notion of reason" (Cohon, 2008:72).

Cohon's view seems correct to me in that I think that Hume's empiricism gives him a self-ascribed justification (if not also an objective justification) not to work with a notion of 'reason' as some sort of faculty or, to use Cohon's terms 'organ of the mind' that *cannot* be identified by means of an appeal to experience, in this case by internal observation, i.e. by introspection. Having said that, there clearly is a fair bit of textual evidence for the faculty-view. In the introduction of his *Treatise* Hume argues:

If therefore the sciences of Mathematics, Natural Philosophy, and Natural Religion, have such a dependence on the knowledge of man, what may be expected in the other sciences, whose connexion with human nature is more close and intimate? The sole end of logic is to explain the principles and operations of our reasoning faculty, and the nature of our ideas: morals and criticism regard our tastes and sentiments: and politics consider men as united in society, and dependent on each other. In these four sciences of *Logic, Morals, Criticism, and Politics*, is comprehended almost every thing, which it can any way import us to be acquainted with, or which can tend either to the improvement or ornament of the human mind. (T: xv-xvi; italics original, my embolding)

And in book II of his *Treatise* for example Hume says: “When any of these passions are calm, and cause no disorder in the soul, they are very readily taken for the determinations of reason, and are suppos’d to proceed from the same *faculty*, with that, which judges of truth and falsehood” (T: 417; my italics; see also T: 193¹³; T: 414; T: 415; T467; T: 610; T: 646).¹⁴

Looking at the type and also quantity of Humean claims about reason as a ‘faculty’ (about quantity: even though not very often, Hume refers to ‘reason’ as a ‘faculty’ on a significant number of occasions, including the introduction of the *Treatise*) there seems to be sufficient ground to take the faculty view of ‘reason’ seriously. Having said that, let us look at the following remark from Hume:

[A]s nature seems to have observ’d a kind of justice and compensation in every thing, she has not neglected philosophers more than the rest of the creation; but has reserv’d them a consolation amid all their disappointments and afflictions. This consolation principally consists in their invention of the words *faculty* and *occult quality*. For it being usual, after the frequent use of terms, which are really significant and intelligible, to omit the idea, which we wou’d express by them, and to preserve only the custom, by which we recal the idea at pleasure; so it naturally happens, that after the frequent use of terms, which are wholly insignificant and unintelligible, we fancy them to be on the same footing with the precedent, and to have a secret meaning, which we might discover by reflection. The resemblance of their appearance deceives the mind, as is usual, and makes us imagine a thorough resemblance and conformity. By this means these philosophers set themselves at ease, and arrive at last, by an illusion, at the same indifference, which the people attain by their stupidity, and true philosophers by their moderate scepticism.

¹³ Note that at least terminologically speaking, here Hume talks not about ‘reason’, but about ‘the understanding’. There are exegetical issues as to the conceptual relationship between (a) ‘reason’ and ‘the understanding’; (b) reason and ‘intellectual faculties’ (when using this compound term, Hume uses the plural form ‘faculties’, while often referring to reason as a ‘faculty’) and (c) ‘the understanding’ and ‘intellectual faculties’ in Hume’s work. (see e.g. T: 267-8, 371; EPM: 32, 172, 291, 293). The concepts ‘reason’, ‘the understanding’ and ‘the intellectual faculties’ were definitely highly conceptually related for Hume. Having said that, Hume, more or less intentionally, may have used the first term less than the other two for an all-encompassing organ of the mind.

¹⁴ Note that Hume describes memory and the imagination as two (other) ‘faculties’ (besides reason) (see e.g. T: 23-4, 183 and especially also T: 117-118, footnote 22). Occasionally Hume also refers to the ‘will’ as a ‘faculty’ (T: 417-8).

They need only say, that any phaenomenon, which puzzles them, arises from a faculty or an occult quality, and there is an end of all dispute and enquiry upon the matter. (T: 224)

This Humean statement seems to express some real antipathy towards the term 'faculty'. Even if that is true however, Hume's antipathy here for (the use of) the term 'faculty' seems compatible with him using a faculty conception to refer to 'reason'. Hume might see the term 'faculty' fit for that/part of which he wants to capture by the term 'reason' (probably without wanting to put too much metaphysical burden on this notion), while also having a concern about the use of the term faculty that reflects a concern about philosophers appealing to some hocus-pocus metaphysical faculty.

4. Reason and mathematics

According to Kant, Hume thought that all mathematical judgements are *analytic*. Hume's account of mathematics gives no reason to think that this is true. Kant seems to have been mistaken there (see Atkinson, 1960, Gotterbarn 1974, Steiner, 1987).¹⁵ Hume *did* think however that mathematical judgements were of an *a priori* type.¹⁶ In *EHU* Hume argues:

All the objects of human reason or enquiry may naturally be divided into two kinds, to wit, *Relations of Ideas*, and *Matters of Fact*. Of the first kind are the sciences of Geometry, Algebra, and Arithmetic: and in short, every affirmation which is either intuitively or demonstratively certain. *That the square of the hypotenuse is equal to the square of the two sides*, is a proposition which expresses a relation between these figures. *That three times five is equal to the*

¹⁵ That might be due to Kant not having read the *Treatise*. Donald Gotterbarn (1974) claims that Kant did not read the *Treatise*. I myself am insufficient confident to commit myself to any view on this matter.

¹⁶ Note that Hume himself indeed used the term *a priori* (see e.g. T: 247; EHU: 27), although much less frequently than Kant.

half of thirty, expresses a relation between these numbers. Propositions of this kind are discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe. Though there never were a circle or triangle in nature, the truths demonstrated by Euclid would for ever retain their certainty and evidence. (EHU: 25)

So for Hume mathematical '*propositions*', and most probably Hume also wants to say '*judgements*', are *a priori*; they "are discoverable by the mere operation of thought". But who makes this claim? Is it Hume the cognitive psychologist and/or Hume the epistemologist? As I read Hume, he clearly is a cognitive psychologist throughout his work on mathematics. I also get the impression however that he has normative views about the presence and nature of knowledge in mathematics. He could be a constructivist or a realist. As to the former, Hume argues that "in an arithmetical operation...both the truth and the assurance are of the same nature" (T: 449). If and when Hume is an epistemological *realist* on mathematics he is one who does *not* believe in a mind-*independent* realism for 'objects of geometry'. Hume argues:

[O]bjects of geometry, those surfaces, lines and points, whose proportions and positions it examines, are mere ideas in the mind; and *not only never did, but never can exist in nature*. They never did exist; for no one will pretend to draw a line or make a surface entirely conformable to the definition: They never can exist; for we may produce demonstrations from these very ideas to prove, that they are impossible. (T: 42-3; cfT: 52)

This might be compatible with Hume being a realist about geometrical mathematical *conclusions*. We can leave this issue aside here. It certainly does seem to be compatible with Hume being a mind-independent realist on mathematical conclusions in the other two areas of mathematics Hume talks about besides geometry: arithmetic and algebra. That is, logically speaking such a compatibility seems possible. Furthermore, from a perspective of textual exegesis we cannot easily exclude Hume being a mind-independent about mathematical

conclusions in arithmetic and algebra. In order to understand this, I first need to lay out a lot more about Hume's writing on mathematics.

As indicated earlier, for Hume mathematical judgements have something important to do with 'demonstration'. But what exactly is not a particularly easy matter. This is so because there is an exegetical challenge that asks us to make sense of the following four elements that Hume seems to associate with 'demonstrative reasoning' and 'demonstration' (note that these two terms *may* not reflect fully identical concepts): (a) something having to do with *proof/proving* (T: 449 and recall the argument about 'knowledge-extension'); (b); something that has to do with an operation of the understanding that involves *intermediate ideas* rather than it being concerned with 'intuition': an immediate grasp of something (T: 70); (c) something having to do with *certainty* as opposed to probability (see e.g. T: 449); (d) something having to do with demonstrative reasoning/demonstration being incompatible with 'contradiction' (T: 43, cf EHU: 163-164). It is beyond the purposes of this chapter to solve the problem of what the terms 'demonstration' and 'demonstrative reasoning' refer to, but for our purposes of getting a good grasp on Hume's conception of reason to subsequently use it in an argument to identify one criterion for there being a non-realist component to moral objectivity, it *is* important to dig a little deeper into the meanings of some of the aspects I have just mentioned.

For a brief discussion about aspect (a) I refer to the previous section in combination with Hume's claim that "in an arithmetical operation...both the truth and the assurance are of the same nature" (T: 449). As to aspect/candidate (b) of the concepts belonging to the terms 'demonstrative reasoning' and demonstration, Hume says the following in the *Treatise*:

It appears, therefore, that of these seven philosophical relations, there remain only four, which depending solely upon ideas, can be the objects of knowledge and certainty. These four are *resemblance, contrariety, degrees in quality, and proportions in quantity or number*. Three of these relations are discoverable at first sight, and fall more properly under the province of **intuition** than **demonstration**. When any objects *resemble* each other, the resemblance will at

first strike the eye, or rather the mind; and seldom requires a second examination. The case is the same with *contrariety*, and with the *degrees* of any *quality*. No one can once doubt but existence and non-existence destroy each other, and are perfectly incompatible and contrary. And tho' it be impossible to judge exactly of the degrees of any quality, such as colour, taste, heat, cold, when the difference betwixt them is very small; yet 'tis easy to decide, that any of them is superior or inferior to another, when their difference is considerable. And this decision we always pronounce at first sight, without any enquiry or reasoning. (T: 70; italics original; embolding mine)

For Hume, there is some sort of important contrast between 'demonstrative reasoning' and 'demonstration' on the one hand and 'intuition' on the other hand. As such, the former has something to do with as Hume calls it himself "interposition of other ideas" (T: 95), or call it 'taking intermediate steps in thought', while the latter has something to do with immediacy/an immediate grasp.

As to aspect (c), concerning the contrast between certainty and probability, as we have seen in the quotation at the top of this section, Hume argues:

All the objects of human reason or enquiry may naturally be divided into two kinds, to wit, *Relations of Ideas*, and *Matters of Fact*. Of the first kind are the sciences of Geometry, Algebra, and Arithmetic: and in short, every affirmation which is either intuitively or demonstratively certain.

Here in fact we see a combination of aspect/candidate (a) and (b). Note furthermore that from Hume's work becomes clear that judgements about matters of fact for Hume are judgements about probability.

Now, things get exegetically difficult because Hume seems to associate demonstrative knowledge with *certainty* as opposed to *probability*. Taking that for granted here, Hume argues that *not all three* areas of mathematics (Hume sees as jointly exhaustive of the subject area of mathematics): arithmetic, algebra and geometry can be associated with *knowledge as*

certainty or even knowledge at all¹⁷. In book 1 of the *Treatise*, Hume excludes geometry from knowledge as reflecting certainty and relates it to probability (T: 71). However, it is not so clear that geometry is exempt from ‘reasoning’/proof that results from ‘interposition of ideas in thought’. Obviously, through history ‘geometry’ did not always refer to the same concept. It is not totally clear to me what it meant to Hume. In any case, one thing that speaks in favour of geometry primarily if not exclusively being excluded from ‘interposition of ideas in thought’ in favour of being concerned with ‘intuition’ is the rather visual nature of geometry. Hume explicitly points to this e.g. on page 71-2 of his *Treatise*. Taking that seriously, note that in terms of a solution to the problem it obviously matters whether geometry is only primarily or exclusively concerned with intuition.

Having mentioned that for Hume there appears to be some sort of real contrast between ‘demonstration’ and ‘intuition’, the exegetical challenge becomes even more interesting and challenging once one keeps open the possibility that (i) demonstration and/or demonstrative reasoning and/or demonstrative knowledge and (ii) ‘intuition’ are compatible. As indicated in § 3.2, they seem to have been for Locke and since Hume seems to have worked at least partly in a Lockean framework when it comes to Hume’s ideas about ‘reason’, they might have been compatible for Hume as well. As mentioned earlier, for Locke knowledge that requires reasoning always relies on intuition in the sense that this intuition is necessary in “all the Connexions of the intermediate *Ideas*, without which we cannot attain Knowledge and Certainty” (Locke, 4.2.1; cited in Owen, 1992:188).

As to aspect (d), i.e. Hume’s ideas about ‘demonstrative reasoning’ and ‘demonstration’ as related to a Humean claim about *contradiction*, in book I of the *Treatise* Hume argues:

Whatever can be conceiv’d by a clear and distinct idea necessarily implies the possibility of existence; and he who pretends to prove the impossibility of its existence by any argument

¹⁷ As to ‘knowledge at all’, there are difficulties here as to how on behalf of Hume best to label judgements with a high probability.

deriv'd from the clear idea, in reality asserts, that we have no clear idea of it, because we have a clear idea. 'Tis in vain to search for a contradiction in any thing that is distinctly conceiv'd by the mind. Did it imply any contradiction, 'tis impossible it cou'd ever be conceiv'd. (T: 43)

This is a passage that reflects the fact that when Hume *explicitly talks about* (note my remark later on in this paragraph) the concept of 'contradiction', his words reflect an endorsement of a *psychological* rather than *logical* notion of contradiction. The difficult exegetical puzzle is when in Hume's work this *psychological* notion of contradiction is a *psychological notion related to* claims about (a) 'truth and falsehood', (b) 'necessity' and (c) 'certainty' for Hume as a *cognitive psychologist* and when this psychological notion of contradiction is a psychological notion for claims about (a) 'truth and falsehood', (b) 'necessity' and (c) certainty for the *epistemologist* Hume. The fact that for Hume the notion of contradiction in passages like the one presented is *psychological rather than logical* neither logically nor from a perspective of textual exegesis implies that it therefore can solely be a *criterion* of the *cognitive psychologist* Hume, i.e. it might be one of the *epistemologist* Hume as well. There is another remark to be made. The fact that for Hume the criterion of (non-)contradiction is psychological in passages like the one given, does *not* imply that Hume does *not* endorse the thesis that there are certain and necessary, *possibly mind-independent*, truths that have as a criterion the logical principle of non-contradiction. It is beyond the purposes this chapter to provide an extensive and deep explanation and justification of this claim. However, observing both a Humean tendency to sharply distinguish and to intentionally (and unintentionally) blur the concepts of 'knowledge'¹⁸ and 'probability'^{19,20}, I want to make the following two tentative remarks about my own claim. First, it does not seem implausible to me to read Hume as saying that there *are objective truths in mathematics that have an*

¹⁸ Hume argues: "By knowledge, I mean the assurance arising from the comparison of ideas" (T: 124).

¹⁹ Hume argues: "By probability, that evidence, which is still attended with uncertainty" (T: 124).

²⁰ Hume argues that all knowledge degenerates into probability (see T: 181).

objective probability of 1 (cf Meeker, 2007). Secondly, such truth may be (partly or even exclusively) mind-independent ones. From a perspective of textual exegesis it seems possible, to ascribe both conclusions to Hume. Both conclusions however, and especially the latter, deserve only a modest status, i.e. they should be regarded as no more (and less) than tentative conclusions/hypotheses to be investigated further.

5. 'Matters of fact' and evaluative distinctions

5.1. What are 'matters of fact' ?

At the top of the previous section I have presented the following statement from Hume:

All the objects of human reason or enquiry may naturally be divided into two kinds, to wit, *Relations of Ideas*, and *Matters of Fact*. Of the first kind are the sciences of Geometry, Algebra, and Arithmetic: and in short, every affirmation which is either intuitively or demonstratively certain. *That the square of the hypotenuse is equal to the square of the two sides*, is a proposition which expresses a relation between these figures. *That three times five is equal to the half of thirty*, expresses a relation between these numbers. Propositions of this kind are discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe. Though there never were a circle or triangle in nature, the truths demonstrated by Euclid would for ever retain their certainty and evidence. (EHU: 25; cfEPM: 287)

I have closed the previous section by some thoughts about contradiction. Hume appeals to this notion as well in the following passage that succeeds his remark above. More or less directly after he has expressed the above, Hume argues:

Matters of fact, which are the second objects of human reason, are not ascertained in the same manner; nor is our evidence of their truth, however great, of a like nature with the foregoing. The contrary of every matter of fact is still possible; because it can never imply a contradiction, and is conceived by the mind with the same facility and distinctness, as if ever so conformable to reality. *That the sun will not rise to-morrow* is no less intelligible a proposition, and implies no more contradiction, than the affirmation, *that it will rise*. We should in vain, therefore, attempt to demonstrate its falsehood. Were it demonstratively false, it would imply a contradiction, and could never be distinctly conceived by the mind. (EHU: 25-6)

So, just as Hume is concerned with contradiction in the area of relations of ideas expressed by mathematical propositions, he is concerned with contradiction in the area of ‘matters of fact’. And in both cases the notion of contradiction seems to be a psychological one rather than a logical one. In what follows I concentrate on Hume’s notion of a ‘matters of fact’. In the next section I connect Hume’s ideas about ‘matters of fact’ with his notion of contradiction. An explanation of this latter relationship should help us understand the relationship between ‘reason’ and impressions’ in Hume’s work. That on its turn should serve a good understanding of Hume’s claim that reason alone cannot make moral distinctions; passions necessarily contribute.

Hume himself says the following about what a ‘matter of fact’ is. Actually that is not quite true, Hume says the following about what an ‘*inference* concerning a matter of fact’ is. Hume argues:

An inference concerning a matter of fact is nothing but the idea of an object, that is frequently conjoin’d, or is associated with a present impression. (T: 626)

Immediately after having made this claim, Hume argues: “This is the whole of it”. I wish things were so easy as Hume gives us the impression. When it comes to Hume’s philosophy, we both need to be careful and pay a lot of attention to what the term ‘matters of fact means’.

We need to be careful in the sense that the meanings Hume (according to me) ascribes to the term do not fit very well with mainstream contemporary views of 'facts'. First, as we can see in the above quotation, for Hume there is a relationship between 'inferences concerning matters of fact' and '*objects*'. I take that in contemporary philosophy, rather than being 'objects', 'facts' are often conceived of as either the *referents of true propositions, beliefs and/or statements* or else the *truth makers of true propositions, beliefs or statements*. To illustrate the former by means of an example, 'The German football team won the world cup in 1954', if true [as a proposition, belief or statement], refers to the fact that that German football team won the world cup in 1954. To illustrate the latter by means of the example, the fact that Germany won the world cup in 1954 is what makes the proposition, belief or sentence 'The German football team won the world cup in 1954' true.

The second thing to watch out for is the following one. As we know, Hume thinks we cannot draw any philosophical conclusions that go beyond our experiences. This may well clash with contemporary mainstream thoughts that facts have something to do with an extra-mental world.

Below, by means of a brief bullet point list consisting of a variety of likely uses of Hume's term 'matters of fact', I point out how according to me Hume uses the term 'matters of fact' in his work. Broadly speaking what explains the significant variety in the list is that in his work about 'matters of fact' Hume seems to adopt both the guise of a social scientist and the guise of an epistemologist. As a *social scientist*, Hume is largely, but not exclusively a *cognitive psychologist*. As a *social scientist* who is *not* best said to come in the guise of a *cognitive psychologist*, Hume observes that ordinary people make factual statements about the world and draw epistemological conclusions about the world. In his typical *sub-guise of a social scientist* - the *cognitive psychologist* - then, Hume describes the psychological mechanisms behind ordinary people's linguistic claims and psychological conclusions.

While Hume is very much a social scientist when he talks about matters of fact, he also seems to be an *epistemologist*. Here it should be pointed out that as an epistemologist

Hume seems to come in *two guises*. Hume certainly is a philosopher having clear and *hard* views about the *limits* of knowledge. However, Hume also seems to come in the guise of a more (i) *positive* (ii) down to earth *practically orientated* philosopher concerned with epistemological truths (or rather epistemological normativity not suffering from all kinds of undesirable epistemological defects) that can be used in daily life.

Here then is my view of how Hume uses the term ‘matters of fact’ in a bullet point summary:

- 1) The term ‘matters of fact’ for Hume does *never* refer to any hard philosophical theoretical commitment about the existence of an extra-mental world and any objects therein (his empiricism prevents him from that).
- 2) For Hume, the term ‘matters of fact’ is one that is used by ‘the vulgar’ in thought and language to refer to *‘something objective’*. Sometimes, but not necessarily the term is used that literally and explicitly.
- 3) Occasionally, for Hume the term ‘matters of fact’ sometimes seems to refer to *extra-mental objects*, namely in the context of Hume referring to a mistake of ‘the vulgar’ when they have beliefs about objects and make claims about the nature of their perceptions.
- 4) The term ‘matters of fact’ seems to be a particular kind of *truth-maker* for the (a) *practically oriented down-to-earth*, (b) *social scientist* Hume concerned with what ordinary people see as true beliefs. The truth-maker however comes from analyses of Hume as a moderate (rather than radically skeptical) *epistemologist*. The truth-maker seems to be a perceptually unflawed impression of sensation in the case of our beliefs in extra-mental objects. In the case of causal relationships it is our frequent experience of two events, influenced by proper employment of our reasoning capacities and a healthy functioning of our imaginative faculty.

- 5) (Related to 4, but leaving the scientific guise out, i.e. keeping (4a) while leaving aside (4b)), the term 'matters of fact' seems to be a particular kind of truth-maker of what for Hume as a moderate rather than radically skeptical epistemologist are true propositions, beliefs and statements. The truth-maker seems to be a perceptually unflawed impression of sensation in the case of our beliefs in extra-mental objects. In the case of causal relationships it is our sufficiently frequent experience of two events, influenced by proper employment of our reasoning capacities and a healthy functioning of our imaginative faculty.
- 6) The term 'matters of fact' *might* also be a referent of epistemologically true propositions and statements for the epistemologist Hume.

I am inclined to say that what Hume observes as a social scientist in terms of ordinary people making claims about there being facts and normative truths can be epistemologically evaluated for Hume by means of 'sufficiently careful philosophical conclusions'. These conclusions are couched in skeptical terms concerning the limitations of what we can know. They come in a *hard-core* form as a result of the epistemological truth/force of Pyrrhonian scepticism (see § 10).

Although I take this to be a little more controversial, Hume's 'sufficiently careful philosophical conclusions' furthermore seem to come in the guise of 'empirically informed inductive conclusions about whose accuracy we as philosophers are confident and that do not suffer from obvious epistemological flaws'. In contrast to the first type of 'sufficiently careful philosophical conclusions', these conclusions, if they can be ascribed to Hume the epistemologist would come in a mitigated rather than hard-core form. A mitigated form that (a) is characterized by good reasoning and a healthy functioning of our imaginative faculty and (b) is compatible with Hume's conviction that all human beings, including philosophers, need to live their lives effectively as practical agents. That latter feature (b) is incompatible

with hard-core Pyrrhonian scepticism, but leaves room for, indeed demands a mitigated form of, such scepticism.

6. 'Reason', 'truth' and contradiction

In this section I focus again prominently on Hume's notion of 'reason', doing so in relation to his ideas about 'truth' and 'contradiction'. First, I present some thoughts about the relationship between 'reason' and 'truth and falsehood'. Subsequently I involve Hume's notion of contradiction.

Let us first consider the following passage (I dub A):

(A) *It seems evident, that reason, in a strict sense, as meaning the judgement of truth and falsehood, can never, of itself, be any motive to the will, and can have no influence but so far as it touches some passion or affection. Abstract relations of ideas are the objects of curiosity, not of volition. And matters of fact, where they are neither good nor evil, where they neither excite desire nor aversion, are totally indifferent; and whether known or unknown, whether mistaken or rightly apprehended, cannot be regarded as any motive to action. (DP: 5.1; cited in Cohon, 2008: 50; cf. 69; italics original)*

Let us also have a look at some other utterances of Hume on 'truth and falsehood'. In the *Treatise* Hume states:

- i. Reason is the *discovery* of truth or falsehood. Truth or falsehood consists in an agreement or disagreement either to the real relations of ideas, or to real existence and matter of fact. Whatever, therefore, is not susceptible to this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. (T: 458; italics mine)

- ii. Truth is of two kinds, consisting either in the discovery of the proportions of ideas, consider'd as such, or in the conformity of our ideas of objects to their real existence. (T: 448)
- iii. What may at first occur on this head is that *nothing can be contrary to truth or reason*, except what has a reference to it, and as the judgements of our understanding only has this reference, it must follow, that *passions can be contrary to reason only so far* as they are *accompany'd* with some judgement or opinion. (T: 415-416; the former two phrases have been italicized by me; the latter word is italicized by Hume)

Cohon summarizes these ideas in the first three steps of what she calls the *Representation Argument* (the other three steps we can leave aside here):

1. Reason is the discovery of truth and falsehood. (see T: 458)
2. Truth or falsehood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact. (see T: 458)
3. Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. (see T: 458; Cohon, 2008: 69)

So, what we see here is that according to Hume 'truth' and 'falsehood' consist in the matching of our ideas with something else: real 'relations of ideas', or 'real existence and matter of fact' (probably Hume intends to use the final two as synonyms here). For the purposes of understanding Hume's concept of 'reason' as a comprehensive notion doing justice to a variety of aspects and descriptions in Hume's work, it seems important to recognize the following difference between step 1 of the *Representation Argument* and the first sentence of passage (A) at the top of this section. In step 1 of the argument Hume talks about reason as the *discovery* of truth and falsehood while in the first sentence of passage (A) Hume talks about reason as the *judgement* of truth and falsehood. This difference *could very*

well reflect the double guise Hume takes in his work. The *former* would refer to the guise of an epistemologist claiming that there is such a thing as normative truth, while also making claims as to how we can know it, but while assuming all the way that there are limitations to what we can know. The *latter dependent on the subject area* may represent both guises. To restrict myself just to two areas here, in the domain of *matters of fact* it seems to be solely a cognitive psychologist who is speaking, while in the domain of mathematics, both guises might very well play a role.

It seems to be the case that from a comprehensive point of view of the different areas of his work on truth and reason (at least this counts for the union of mathematics and matters of fact; perhaps also more widely), the two mentioned intellectual guises meet each other in certain Humean ideas about contradiction. To understand this, let us consider the following two passages from Hume's work:

- I. When we infer the existence of an object from that of others, some object must always be present either to the memory or senses, in order to be the foundation of our reasoning; since the mind cannot run up with its inferences *in infinitum*. Reason can never satisfy us that the existence of any object does ever imply that of another; so that when we pass from the impression of one to the idea or belief of another, we are not determin'd by reason, but by custom or a principle of association. (Hume, T:97)

- II. In a word, if we proceed not upon some fact, present to the memory or senses, our reasonings would be merely hypothetical; and however the particular links might be connected with each other, the whole chain of inferences would have nothing to support it, nor could we ever, by its means, arrive at the knowledge of any real existence. If I ask, why you believe any particular matter of fact, which you relate, you must tell me some reason; and this reason will be some other fact, connected with it. But as you cannot proceed after this manner, *in infinitum*, you must at last terminate in some fact, which is present to your memory or

senses; or must allow that your belief is entirely without foundation. (EHU: 46; *cf* T: 82-3, 91, 97, 517-8; DNR: 11.4 , 11.17)

In both passages Hume appeals to an inability of 'reason(ing)' to result in terminating processes. But how exactly? It looks like first and foremost Hume is giving us a bit of psychological theory. That is, he argues that whatever is going on in our minds in terms of some sort of inferential reasoning process must trace back to some kind of impression. Implicitly, Hume is also making an *observation* as a cognitive psychologist: us having an impression is what happens in our minds.

As an instrument for understanding how the two Humean intellectual guises meet each other in Hume's work on reason, truth and contradiction, let me now present a passage from Hume's *moral* philosophy that also reflects some Human view about an *inability of reason to terminate a process of reasoning*. In *EPM* Hume argues:

It appears evident that the ultimate ends of human actions can never, in any case, be accounted for by *reason*, but recommend themselves entirely to the sentiments and affections of mankind, without any dependence on the intellectual faculties. Ask a man *why he uses exercise*; he will answer, *because he desires to keep his health*. If you then enquire, *why he desires health*, he will readily reply, *because sickness is painful*. If you push your enquiries farther, and desire a reason *why he hates pain*, it is impossible he can ever give any. This is an ultimate end, and is never referred to any other object.

Perhaps to your second question, *why he desires health*, he may also reply, that *it is necessary for the exercise of his calling*. If you ask, *why he is anxious on that head*, he will answer, *because he desires to get money*. If you demand *Why? It is the instrument of pleasure*, says he. And beyond this it is an absurdity to ask for a reason. It is impossible there can be a progress *in infinitum*; and that one thing can always be a reason why another is desired. Something must be desirable on its own account, and because of its immediate accord or agreement with human sentiment and affection. (EPM: 293)

The passage above seems to contain a vital argumentative point to support Hume's claim that moral distinctions are not derived from reason. That is, in the form of a Humean claim about some impotence of reason in terms of it having an inability to result in something that terminates a reasoning process. Having said that, this seems to be one of the passages in Hume's work about which one wonders strongly whether Hume's primary aim is to give an account of the nature of moral distinctions or whether his primary aim is to present an empirical investigation of the mind of men by giving a descriptive analysis of human mental states. The first sentence of the passage seems compatible with both a descriptive and normative project. The focus on the mental states, or rather linguistic utterances of human agents that follows suggests the former. The penultimate sentence where he uses the term 'desirable' –an apparently normative term- suggests the latter. In any case, Hume concludes that it is pain and pleasure that stop an endless regress. He also concludes that our intellectual faculty of reason does not have the necessary features to stop an endless regress that functions as the ultimate foundation for our moral distinctions. It looks like there is no such thing for Hume as a long string of intuitions of assent being able to result in a normative conclusion.

The similarity between the passages about *matters of fact* and the *moral* passage seems to be one that is important to understand Hume's notion of 'reason' and to explain the variety of aspects there seem to be to the notion. On basis of what I have said in earlier sections about Hume's notions of reason, I now present an overall view of Hume's notion of 'reason'. That overall view will incorporate a justification of my claim that the cognitive psychologist Hume and the epistemologist Hume meet each other in Hume's thoughts about reason, truth and (non-) contradiction.

All aspects considered, it seems best to arrive at the following tentative conclusions about 'reason' in Hume's work. For Hume 'reason' is essentially a *reasoning process/activity*. Possibly (sometimes) for Hume this reasoning process is part of a *faculty view* of reason. Here then the notion of contradiction becomes important for both the area of mathematics

and the areas of matters of fact. I think Hume wants to argue that it is only when our faculty of reason *cannot* conceive of contrary instances that it can give us *a priori* knowledge as it can in mathematics. When reason *does* give us a *a priori knowledge* the knowledge results from *a priori reasoning processes* that are terminated by our faculty of reason by means of an *a priori act of judgement* resulting in an *a priori product of judgement* reflecting a *a priori knowledge*.

In regard to the case of matters of fact and contradiction, I think Hume as a skeptical epistemologist wants to claim that reason as a chain of arguments is essentially incapable of giving us any certain truth about objects that seem to be part of the external world *because it is always possible for the mind to conceive of contrary instances*. As we have seen, Hume argues:

The contrary of every matter of fact is still possible; because it can never imply a contradiction, and is conceived by the mind with the same facility and distinctness, as if ever so conformable to reality. *That the sun will not rise to-morrow* is no less intelligible a proposition, and implies no more contradiction, than the affirmation, *that it will rise*. We should in vain, therefore, attempt to demonstrate its falsehood. Were it demonstratively false, it would imply a contradiction, and could never be distinctly conceived by the mind. (EHU: 25-6)

I take it that the final sentence of this passage at least is compatible with, if not exclusively applies to Hume's ideas about true mathematical judgements. Here then his thoughts about truth and non-contradiction will refer at least partly to a psychological criterion for a judgement of a cognitive psychologist. While they refer 'at least partly' to a psychological criterion, referring to and relying on what I have said in § 3 and 4, Hume's thoughts about (non-)contradiction may also represent an endorsement of a (*mind-independent*) logical form of contradiction advocated by Hume the epistemologist.

In § 8, I discuss Hume's moral philosophy, including his views about 'reason' in relation to morality. Before we can well deal with his moral philosophy, there is another discussion awaiting us however. We need to discuss Hume's views on 'passions'.

7.Humean passions

7.1. The unrepresentative character of impressions

7.1.1. On the concept 'representation' in Hume's work

One of the more important points established by Hume in the very first section of the *Treatise* is what Cohon and Owen (UM) call the *Priority Principle*: "all our simple ideas in their first appearance are deriv'd from simple impressions, which are correspondent to them, and which they exactly represent." (T: 4, cited in Cohon & Owen, UM). As Cohon and Owen point out:

Simple impressions and ideas come in resembling pairs, and as the former cause the latter, the latter represent the former. Ideas represent, and what they represent is impressions. Impressions don't, it appears, represent at all. At least, they don't represent other perceptions of the mind.

Note that to say that impressions are not copies of other, precedent perceptions is not to deny that they have no causes. Nor is it yet to deny that they might resemble, copy or represent their causes (Cohon & Owen, UM: 4). Hume just has no commitment to that because of the limiting side of his empiricism. Hume himself says the following about that:

As to those *impressions*, which arise from the senses, their ultimate cause is, in my opinion, perfectly inexplicable by human reason, and 'twill always be impossible to decide with certainty, whether they arise immediately from the objects, or are produc'd by the creative power of the mind, or are deriv'd from the author of our being. (T: 84)

The *Priority Principle* reflects the main idea of ‘representation’ in Hume’s work. Granting that, and while there is a notion of ‘correspondence’ inherent to the *Priority Principle*, the conclusion should be drawn that Hume’s main notion of representation is different from (contemporary) correspondence theories of truth with a theoretical commitment to the existence of an extra-mental world. First, for Hume the notion of correspondence inherent to the *Priority Principle* is not one that is directly concerned with truth and falsehood. Secondly, none of Hume’s mental items are mental signs or symbols that reflect extra-mental reality.

While Hume’s notion of ‘representation’ does *not* refer to an extra-mental reality and while Hume’s *primary use* of the term seems to refer to a particular relationship between ideas and impressions, there might be another use of the term that corresponds to what I have said in the previous section about truth and falsehood. By this I mean the following. *If* Hume draws indeed –as I tentatively believe– positive epistemological conclusions in one or more subject areas he writes about then we might have to say that whatever for Hume can be normatively true and false is ‘represented’ by certain mental items, ‘perceptions of the mind’ as Hume calls them. There is textual evidence that speaks in favour of such a use of the term. Having said that, on basis of a comprehensive reading of his work, I am not entirely sure that such a use of the term ‘representation’ is in a significant sense part of Hume’s work. As to textual evidence, on page 84 of the *Treatise* Hume argues:

As to those impressions, which arise from the senses, their ultimate cause is, in my opinion, perfectly inexplicable by human reason, and ’twill always be impossible to decide with certainty, whether they arise immediately from the object, or are produc’d by the creative power of the mind, or are deriv’d from the author of our being. Nor is such a question any way material to our present purpose. We may draw inferences from the coherence of our perceptions, whether they be true or false; whether they represent nature justly, or be mere illusions of the senses. (T: 84)

In this passage one might even read Hume as a potential correspondence theorist of truth, with commitments to an extra-mental world. On basis of his work as a whole however one should not regard Hume in this way. Indeed one should strongly and actively refrain from drawing such a conclusion, because of the nature of Hume's philosophical project as laid out in § 2.

7.1.2. The unrepresentative character of 'passions'

In what follows, I discuss the relationship between 'reason', 'passion' and 'representation'. In § 6, I presented part of Cohon's *Representation Argument*:

1. Reason is the discovery of truth and falsehood.
2. Truth or falsehood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact.
3. Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. (Cohon, 2008: 69)

So, whatever corresponds either to real relations of ideas or to real matters of fact can be true and false according to Cohon's Hume (a claim I agree with). Now, Hume argues:

A passion is an original existence, or, if you will, modification of existence, and contains not any representative quality, which renders it a copy of any other existence or modification... 'Tis impossible... that this passion can be oppos'd by, or be contradictory to truth and reason; since this contradiction consists in the disagreement of ideas, consider'd as copies, with those objects, which they represent. (T: 415)

So, what does Hume mean when he says that a passion does not contain any representative quality? What he seems to mean is that the passion produced is not a copy of its cause in the way an idea is a copy of the impression from which it is derived. Note that part of what it is to be an impression is to be an original, that from which copies are made. Passions, like all impressions, are not copies of anything else (see Cohon & Owen, UM: 9-10) In *particular, they are not copies of other items, as ideas are copies of the impressions that cause them and that they consequently represent* (Cohon, 2008: 19). Cohon argues: “*They are not signs or symbols of a reality beyond them*” (Cohon, 2008: 19; *my italics*). Now, there seems to be something ambiguous, incomplete or even wrong about Cohon’s views in italics above. The problem with them first of all is that it is not clear what Cohon means by ‘reality’ in the quotation above. If the term ‘reality’ here (also) refers to something that is opposed to fiction rather than simply to a feature of the nature of an impression (being ‘real’), the problem with Cohon’s views *could already be* that often a single impression for Hume does not equal that which for Hume deserves the title ‘reality’. If the word ‘reality’ refers to something that is opposed to fiction, there is *certainly* a problem with Cohon’s views in italics given that sometimes for Hume *not even a complex of impressions* equals that which for Hume deserves the title ‘reality’. Cohon’s first half of the representation argument about truth and reason seems to argue that truth and falsehood are about agreement and disagreement between an idea (or complex of ideas) that is a copy of that which the copy represents. In regard to passions then Cohon’s argument seems to be that a passion is not a copy and does not refer and that contradicting reason and truth requires disagreement between an idea (or complex of ideas) that copies that real something which it is a copy of where this real something seems to be a truth-maker. Here then it should be noted that the ‘reality’/the object that for Hume is a truth maker is *at least not always* a single impression and at least not always a complex of impressions. It is *certainly not* a single impression if (a) we regard Hume as an *epistemologist* about (b) matters of *causality*. In the *epistemological case of causal matters of fact*, the truth-makers are repeated experiences of impressions of

sensation *accompanied by proper reasoning activity and results of a healthy functioning imaginative faculty with its capacity for making a variety of types of associations.* And in the *cognitive psychological case of causal matters of fact* the explanatory causal story also appeals to repeated impressions plus the principles of association. In the case of *objects* the epistemologist Hume might allow for a single impression having epistemological power by mean of some sort of trust in our natural constitution (see T: 183). If we regard Hume as a *cognitive psychologist*, probably the truth-maker sometimes is a single impression, while sometimes only repeated impressions will do the job.

7.2. The 'original' existence of passions

Central to Hume's philosophy about passions are the concepts of 'pleasure' and 'pain'. Hume argues:

The chief spring or actuating principle of the human mind is pleasure or pain; and when these sensations are remov'd, both from our thought and feeling, we are, in a great measure, incapable of passion or action, of desire or volition. (T: 574; cfT: 118)

Before I say a little more about pleasure and pain, I first discuss Hume's ideas about 'impressions of sensation' and 'impressions of reflexion'.

In book I of the *Treatise* Hume divides impressions into two kinds: impressions of sensation and impressions of 'reflexion'. About the first, Hume argues that "[they] arise...in the soul originally, from unknown causes" (T: 7). Impressions of reflexion by contrast are derived "in a great measure" from our ideas (T: 7). For Hume impressions of reflexion are antecedent to their correspondent ideas, but posterior to those of sensation, and derived from them. (T: 8). Hume elucidates this idea as follows:

An impression first strikes upon the senses, and makes us perceive heat or cold, thirst or hunger, pleasure or pain of some kind or other. Of this impression there is a copy taken by the mind, which remains after the impression ceases; and this we call an idea. This idea of pleasure or pain, when it returns upon the soul, produces the new impressions of desire and aversion, hope and fear, which may properly be called impressions of reflexion, because derived from it. These again are copied by the memory and imagination, and become ideas; which perhaps in their turn give rise to other impressions and ideas. (T: 7-8)

In book II Hume engages with exactly the same subject matter²¹, although he uses different terms to refer to the distinction between impressions addressed in book I. Rather than about *impressions of sensation* and *impressions of reflexion*, Hume talks about *original* and *secondary impressions*. He argues:

Original impressions or impressions of sensation are such as without any antecedent perception arise in the soul, from the constitution of the body, from the animal spirits, or from the application of objects to the external organs. Secondary, or reflective impressions are such as proceed from some of these original ones, either immediately or by the interposition of its idea. Of the first kind are all the impressions of the senses, and all bodily pains and pleasures: Of the second are the passions, and other emotions resembling them. (T: 275; my italics)

So passions and “other emotions resembling them” (it is unclear to me how to understand this) are ‘impressions of reflection’ while bodily pains and pleasures are impressions of sensation.²² About bodily pains and pleasures Hume makes the following further remark that

²¹ Hume himself says: “This division of the impressions is the same with that which I formerly made use of when I distinguish’d them into impressions of *sensation* and *reflection*” (T: 275).

²² Hume then goes on arguing that “’Tis certain, that the mind, in its perceptions, must begin somewhere; and that since the impressions precede their correspondent ideas, there must be some impressions, which without any introduction make their appearance in the soul” (T: 275) Hume does not continue to say something about these impressions’ natural and physical causes (they have according to Hume, T: 275-6). An examination (and exposition) of these causes, Hume argues would lead him too far from his present subject (see T: 275-6).

will turn out to be of some indirect importance (to be explained later in this section) in regard to his moral philosophy. Hume argues that bodily pleasures and pains are:

the source of many passions, both when felt and consider'd by the mind; but arise originally in the soul, or in the body, whichever you please to call it, without any preceding thought or perception. A fit of the gout produces a long train of passions, as grief, hope, fear; but is not deriv'd immediately from any affection or idea. (T: 276)

In the above quotation Hume mentions that bodily pleasures and pains “arise originally in the soul, or in the body, whichever you please to call it”. This remark is important in relation to his moral philosophy and it is so because of a conceptual ambiguity in Hume’s use of the term ‘original’.

When Hume talks about pleasure and pain he does not always talk about the same thing. Sometimes that is more obvious than at other times. As we have seen there is talk about *bodily* pleasures and pain. There is however also a large amount of talk about ‘mental’ pleasures and pain, most prominently in the sense that when concerned with moral matters as a cognitive psychologist Hume argues that character traits and actions we human beings *approve of* are either those that *produce pleasure* for others (individuals or society) or for the possessor himself or else are action that are *useful* to the possessor himself or others. On the negative side, we disapprove of character traits²³ and actions that yield pain (harm/uneasiness) to the person who has them or others.

What is also included in the all-encompassing talk about ‘pleasure’ and ‘pain’, are Humean claims about the existence of “natural impulses” and “natural instincts” in human beings. Below I give two quotations –(A) and (B)- that are similar in content. The first

²³ Hume describes a character trait as “a durable principle of the mind” (T: 575; cf T: 411–12).

directly supports the claim that there is a link between ‘pleasure’ and ‘pain’ and ‘natural impulses and instincts’ in Hume’s work:

(A) *Beside good and evil, or in other words, pain and pleasure*, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites. These passions, properly speaking, produce good and evil, and proceed not from them, like the other affections. (T: 439; my italics)

(B) Now ‘tis certain that, there are certain calm desires and tendencies, which, tho’ they be real passions, produce little emotion in the mind, and are more known by their effects, than by the immediate feeling or sensation. These desires are of two kinds; either *certain instincts originally implanted in our natures*, such a benevolence and resentment, the love of life and kindness to children; or the general appetite to good, and aversion to evil, consider’d merely as such. (T: 417; my italics)

As mentioned these two passages are similar in content. We need them together in order to get my claim off the ground that there is an ambiguity in Hume’s talk about the term ‘original’. Indeed we need the two quotations above (on the one side) plus Hume’s claim that bodily pleasures and pains “arise originally in the soul, or in the body, whichever you please to call it”. I will now offer an explanation of the ambiguity.

As we have seen Hume regards bodily pleasures and pains as ‘impressions of sensation’. When Hume talks about bodily pleasures and pains being impressions of sensation Hume uses the term ‘original’ partly or even exclusively as a necessary defining feature of an impression. That is, when Hume says that bodily pleasures and pains arise ‘originally’ in the soul, he seems to mean that these bodily pleasures and pains are perceptions ‘from which no further copies can be made’. Importantly, when it comes to *this* use of the term ‘original’, what typically, if not standardly and necessarily, is implied for Hume is the view that the

impressions he talks about arise from unknown causes (see T: 7). That is, when it comes to the mentioned use of ‘impression’ there is *no Humean commitment to the causes of the impressions*.

The way just mentioned seems to be the most dominant way in which Hume uses the term ‘original’. It does not seem to be the only way however. Referring to the quotations about natural instincts and impulses, rather than ‘from which no further copies can be made’, for Hume the term ‘original’ sometimes seems to mean: ‘from our natural constitution’. When Hume uses the term ‘original’ in this way, in contrast to what counts for the first way, *a cause is implied*. Given that Hume links our mental discernment of good and evil to our natural constitution (in a variety of ways, see the quotations above), while he is mostly a moral *psychologist* (rather than *biologist* or *aetiologist*), one wonders –at least I wonder– what exactly Hume wants to say about pleasure and pain as bits of our innate constitution (i. as actualized forms and ii. as pre-dispositions) necessarily involved in the *mental* discernment of moral distinctions, distinctions between good and evil. If there are such bits, the first question is how they are similar in terms of force with which they strike the soul as bodily impressions. There may be different things to say for different bits of our innate constitution. Secondly, the question is how they are conceptually similar to and different from bodily impressions of pleasure and pain.

7.3. Reason and calm passions

Space prevents me from discussing all Hume says about passion. In this section and the next I will address two further subtopics. In this section I discuss Hume’s notion of calm passions. The justification for that is that Hume often linguistically relates talk about ‘calm passions’ to talk about ‘reason’. In the next section, I discuss Hume’s account of desire. This I do with an

eye on the next chapter where I focus on the intellectual risks of misinterpreting Hume for contemporary meta-ethics.

In the very beginning of book 2 of the *Treatise* Hume argues:

The reflective impressions may be divided into two kinds, *viz.* the *calm* and the *violent*. Of the first kind is the sense of beauty and deformity in action, composition and external objects. Of the second are the passions of love and hatred, grief and joy, pride and humility. (T: 276)

Hume notes that the division is far from exact (T: 276). He also notes that the causes as well as the effects of both the violent and the calm passions are “pretty variable” and “depend in a great measure, on the peculiar temper and disposition of every individual (T: 437).

As mentioned, Hume regularly links the calm passions to a notion of ‘reason’. We need to be very careful however as to how to interpret Hume here. I think we should *not* interpret him as Kenneth Westphal does. Westphal argues “Of course Hume held that ‘reason’ is nothing other than ‘calm’ passion or passions” (Westphal, 2010: 125). One way, the relevant way here, in which this claim seems problematic is that it seems to conflate Hume’s philosophical notion of reason with his critical utterances about what ordinary people think about their own mental states. It needs to be pointed out that Hume seduces us into thinking that Westphal’s view is true. In his *Treatise* Hume argues:

What we commonly understand by *passion*, is a violent and sensible emotion of mind, when any good or evil is presented, or any object, which, by the original formation of our faculties, is fitted to excite an appetite. By *reason* we mean affections of the very same kind with the former; but such as operate more calmly, and cause no disorder in the temper: Which tranquility leads us into a mistake concerning them, and causes us to regard them as conclusions only of our intellectual faculties. (T: 437)

And a little later in the *Treatise* he argues:

This language will be easily understood, if we consider what we formerly said concerning that *reason*, which is able to oppose our passion; and which we have found to be nothing but a general calm determination of the passions, founded on some distant view or reflection .(T: 583)

It must be said that the ‘we’ in the quotation above is an incredibly nasty one and one that could have seduced Westphal into thinking that Hume equates ‘calm passions’ with ‘reason’ and one that might seduce one into thinking that Westphal is right. However, on basis of Hume’s work as a whole and on basis of further particular passages on calm passions I strongly believe we should not conclude from the above passages that Hume wants to equate ‘reason’ with ‘calm passions’. When Hume gives us the impression to equate reason with calm passions, he tells us something about our vulgar notion. What he tells us is that *because* the calm and tranquil action of the mind in reasoning is similar to the operation of calm desires, those “who judge of things from the first view and appearance” confound the two, and conclude that reason has an influence on action (T: 417, cited in McIntyre, 2000; *cf* 419, EPM: 239). Hume wants to say that *vulgarly* calm passions are called ‘reason’ (see T: 419), because of their similar influence on the mind. He does not want to say that reason and calm passions are conceptually or metaphysically similar. They have some resemblance in psychology in that both are not violent. That much is true, but that is also where the resemblance ends.

7.4. Hume on ‘desire’

For Hume a desire is a particular type of passion. Unfortunately, Hume is not clear on what exactly a desire is, partly through him *not* being *precise* about the concept (see IN: 11; T: 382,

417; cfT: 7).²⁴ What *we can* get from Hume's work is that a desire is a 'direct' passion. Hume makes a distinction between direct and indirect passions. For Hume, 'direct passions' are passions "as arise immediately from good or evil, from pain or pleasure" (T: 276). Under the direct passions, Hume lists: desire, aversion, grief, joy, hope, fear, despair and security (T: 277). 'Indirect passions' by contrast are those that "proceed from the same principles, but by the conjunction of other qualities" (T: 276). Passions that fall into this category are: pride, humility, ambition, vanity, love, hatred, envy, pity, malice, generosity "with their dependents". (T: 276-7). Hume's account of direct passions and indirect passions is rather complex. The complexity of the latter shines through the extensive complex explanation Hume provides of it in his work. Hume does not deal much with the former and gives some at face value clear descriptions of it. But there seems to be much more complexity behind Hume's account of direct passions than a first look at his work reveals. There is no easy and straightforward theory about Hume's account of direct passions (for a substantive explanation and justification of this claim see McIntyre, 2000).

One more thing that can be said about Hume's conception of desire is that he makes a distinction between principle and subordinate desires. Hume argues:

any principal desire may be attended with subordinate ones, which are connected with it, and to which if other desires are parallel, they are by that means related to the principal one. Thus hunger may oft be consider'd as the primary inclination of the soul, and the desire of approaching the meat as the secondary one; since 'tis absolutely necessary to the satisfying that appetite. If an object, therefore, by any separate qualities, inclines us to approach the meat, it naturally increases our appetite; as on the contrary, whatever inclines us to set our victuals at a distance, is

²⁴ Below are two Humean expressions that indicate a lack of precision:

"We are, therefore, to look for instances of this peculiar relation of impressions only in such affections, as are attended with a certain appetite or desire; such as those of love and hatred."(T: 382)

"*Benevolence* or the appetite, which attends love, is a desire of the happiness of the person below'd, and an aversion to his misery." (T: 382)

contradictory to hunger, and diminishes our inclination to them. Now 'tis plain that beauty has the first effect, and deformity the second: Which is the reason why the former gives us a keener appetite for our victuals, and the latter is sufficient to disgust us at the most savoury dish, that cookery has invented. All this is easily applicable to the appetite for generation. (T: 394-5)

Be that a conceptual distinction one finds in Hume's work about desire, this conceptual distinction is textually related to the lack of precision about the concept. In the abovementioned quotation, Hume seemingly uses the word 'primary desire', 'inclination' and 'appetite' in an interchangeable way. Generally, it has to be concluded that Hume is not very precise and clear about what a desire is. While that seems to be a fact, there is some limited textual evidence that at least indicates that Hume wants to separate 'desire' (as a noun) from 'willing'. Hume argues:

The act of the mind, exprest by a promise, is not a *resolution* to perform any thing: For that alone never imposes any obligation. Nor is it a *desire* of such a performance: For we may bind ourselves without such a desire, or even with an aversion, declar'd and avow'd. Neither is it the *willing* of that action, which we promise to perform: For a promise always regards some future time, and the will has an influence only on present actions (T: 516-7; cfEHU: 18).

For Hume the type of distinction that applies to the noun 'desire' and to 'willing' also seems to apply to the *verbs* 'to desire' and 'to will'.

8. Hume's view on morality

8.1. Morality and evaluative distinctions

In this chapter, I have discussed Hume's view of 'reason' in detail. I have focused on 'reason' as '*reasoning*' and also on how 'reason' as '*reasoning*' relates to certain types of 'judgment'.

With regards to reason as reasoning and the latter's relation to 'judgment', for Hume the moral case in one way is similar to the case about 'matters of fact', while it is also different in some significant way. It is similar because Hume with regard to both the moral case and the matters of fact case wants to say that psychologically speaking and normatively speaking evaluative distinctions in both areas are constitutionally dependent on certain types of impressions. The moral case and the case about 'matters of fact' are different in that the nature of normative distinctions is significantly different for the case about 'matters of fact' and the moral case. This counts both for the cognitive psychologist Hume and for the moral epistemologist Hume

It is not an easy task to determine in what guises Hume shows up in his moral philosophy. Hume certainly is a *cognitive moral psychologist*. He certainly is also a *sceptical* and *polemical philosopher* uttering certain views about the limitations of reason. In contrast to other areas of his work, in the domain of morality Hume strongly works with a notion of 'reason' that is a philosophical object of challenge for him. More about this directly. Besides the two guises mentioned, Hume might also adopt the guise of a normative ethicist and/or constructive moral epistemologist about the nature of moral normativity. As to the latter, I assume here for the sake of the argument and possible reality that Hume thinks that there are real normative distinctions to be made in the moral area.

Let me say something more about Hume's polemical guise in regard to moral matters. Hume aims to challenge ideas from his contemporaries, in particular Samuel Clarke. Hume's ideas about the limits of reason in regard to moral matters should be understood against the background of theories about the relationship between God and morality as endorsed in the 17th and 18th century. I will say something about that now, albeit only briefly.²⁵

In the 17th and 18th century the religious-cum moral environment in what is now called the United Kingdom was such that many people tried to save God from one horn of the Euthyphro dilemma. That is, many tried to make sure that 'voluntarism' or -using

²⁵ For a more extensive account see Jerry Schneewind's (2000).

another term for the same idea - 'divine command theory' does not imply a complete *carte blanche* for God as an omnipotent being. It was felt that if God as the highest being with an omnipotent power determines morality then he could be a tyrant. It was felt that something needed to be done in order to prevent the possibility that God's omnipotent will could be compatible with him being a tyrant. Some theory needed to be developed that prevented the possibility of God being a tyrant, a theory that somehow needed to be compatible with the omnipotence and goodness of God. God's will must be governed by eternal measures of right and wrong, but somehow without them being in conflict with God's essential features, especially also goodness and omnipotence. The view that then was put forward was the idea that there are necessary eternal truths, truths that would have to be recognized by any rational agent, no matter how situated, and so by God as well (see Schneewind, 2000). Let me now illustrate this contextual information by means of a passage from Hume's Treatise:

Those who affirm that virtue is nothing but a conformity to reason; that there are eternal fitnesses and unfitnesses of things, which are the same to every rational being that considers them; that the immutable measures of right and wrong impose an obligation, not only on human creatures, but also on the Deity himself: *All these systems concur in the opinion, that morality, like truth, is discern'd merely by ideas, and by their juxtaposition and comparison.* In order therefore, to judge of these systems, we need only consider, whether it is possible, from reason alone, to distinguish betwixt moral good and evil, or whether there must concur some other principles to enable us to make that distinction.²⁶ (T: 456-7; my italics)

²⁶ Compare: "According to the principles of those who maintain an abstract rational difference betwixt moral good and evil, and a natural fitness, and unfitness of things, 'tis not only suppos'd, that these relations, being eternal and immutable, are the same, when consider'd by every rational creature, but their *effects* are also suppos'd to be necessarily the same; and 'tis concluded they have no less, or rather a greater, influence in directing the will of the deity, than in governing the rational and virtuous of our own species. These two particulars are evidently distinct. 'Tis one thing to know virtue, and another to conform the will to it. In order, therefore, to prove, that the measures of right and wrong are eternal laws, *obligatory* on every rational mind, 'tis not sufficient to shew the relations upon which they are founded: We must also point out the connexion betwixt the relation and the will; and must prove that this connexion is so necessary, that in every well-disposed mind, it *must* take place and have its influence; tho' the difference betwixt these minds be in other respects immense and infinite. Now besides what I have already prov'd, that even in human nature no relation can ever alone produce any action; besides this, I say, is has been shewn, in treating of the understanding, that there is no connexion of cause and effect, such as this is suppos'd to be, which is

So, what we should keep in mind is that Hume's thoughts *about morality* and *from the angle of* his writings on morality and *reason*, i.e. rather than from the angle of morality and *sentiment*, are at least partly, if not exclusively, polemical. As polemical thoughts Hume's thoughts are directed at 17th and 18th century ideas about "eternal fitnesses and unfitnesses of things" and "immutable measures of right and wrong" (Hume *probably* regarded these terms as interchangeable).

8.2. Morality and the 'features true and false'

For Hume passions are a necessary component of moral distinctions. In § 7, I have argued that for Hume passions cannot be 'reasonable' or 'unreasonable'. In § 6, we discussed Hume's ideas of 'truth' and 'falsehood'. The issue of whether for Hume there are such things as 'truth' and 'falsehood' in the moral domain is an issue that is rather hard to settle. Reasons are as follows.

- (1) There seems to be strong, indeed conclusive textual evidence that suggests that for Hume moral *actions* (as well as passions) cannot be true or false. It very much looks like for Hume *actions* cannot be truth-bearers.

discoverable otherwise than by experience, and of which we can pretend to have any security by the simple consideration of the objects. All beings in the universe, consider'd in themselves, appear entirely loose and independent of each other. 'Tis only by experience we learn their influence and connexion; and this influence we ought never to extend beyond experience.

Thus it will be impossible to fulfill the first condition required to the system of eternal rational measures of right and wrong; because it is impossible to shew those relations, upon which such a distinction may be founded; And 'tis as impossible to fulfill the *second* condition; because we cannot prove *a priori*, that these relations, if they really existed and were perceiv'd, would be universally forcible and obligatory" (T: 465-6; italics are Hume's; embolding is mine; cf 413, 463; see also ST:9).

(2) Having said that, secondly, there is *very limited*, but *some* textual evidence that suggests that for Hume moral *thought and/or language* unlike actions *can* be true or false.

(3) Thirdly, there seems to be conclusive textual evidence that suggests that *mental judgements and language* with moral content as well as *moral actions* all *have moral qualities*. But, definitely in the case of actions (and probably in the case of mental judgements and language), these are not qualities we can refer to by means of the terms 'true' and 'false'. And neither can they be referred to by the label 'reasonable' and 'unreasonable.'

To illustrate the thoughts above let us look at a longer passage from Hume's *Treatise*:

Reason is the discovery of truth or falsehood. Truth or falsehood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact.

Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. Now 'tis evident our passions, volitions and actions, are not susceptible of any such agreement or disagreement; being original facts and realities compleat in themselves, and implying no reference to other passions, volitions and actions. 'Tis impossible therefore, they can be pronounced either true or false, and be either contrary or conformable to reason.

This argument is of double advantage to our present purpose. For it proves *directly*, that actions do not derive their merit from a conformity to reason, nor their blame from a contrariety to it; and it proves the same truth more *indirectly*, by shewing us, that as reason can never immediately prevent or produce any action by contradicting or approving of it, it cannot be the source of the distinction betwixt moral good and evil, which are found to have that influence. Actions may be laudable or blameable; but they cannot be reasonable or unreasonable: Laudable or blameable, therefore, are not the same with reasonable or unreasonable. The merit and demerit

of actions frequently contradict, and sometimes controul our natural propensities. But reason has no such influence. Moral distinctions, therefore, are not the offspring of reason. Reason is wholly inactive, and can never be the source of so active a principle as conscience, or a sense of morals. (T: 458)

What Hume says in this passage is first, that “reason is the discovery of truth and falsehood”. Secondly, he argues that “[t]ruth or falsehood consists in an agreement or disagreement either to the *real* relations of ideas, or the *real* matters of fact”. Thirdly, passions, volitions and actions are not susceptible of such an agreement or disagreement. What therefore follows are Hume’s conclusions (i) “passions, volitions and actions cannot be pronounced true or false” and (ii) passions, volitions and actions cannot be conformable or contrary to reason, i.e. be ‘reasonable’ or ‘unreasonable’.

Hume, furthermore (fifthly), in this passage, seems to argue for an identity-relationship between ‘being reasonable’ (unreasonable) and being susceptible to truth and falsehood (*cf*T: 415). Sixthly, Hume argues that actions do have ‘merit’ and that they can be ‘laudable’ and ‘blameable’. What we have to conclude from that is that there is a non-identity assumption in Hume about the properties ‘good’ and ‘evil’, ‘merit’ and ‘demerit’, ‘laudable’ and ‘blamable’ (on the one hand) and reasonable and ‘unreasonable’ (on the other hand).^{27 28}

²⁷ Having said that, in regard to the abovementioned points - especially 5 and 6 - I invite the reader to an exegetical challenge generated by two passages from Hume’s work, and seemingly *only* those passages (I have not found any similar one’s that contribute to the challenge). The challenge comes partly from the individual passages and partly from the union of both. Below I present and compare and contrast a passage from *EPM* with one from Hume’s ‘A dissertation on the passions’. In *EPM*, Hume argues:

“The end of all moral speculations is to teach us our duty; and, by proper representations of the deformity of vice and beauty of virtue, beget correspondent habits, and engage us to avoid the one, and embrace the other. But is this ever to be expected from inferences and conclusions of the understanding, which of themselves have no hold of the affections nor set in motion the active powers of men? *They [inferences and conclusions of the understanding] discover truths: but where the truths which they discover are indifferent, and beget no desire or aversion, they can have no influence on conduct and behaviour...* What is intelligible, what is evident, what is probable, what is *true*, procures only the cool assent of the understanding; and gratifying a speculative curiosity, puts an end to your researchers.” (EPM: 172; my italics)

And in ‘A dissertation on the passions’ Hume says something very similar, though at least linguistically somewhat different:

“It seems evident, that reason, in a strict sense . . . can never, of itself, be any motive to the will, and can have no influence but so far as it touches some passion or affection. *Abstract relations* of ideas are the objects of curiosity, not of volition. And *matters of fact, where they are neither good nor evil, where they neither excite desire nor aversion*, are totally indifferent; and whether known or unknown, whether mistaken or rightly apprehended, cannot be regarded as any motive to action.” (DP: 5.1)

As to the similarities, we can conclude the following. First, in both passages Hume talks about something that seems to be related to his notion of ‘truth’. Secondly, in both passages we see a remark appearing about that ‘truth related something’ *not* generating desire or aversion. The noteworthy difference however between these apparently very similar passages is that in the passage from *EPM* Hume talks about ‘truths’ while in the passage from ‘the Doctrine of Passions’ Hume talks about ‘matters of fact’ instead of ‘truths’. The union of both this difference and the similarity of the passages might have to invite us to consider and reconsider what the notion of ‘matters of fact’ amounts to for Hume and how it relates to ‘truth and falsehood’. There is some reason to interpret those passages as saying that while some truths and matters of fact can have no influence on conduct and behaviour/do neither beget desire nor aversion, other truths and matters of fact can. And here then one could read those passages also as saying that such truth and fact can be of a moral kind (*cf* Cohon, 2008: 50).

Given then that Hume’s general inclination seems to be to identify the ‘reasonable’ and ‘unreasonable’ with the ‘true’ and the ‘false’ (for indirect support see T: 415), this should also make us have some further thoughts about the relationship between matters of fact and the reasonable and unreasonable. Here then we face the fact that Hume argues that truths are of two kinds: either they consist in matters of facts or relations of ideas. And these two categories seem to be exhaustive of truths. Now, throughout his work, Hume gives the impression that the former relates strongly, indeed exclusively to truth about objects we can discover by means of causal reasoning. Truths concerning relations of ideas fall within the province of mathematics by being demonstrable (note the complexities identified in §§ 3 and 4). This brings us to the following conclusion: if we stick to the two categories as being exhaustive of ‘truths’, and if, for the sake of the argument, we assume that there are such things as *moral* truths (and let us assume that they can be called ‘moral truth’ and ‘moral matters of fact’), then such truths *must* join the truths about objects we discover by causal probabilistic reasoning –that can never go beyond sense-data- in the category of matters of fact. It is excluded that for Hume moral truths fall within the province of the other category. He is at pains to argue against that.

When reflecting on the relationship between ‘matters of fact’ and morality in Hume’s work, there is a further matter to focus on in the form of the following famous passage:

“*Nor* does this reasoning only prove, that morality consists not in any relations, that are the objects of science; but if examin’d, will prove with equal certainty, that it consists not in any *matter of fact*, which can be discover’d by the understanding. This is the *second* part of our argument; and if it can be made evident, we may conclude, that morality is not an object of reason. But can there be any difficulty in proving, that vice and virtue are not matters of fact, whose existence we can infer by reason? Take any action allow’d to be vicious: Wilful murder, for instance. Examine it in all lights, and see if you can find that matter of fact, or real existence, which you call *vice*. In which-ever way you take it, you find only certain passions, motives, volitions and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflection into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action. Here is a matter of fact; but ’tis the object of feeling, not of reason. It lies in yourself, not in the object. So that when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it. Vice and virtue, therefore, may be compar’d to sounds, colours, heat and cold, which, according to modern philosophy, are not qualities in objects, but perceptions in the mind: And this discovery in morals, like that other in physics, is to be regarded as a considerable advancement of the speculative sciences; tho’, like that too, it has little or no influence on practice. Nothing can be more real, or concern us more, than our own sentiments of pleasure and uneasiness; and if these be favourable to virtue, and unfavourable to vice, no more can be requisite to the regulation of our conduct and behaviour.” (T: 468-469)

In this passage Hume strongly relates the term ‘matter of fact’ to a sentiment of disapprobation.

8.3. Reason as the subordinate guide of the passions

Hume famously argued that: “Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them” (T: 414-5). As to this famous metaphor, it can easily be misunderstood in the sense of ascribing to Hume an overdose of intention to denigrate of reason. When Hume argues that “Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them” (T: 414-5), he indeed seemingly wants to be critical about the powers of reason and perhaps even denigrate reason. However, Hume is very well aware of the fact that rather than being of no noteworthy use, reason in the domain of morality is important in moral *thought* and (potentially) subsequently for *moral action*. This is helpfully elucidated by Nuyen.

When Hume says that reason is the slave of passions, he does not say thereby that reason is unimportant. He is saying merely that reason alone does not move one to act, The force that propels one to action is the passion, whether it be love, or anger, or pride, or envy, or fear, or desire. Reason alone does not provide the motive. However, it is reason that does the groundwork, analyzing facts, perceiving relations, and drawing conclusions, all of which go to determine which action to perform, but the action will not be performed unless one of the passions is also present. Just because reason is the slave does not indicate that its work is not relevant. To extend the slavery metaphor, we may observe that slave owners supported slavery for the very reason that the slave's work was found indispensable; analogously, it must be observed that Hume is most unlikely to hold that reason plays no part in human action. To complete the analogy, we should say that while much could be achieved with slaves, it was the

²⁸ What about the concepts belonging to the terms ‘rational’ and ‘irrational’, terms that are popular in contemporary meta-ethics and that Hume does use, but only occasionally. My inclination is to think that Hume’s use of the adjectives ‘rational’ and ‘irrational’ is similar to, if not identical with, Hume’s use of the adjectives ‘reasonable’ and ‘unreasonable’. It is beyond the confines of my study to provide a more in-depth account of this scholarly problem.

masters who conceived the plan, and the masters must rank higher in the causal order. (Nuyen, 1984: 27)

So what we see here is that Hume did not want to claim at all that reason does not do any work in the area of morality. Indeed, at places, Hume even gives the impression that it is a necessary condition for moral decision-making. See for example the following passage from the *Treatise*:

One principal foundation of moral praise being supposed to lie in the usefulness of any quality or action; it is evident, that *reason* must enter for a considerable share in all decisions of this kind; since nothing but that faculty can instruct us in the tendency of qualities and actions, and point out their beneficial consequences to society and to their possessor. (EPM: 285; cf172-3)

So for Hume reason, in the business of morality, has the role of a useful assistant.

I think it is fair to approach the positive role of reason for morality also from the following angle: for Hume reason is important to avoid being subject to epistemic defects. Hume is unsympathetic to judgements and beliefs that result from a lack of proper employment of our capacities for deliberation and reflection. Superstition, prejudice and related practices are disapproved of by Hume (see especially his essay ‘On the standard of Taste’). We must avoid holding beliefs that result from such practices. In *EHU* he argues:

It must...be confessed, that [a moderate form of Pyrrhonian skepticism] is a necessary preparative to the study of philosophy, by preserving a proper impartiality in our judgments, and weaning our mind from all those prejudices, which we may have imbibed from education or rash opinion. To begin with clear and self-evident principles, to advance by timorous and sure steps, to review frequently our conclusions, and examine accurately all their consequences; though by these means we shall make both a slow and a short progress in our systems; are the only methods, by

which we can ever hope to reach truth, and attain a proper stability and certainty in our determinations. (EHU: 150; more on Pyrrhonian skepticism in § 10)

So proper employment of our reasoning faculties for Hume here seems to be a normatively mandatory source for getting rid of undesirable sources of belief such as superstition and prejudice. Having said that, what counts in the case of reason being a ‘useful guide/ provider of evidence for moral matters, counts here as well: it is not reason, but a passion that actually is and necessarily must be the ultimate mover of a moral action; reason is not such an ultimate mover and it could never possibly be.

9. Further elements of Hume’s moral philosophy

9.1. Hume’s account of ‘sympathy’

Hume’s moral philosophy is characterized by an important role for the principle of sympathy. We need to be extremely careful however not to connect this too quickly to our folk notion of what sympathy is. I take it that what we in daily life mean by ‘sympathy’ is something like a ‘warm, loving feeling of compassion’. Hume hardly uses the term in this way in his *Treatise* however. Having said that, in *EPM* Hume seems to be turning more towards our folk notion of sympathy.

Hume’s *Treatise* account of sympathy reflects the idea that passions are communicated from one mind to another, while the receiver comes to hold the passion of the other in some sort of veridical way. In a nutshell, Hume’s idea seems to be based in three components. First, the idea that humans fundamentally have a similar psychological make-up. Secondly, the idea at all times each of us possesses a maximally vivid and forceful impression of himself (see Cohon, 2008: 129). Thirdly, Hume’s ‘associationism’: his philosophical thoughts about a variety of associate mechanisms that take place in the mind. I

will now say a more as to how these three components constitute and reflect Hume's ideas about sympathy in the *Treatise*.

As mentioned, for Hume human beings have a fundamentally similar psychological and bodily make-up (T: 575-6; *cf*T: 318; OC: 4²⁹). This includes them being similar in their experience of passions. This similarity in psychological make-up makes it possible that the passions in the mind of any person I contemplate travel to my mind. For Hume there are three associative mechanisms in the human mind: resemblance, contiguity and cause and effect. According to Hume's associationism, the liveliness of one perception will be automatically transferred to those other perceptions in the mind that are related to it by resemblance, contiguity, or cause and effect. There is so much liveliness transmission that the idea of the passion in my mind (the mere thought of it) becomes an impression and I actually experience a veridical form of the passion the other person is experiencing. This is the general idea of Hume's *Treatise* account of sympathy. Now, according to Hume, clearly some people resemble one more than others in some specific respect such as age or nationality. If the other person resembles me not only in being human but in some further respect, or is contiguous with me in space or time, or bears me any causal (that is, familial) relation, this will enhance the enlivening process, making my passion livelier (see Cohon, 2008: 129-30). Because of the resemblance between us or alternatively my proximity to the observed person, some of the great liveliness of the impression of myself is transmitted to my idea of the other person's sentiment.

In general we can say that in the *Treatise*, for Hume 'sympathy' is a communication mechanism, something that comes close to or even reflects what we would nowadays call 'empathy'. At some points in his text Hume gives the impression that 'sympathy' is not a direct form of passion-communication. Hume argues:

²⁹ See page 170 of *EPM* for a view that at face value speaks against this.

No passion of another discovers itself immediately to the mind. We are only sensible of its causes or effects. From *these* we infer the passion: And consequently *these* give rise to our sympathy.³⁰
(T: 576)

Hume explains:

When I see the *effects* of passion in the voice and gesture of any person, my mind immediately passes from these effects to their causes, and forms such a lively idea of the passion, as is presently converted into the passion itself. In like manner, when I perceive the *causes* of any emotion, my mind is convey'd to the effects, and is actuated with a like emotion. Were I present at any of the more terrible operations of surgery, 'tis certain, that even before it begun, the preparation of the instruments, the laying of the bandages in order, the heating of the irons, with all the signs of anxiety and concern in the patient and assistants, wou'd have a great effect upon my mind, and excite the strongest sentiments of pity and terror. (T: 576)

Having said that, there is also some textual evidence that speaks against 'sympathy' *not* being a direct mechanism of passion-communication. Hume argues:

'Tis evident, that *sympathy*, or the communication of passions, takes place among animals, no less than among men. Fear, anger, courage and other affections are frequently communicated from one animal to another, without their knowledge of that cause, which produc'd the original passion. Grief likewise is receiv'd by sympathy; and produces almost all the same consequences, and excites the same emotions as in our species. The howlings and lamentations of a dog produce a sensible concern in his fellows. And 'tis remarkable, that tho' almost all animals use in play the

³⁰ Compare "When any affection is infus'd by sympathy, it is at first known only by its effects, and by those external signs in the countenance and conversation, which convey an idea of it. This idea is presently converted into an impression, and acquires such a degree of force and vivacity, as to become the very passion itself, and produce an equal emotion, as any original affection" (T: 317).

same member, and nearly the same action as in fighting; a lion, a tyger, a cat their paws; an ox his horns; a dog his teeth; a horse his heels: Yet they most carefully avoid harming their companion, even tho' they have nothing to fear from his resentment; which is an evident proof of the sense brutes have of each other's pain and pleasure. (T: 398)

The abovementioned passages seem to be in conflict as to whether or not 'sympathy' in the *Treatise* is a direct form of passion-communication or not. The passage directly above about animals seems to provide evidence in favour of it being a direct form of passion-communication, while the passage cited earlier in this section seems to provide evidence against this view given that Hume argues that "no passion of another discovers itself immediately to the mind" and, related, where he states that it is only via the causes and effect that we feel the passion of another.

The mechanism of emotional contagion that characterizes Hume's *Treatise* account of sympathy is also visible in *EPM*. (A comparative example is in this footnote.³¹) But, when having a look at *EMP*, while every now and then one spots this similarity, one cannot but conclude that Hume's use of the term 'sympathy' is very different in his *EPM* compared to the *Treatise*.³² In *EPM* one finds much more pro-social connotations of the term, every now and then mixed with the *Treatise* notion. Further to the simple term 'sympathy', in *EPM*

³¹ Hume in the *Treatise*: 'Tis evident, that *sympathy*, or the communication of passions, takes place among animals, no less than among men. Fear, anger, courage and other affections are frequently communicated from one animal to another, without their knowledge of that cause, which produc'd the original passion. Grief likewise is receiv'd by sympathy; and produces almost all the same consequences, and excites the same emotions as in our species. The howlings and lamentations of a dog produce a sensible concern in his fellows. And 'tis remarkable, that tho' almost all animals use in play the same member, and nearly the same action as in fighting; a lion, a tyger, a cat their paws; an ox his horns; a dog his teeth; a horse his heels: Yet they most carefully avoid harming their companion, even tho' they have nothing to fear from his resentment; which is an evident proof of the sense brutes have of each other's pain and pleasure" (T:398).

Hume in *EPM*: "Have we any difficulty to comprehend the force of humanity and benevolence? Or to conceive, that the very aspect of happiness, joy, prosperity, gives pleasure; that of pain, suffering, sorrow, communicates uneasiness? The human countenance, says Horace, borrows smiles or tears from the human countenance." (EPM: 220).

³² Occasionally, e.g. on page 604 of the *Treatise* one finds a pro-social notion of 'sympathy' or at least a sign of it. See e.g. (T: 604).

Hume uses a variety of compound terms and phrases that also point to a pro-social connotation of the term ‘sympathy’. For example, Hume talks about: “social sympathy” (EPM: 224) and “tender sympathy” (EPM: 178). Hume also says: “The same social sympathy, we may observe, or fellow-feeling with human happiness or misery” (EPM: 260); “our hearts are immediately caught, our sympathy enlivened, and our cool approbation converted into the warmest sentiments of friendship and regard” (EPM: 230); “[T]he natural symptoms, tears and cries and groans, never fail to infuse compassion and uneasiness” (EPM: 220).

Hume’s *EPM* account of ‘sympathy’ is quite puzzling. It is clear that there is more of a pro-social connotation to the term than in the *Treatise*, but the pro-social connotation is not always free from the *Treatise* meaning. This should be understood as saying that the two meanings get intertwined on one and the same occasion rather than Hume using sometimes the one meaning and at another time the other. The latter does happen, but only very rarely does a pure form of emotional contagion occur in the *Treatise*.

As to Hume’s puzzling notion of empathy in *EPM*, I finally point out that things, at least for me, get particularly puzzling when reflecting on Hume’s use of the verb ‘to sympathize’ (see e.g. EPM: 258; T: 316). When he uses this verb, it is often very ambiguous between passion-communication notion of the *Treatise* and pro-social notions. That raises the questions whether he means one of the two and if so which one or else, whether he wanted the term to cover a concept that embraces both the communication mechanism notion of the *Treatise* and the pro-social connotations prevalent in *EPM*.

9.2. The common point of view: a brief introduction

In this section and the next, I comment on some further specific aspects of Hume’s philosophy. This should aid a comprehensive understanding of Hume’s account of morality. That on its turn might help to deeply understand the role reason and passion play in Hume’s

philosophy. In this section, I address Hume's notion of 'the common point of view', while in the next I focus on some of Hume's ideas on justice and utility.

As to the common point of view, in book III of the *Treatise* Hume argues that when we make distinctions between morally good and evil, we don't do that from our own subjective idiosyncratic perspective (*cf*T: 581). Rather, Hume claims, "we fix on some steady and general points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation" (T: 581-2). He argues:

In general, all sentiments of blame or praise are variable according to our situation of nearness or remoteness, with regard to the person blam'd or prais'd, and according to present disposition of our mind. But these variations we regard not in our general decisions ... We consider not whether the person...be our acquaintance or strangers, countrymen or foreigners. Nay, we overlook our own interest in those general judgements; and blame not a man for opposing us in any of our pretensions, when his own interest is particularly concern'd. We make allowance for a certain degree of selfishness in men; because we know it to be inseparable from human nature, and inherent in our frame and constitution. By this reflection we correct those sentiments of blame, which so naturally arise upon any opposition. (T: 582-83; *cf*T: 586-7, 591; EPM: 228-9: 272-3, 308-309)

Hume introduces the common point of view in order to account for the fact that our moral judgments tend to remain constant, to converge with those of other people, and not to vary as idiosyncratically as one might expect given that they are manifestations of our individual feelings (Cohon, 1997, 2008).

It is also with regard to this aspect of Hume's philosophy that the exegetical question arises as to what the relationship is between Hume's words on the matter and (a) a descriptive science of men and/or (b) a normative view (*cf*Korsgaard, 1997 Cohen, 1990, Magri, 1996). When talking about the common or general point of view, Hume's talk is

largely descriptive, but there is an important question as to how that maps onto Hume's views on the common point of view. Does he really just want to be the cognitive psychologist? Or does he, primarily or in a secondary way, want to say that the general point of view is something mandatory we should adopt to correct our own idiosyncratic moral perspective? How exactly to make up the balance between cognitive psychology and normativity is a hard scholarly problem. Offering a conclusive answer to this question is outside the confines of this study.

9.3. Justice and utility: a brief introduction

For Hume 'justice' is an artificial construction that serves the purpose of regulating society. "The use and tendency of [justice] is to procure happiness and security, by preserving order in society" (EPM: 186; cf261). In Hume's work, the concept 'justice' has a clear relationship with the concept of 'the useful' or (public) 'utility', these two being terms that refer to a concept that fulfils an important role in Hume's philosophy. Hume argues that usefulness is agreeable to us and engages our approbation. He takes this to be a matter of fact, confirmed by daily observation (he even states that explicitly on EPM: 218). Be that a matter of fact, Hume asks:

But, *useful*? For what? For somebody's interest, surely. Whose interest then? Not our own only: For our approbation frequently extends farther. It must, therefore, be the interest of those, who are served by the character or action approved of; and these we may conclude, however remote, are not totally indifferent to us. By opening up this principle, we shall discover one great source of moral distinctions. (EPM: 218)

This idea then of our approbation of utility stretching farther than that of our own interest, in combination with Hume's view about justice makes Hume regard the relationship between justice and utility in the following way:

Thus we seem, upon the whole, to have attained a knowledge of the force of [the principle of public utility]..., and can determine what degree of esteem or moral approbation may result from reflections on public interest and utility. The necessity of justice to the support of society is the sole foundation of that virtue; and since no moral excellence is more highly esteemed, we may conclude that this circumstance of usefulness has, in general, the strongest energy, and most entire command over our sentiments.³³ (EPM: 203-04)

For Hume utility is "the foundation of the chief part of morals, which has a reference to mankind and our fellow-creatures." (EPM: 231). For Hume utility seems to be an innate sentiment. I think Hume should be read as saying that this sentiment, together with the natural sentiment of benevolence and something Hume calls the feeling of 'humanity' (a notion he hardly uses in the *Treatise*, but quite frequently in *EPM*) engages us to pay to the interest of mankind and society.

10. Hume's views on 'belief'

10.1. An introduction

³³ There is much more to say about Hume's account of justice for which I do not have space in this chapter. One further aspect of Hume's account of justice is represented by Hume's speculations about justice and *possible worlds*. Hume argues that "The rules of equity or justice depend entirely on the particular state and condition in which men are placed (EPM: 188). Hume imagines both worlds in which we have a different psychological constitution and worlds in which there are changes in external circumstances and argues that justice in these circumstances is something useless. And with regard to that he says "By rendering justice totally *useless*, you thereby totally destroy its essence, and suspend its obligation upon mankind". (EPM: 188). In relation to the topic of justice, Hume also offers a specific quite extensive account of property (For a more extensive treatment of Hume's ideas on justice see e.g. Flew (1976), Whelan, (1994), Cohon (1997), Taylor (1998), Watkins Tate 2005, Westphal (2010).

Michael Gorman (1993: 92) has helpfully drawn up the following list of types of descriptions of beliefs Hume offers in the *Treatise* and *EHU*:

1. An idea conceived in a certain manner. (e.g., EHU: 49, T: 96)
2. That certain manner of conception itself. (e.g., EHU 49, T: 97)
3. An idea that feels a certain way. (e.g., EHU: 48, T:103)
4. That certain feeling itself. (e.g., EHU: 49, T: 624³⁴, T: 629)
5. An idea that has a great influence on the mind. (e.g., T: 118-20)
6. An act of mind rendering realities influential on the mind. (e.g., EHU: 49, T: 629)
7. A lively idea related to an impression. (e.g., T: 96)
8. A lively manner of conceiving an idea, which manner arises from an impression. (e.g., EHU: 50); something that makes ideas forceful and vivacious. (e.g., T: 101, T: 627)
9. Something that makes ideas forceful and vivacious. (e.g., T: 101, T: 627)

Further to this, below I give four verbatim descriptions about belief one finds in Hume's work. In appendix I of the *Treatise* Hume says:

When I would explain this manner, I scarce find any word that fully answers the case, but am obliged to have recourse to everyone's feeling, in order to give him a perfect notion of this operation of the mind. An idea assented to feels different from a fictitious idea that the fancy alone presents to us; And this different feeling I endeavour to explain by calling it a superior force, or vivacity, or solidity or firmness or steadiness. This variety of terms, which may seem so unphilosophical is intended only to express that act of the mind which renders realities more present to us than fictions, causes them to weigh more in the thought and gives them a superior influence on the passions and imagination. Provided we agree about the thing, 'tis needless to dispute about the terms. ... I confess that 'tis impossible to explain perfectly this feeling or manner

³⁴ "Belief consists merely in a certain feeling or sentiment; in something that depends not on the will, but must arise from certain determinate causes and principles of -which we are not masters". (T: 624)

of conception. We may make use of words that express something near it. But its true and proper name is belief, which is a term that everyone sufficiently understands in common life. And in philosophy we can go no farther than assert that it is something felt by the mind, which distinguishes the ideas of the judgment from the fictions of the imagination. It gives them more force and influence; makes them appear of greater importance; infixes them in the mind; and renders them the governing principles of all our actions. (T: 629)

Here is another description from book I of the *Treatise*:

Belief consists merely in a certain feeling or sentiment; in something that depends not on the will, but must arise from certain determinate causes and principles of -which we are not masters. (T: 624)

And one more:

I am able to conclude from this idea, that such an impression did once exist; and as this conclusion is attended with belief, it may be ask'd, from whence are the qualities of force and vivacity deriv'd, which constitute this belief? And to this I answer very readily, *from the present idea*. For as this idea is not here consider'd, as the representation of any absent object, but as a real perception in the mind, of which we are intimately conscious, it must be able to bestow on whatever is related to it the same quality, call it *firmness, or solidity, or force, or vivacity*, with which the mind reflects upon it, and is assur'd of its present existence. The idea here supplies the place of an impression, and is entirely the same, so far as regards our present purpose. (T: 106)

And finally one from *EHU*:

The difference between fiction and belief lies in some sentiment or feeling which is annexed to the latter, not to the former, and which depends not on the will, nor can be commanded at pleasure. (EHU: 48)

If one looks at the descriptions, i.e. both Gorman's briefer descriptions and the verbatim passages above, there are two things that become very clear. First, that for Hume there is a clear link between a 'belief' and a particular –lively- *manner of conception*. Secondly, that for Hume there is an essential link between 'beliefs' and 'feeling'.³⁵

10.2. Mental states with a strong conviction coefficient

³⁵ Some scholars seem to equate 'judgements' and 'beliefs', something I think we should not do. In her (2008) Cohon for example cites the following passage (one we have seen already in § 6 and 8.2) from Hume's 'Dissertation on the passions':

"It seems evident, that reason, in a strict sense, as meaning the judgement of truth and falsehood, can never, of itself, be any motive to the will, and can have no influence but so far as it touches some passion or affection. *Abstract relations* of ideas are the objects of curiosity, not of volition. And *matters of fact*, where they are neither good nor evil, where they neither excite desire nor aversion, are totally indifferent; and whether known or unknown, whether mistaken or rightly apprehended, cannot be regarded as any motive to action." (DP: 5.1; Cohon, 2008: 50)

Cohon then comments on this passage by saying that "This wording strongly suggests an identification of reason with *judgments or beliefs*: it makes it seem that for Hume reason in a strict sense just is a set of judgments, so that the very judgments themselves—the beliefs—are identical with, or are parts of or instances of, reason" (Cohon, 2008: 69; my italics). Before I comment on this claim from Cohon, let us also consider the following claim from Ingmar Persson: "It is plausible to take 'reason' to be that faculty that manifests itself in the formation of true-or-false judgements or beliefs, for in Hume's own words it is that 'which judges of truth and falsehood'" (Persson, 1997:189).

I do not quite understand (at least not from a justificatory point of view) why it is Cohon and Persson equate 'judgements' with 'beliefs', if that is indeed what they do. Persson rather clearly seems to do and I am inclined to think that Cohon does so as well (if she does this is something that repeats itself throughout her book). Assuming that they do, I think they should not. In my opinion, there is no good reason to think, let alone *conclusive* reason to think, that for Hume 'judgement' and 'beliefs' conceptually are similar things. Neither does there seem to be conclusive reason to think that as psychological states 'judgements' and 'beliefs' are necessarily similar things. Hume counts both 'beliefs' and 'judgements' as perceptions of the mind. Be that as it seems to be, it could very well be the case that some judgements, but not all of them, count as beliefs. And it may also be the case that some beliefs or even all beliefs are judgements. It may also be the case that there is no mental perception that can be called a 'belief' and that can also be called a 'judgement'. Solving this issue is beyond the purposes of this chapter and not an easy matter, especially also given that there is no elaborate account of 'judgement in Hume's work' (while there is a highly elaborate account of belief).

Gorman concludes that inherent to his list of eight different descriptions are two main ways in which Hume lays out his account of belief;³⁶ two ways that correspond to the two essential features I have pointed to above. The first way Gorman calls the *Manner of Conception Theory*. The *Manner of Conception Theory* is supposed to reflect the view that a 'belief' is a particular sort of idea that is characterized by the fact that it is *conceived* in a lively way. The second way is the *Feeling Theory*. In contrast to the *MCT*-descriptions, Hume's *FT*-descriptions reflect the idea that a 'belief' is either an idea that feels a certain way or else that belief is that feeling itself.

I think Gorman's division is sensible and accurate. In what follows I attempt to make sense of the relationship between the two types of descriptions in terms of the idea of the nature of beliefs Hume had in mind. Possibly we should not say that the two descriptions refer to the same concepts, but they seem to be two different ways of describing (almost?) the same phenomenological state.

That phenomenological state can probably be best described by means of an appeal to what I take to be a helpful term from Laird (1993). Laird talks about Hume's beliefs having a strong 'conviction-coefficient'. For Hume, 'beliefs' are a sub species of the genus 'ideas', but their nature seems to be very close to impressions. Recall that about impressions Hume says that they "always actuate the soul and that in the highest degree" (T: 118). For Hume this claim seems to imply that impressions have a high conviction coefficient. Indeed it must be said that for Hume impressions are the type of atomic mental items that have *the highest* conviction coefficient (see Laird, 1939: 429).

As a species of the genus 'idea', what Hume calls 'beliefs' are no dry, flat, cold ideas that pop up in one's mind. Rather, while being part of the genus category 'ideas', beliefs as a species of ideas have a strong intensity resembling the intension of impressions and they also

³⁶ Or rather, Gorman argues that the first four formulations address the question 'what is a belief?', while these formulations can be reduced to two theories. Gorman argues that the second formulation is a poor version of the first formulation, while the fourth is a poor version of the third formulation. As to description 5-8, Gorman argues that description 5 and 6 address the question 'What does a belief do?', while description 7 and 8 take for granted the statements in the first and second categories respectively and address the further question, What causes beliefs? Gorman argues that description 9 is probably a case of bad writing.

have a particular experiential flavour. Note that about this experiential flavour Hume says, and says no more than, that it is a “*je-ne-scai-quoi*, of which ’tis impossible to give any definition or description” (T: 106). Hume himself thinks that everyone sufficiently understands this experiential feeling that cannot be well described (T: 106). Whether he is right on that I leave up to the reader, informed by his own (belief) experience.

Up until now, I have focused on two descriptions one finds in Gorman’s list of eight: belief as a particular feeling and belief as a particular manner of conception. I would like to pick out one more, namely description 7, that occurs in book I of the *Treatise*. Hume offers this description in capital letters, most probably indicating that it was of some sort of special importance to him: “An opinion... or belief may be most accurately defin’d A LIVELY IDEA RELATED TO OR ASSOCIATED WITH A PRESENT IMPRESSION” (Hume, T:96). So how is it we should understand this claim? Probably as follows. Our faculty of imagination can produce ideas. Also my, CJ’s, faculty of imagination can do that, even while writing this bit of text. Let’s say that my imagination does this right now and that the idea produced is that a house is on fire. Right now this idea is pretty cold and without much vivacity. Not much is happening in my mind right now. Now, it could be the case that once I look out of the window, I see smoke. If that happens, my mental state will change quickly, because of some relevant past experience. The smoke will remind me of the time I, with my eyes, really saw a house on fire. Memories pop up and while they do, I (most probably) automatically associate the smoke I see now with the thought that there is a house on fire. By that time, I will have a vivid idea in my mind about a house being on fire that dependent on further circumstances may reflect a belief that it is on fire (example adapted from Nuyen: 1988: 376). As to the qualifier ‘dependent on further circumstances’, for Hume it is not the case that every vivid idea in a person’s mind is necessarily a belief. If there is enough material in someone’s mind that in combination with her current sense-impressions can and do eliminate the immediate association then the person will not come to believe that a house is on fire (or if she was convinced for a short moment that a house is on fire, she will stop believing that it is).

10.3. Belief-producing causes

For Hume beliefs can arise from a wide variety of different causes. One way to organize beliefs from a perspective of their causal history is by seeing them arising from one of the three following causes: (1) abstract reasoning, (2) association and (3) education. Below I offer an introduction into the nature of the beliefs that fall in these three categories.

For Hume reasoning as a process/activity of 'interposition of ideas' is a *natural cause* of belief. Note however that *if* Hume can also be seen as an epistemologist, besides a natural cause, abstract reasoning, in one or more areas could also possibly be a *justifying ground for belief*. In any case, whatever guise(s) we ascribe to Hume, for Hume, the outcomes of abstract reasoning *as we use it in mathematics* can impress us with the 'evidence' or even 'certainty' of what has been discovered/proved (T: 70-71; EHU: 25-26, 163-164; cited in Falkenstein, 1997). Hume remarks that the strength of the belief generated by reasoning correctly about mathematical matters ordinarily equals the vivacity of memory, and exceeds the strength of the beliefs induced by causal association:

This force and this vivacity are most conspicuous in the memory; and therefore our confidence in the veracity of that faculty is the greatest imaginable, and equals in many respects the assurance of a demonstration. The next degree of these qualities is that deriv'd from the relation of cause and effect; and this too is very great, especially when the conjunction is found by experience to be perfectly constant, and when the object, which is present to us, exactly resembles those, of which we have had experience. (T: 153)

Let us now turn to the second category of causes: beliefs arising from association.

As Falkenstein argues, by far the most important object of investigation for Hume in terms of belief-forming mechanisms is association. The associative belief-forming mechanisms relate an imagined idea with some presently-had impression or memory. In the process of belief-formation, associative mechanisms inflict a portion of the vivacity of our

impressions or memories onto the imagined idea. If enough vivacity is injected into the idea, we come to believe it, with a degree of conviction that is proportioned to the quantity of vivacity (Falkenstein, 1997: 34). For Hume there are three sorts of associative relations. First, resemblance. Secondly, contiguity. Thirdly, constant conjunction. Hume maintains that associative relations of resemblance and contiguity do not reflect enough vivacity to induce belief.³⁷ Constant conjunction however *does* have this capacity. Noteworthy is that it can reflect varying quantities of vivacity, depending on how often the events are associated in experience, and whether there are ever any exceptions to their conjunction. Events that are frequently and invariably observed to occur in succession produce a “perfect habit” of association and, where this habit is in place, a great deal of vivacity can be reflected to the associated idea.

Space prevents me from giving an exhaustive account of beliefs by association. For more detail I refer to Falkenstein (1997). Having said that, in what follows directly I comment briefly on one particular subtype of ‘belief by association’, namely belief as arising from associations with passions.

Besides being induced by the transmission of vivacity from a previous sensory impression or memory, belief can also be induced by the transmission of vivacity from a subsequently produced sentiment or emotion (see Falkenstein, 1997). This is what happens in the case of the person suspended over an abyss in an iron cage (T: 148-149). The idea of an abyss is naturally associated with that of falling, and the impression of an abyss naturally transfers some vivacity to this associated idea, so that falling is believed to be a possibility. Ordinarily, this inference would be tempered by the contrary realization that falling is not possible through the solid bars of the cage; however, in the first instant, when the idea of falling is enhanced, it arouses a feeling of fear. That feeling increases our sensitivity to the

³⁷ Although the mechanisms of resemblance and contiguity do not normally transmit enough vivacity to induce belief, Hume takes it that they do have some effect, and as a result can enhance or mitigate the strength of beliefs formed by other mechanisms, or even modify their content. The most striking instances of this occur in the case of resemblance (Falkenstein, 1997: 37).

danger and makes us enhance the idea of falling (Falkenstein, 1997: 39). Hume himself describes this in the following way:

[The] passion returns back upon the imagination and **in**livens the idea; which lively idea has a new influence on the passion, and in its turn augments its force and violence; and both [the] fancy and affections, thus mutually supporting each other, cause the whole to have a very great influence. (T: 148-149)

That much about 'beliefs arising from association'. I now turn to the third category of belief-generating causes. A third cause for belief-formation is education. Hume takes it that beliefs induced in the student through education can take such deep root, that "tis impossible for us, by all the powers of reason and experience, to eradicate them; and this habit not only approaches in its influence, but even on many occasions prevails over that which arises from the constant and inseparable union of causes and effects" (T: 116). Reflected by the idea that 'beliefs about education' make up their own class, beliefs formed by education are not produced by association mechanisms (and neither are they produced by abstract reasoning).

Hume thinks that the human mind is so constituted that the *mere repetition* of an idea can enhance the vivacity with which it is entertained to the point where it comes to be believed. This phenomenon is witnessed in e.g. liars "who by the frequent repetition of their lies, come at last to believe and remember them, as realities" (T: 86). Its most common manifestation however is in education.

When Hume talks about beliefs arising from education, he compares it to the case of belief induced by causal inference. Causal inference is based on two factors: (a) the repeated experience of a conjunction between one object and another, which leads us to associate the one with the other and (b) the tendency of the mind to readily think of associated objects, so that when an impression of the one occurs, an idea of the other readily takes its place, in the process taking on a good deal of the vivacity possessed by the former (see Falkenstein, 1997). But what, Hume asks –in the context of writing about beliefs that arise from education –, if

this process were short-circuited, and the association with a second object left out, so that “a mere idea alone, without any of this curious and almost artificial preparation shou’d frequently make its appearance in the mind?” Hume answers his own question by saying that: “this idea must by degrees acquire a facility and force; and both by its firm hold and easy introduction distinguish itself from any new and unusual idea” (T: 116). Here, no association with some other object or transfer of vivacity from an impression of this object onto the idea is envisioned. The mere repetition of the idea serves to enliven it (Falkenstein, 1997).

Somewhat later in the *Treatise* Hume offers an adjustment of his account of ‘beliefs induced by education’. The qualified account follows from Hume’s thought that “a person, who wou’d *voluntarily* repeat any idea in his mind, tho’ supported by one past experience, wou’d be no more inclin’d to believe the existence of its object, than if he had contented himself with one survey of it” (T: 140). As Falkenstein points out, Hume is here concerned to explain why the belief in a causal relation could not be induced by having just one experience and then voluntarily recalling it a number of times. When he offers his qualification, Hume argues that the belief induced by education does *not* arise through any sort of repetition of the idea, but through its “*undesigned*” repetition at frequent intervals over a long tract of time (T: 140, Hume’s italics). This seems to mean that the student needs to come across the idea in a way that is unintended by him, as when hearing it from a number of others or coming across it in different books, not merely repeat it as an exercise or a response to questioning (see Falkenstein, 1997). The conclusion that follows from what I have just explained is that for Hume the case of the student who acquires a belief by education is different from the case of the liar who comes to believe his own lies through voluntary frequent repetition. That is, in terms of the psychological mechanisms that are involved. Besides the two cases for Hume being different in this way, for Hume the two cases also seem to be different in terms of pathological psychology. For Hume. the case of the liar is a pathological case-arising from some defect in our cognitive constitution. A non-

pathological healthy mind is aware of its volition in producing the repetition as characteristic of the mental state of the liar. This realization then leads it to consider the idea to be its own invention, regardless of how much the idea may be enlivened by repetition (see Falkenstein, 1997). This type of awareness is absent in the student who acquires beliefs through education. Be that the case, this particular absence of what Hume regards as a case of healthy mental functioning does not seem to mean that for Hume the educational case is a pathological one.

10.4. The role of reason in Hume's account of belief:

10.4.1. Reason as a provider of evidence

In the previous section, we have seen that beliefs can arise from abstract *mathematical* reasoning. In this section, I say something different and more generally about the relationship between Hume's concept of 'reason' and his account of 'belief'.

Sometimes, but only very occasionally, Hume gives his readers *some* reason to think that 'belief' has nothing to do with reason. Hume for example says: "When the mind...passes from the idea or impression of one object to the idea or belief of another, it is not determin'd by reason, but by certain principles, which associate together the ideas of these objects, and unite them in the imagination" (T: 92). And a little later in the *Treatise*, Hume tells us that we can easily "discover by experiments" that belief "arises immediately, without any new operation of the reason or imagination" (T: 102). That 'reason' has nothing to do with 'belief' however is not what Hume wants to argue for however. Below I elucidate and defend this claim.

As discovered by Nuyen (1988), on page 102 of the *Treatise* Hume argues: "belief ... arises immediately, without any *new* operation of the reason or imagination" (emphasis original). As Nuyen (1988) argues, this gives an indication that Hume does *not* say that belief

involves *no* operation reason *at all*; rather his claim seems to be that it involves no *new* operation of reason. We need our capacity to reason to draw the inference that *p*, but from *p* to the belief that *p*, no *new* reasoning is required (see Nuyen, 1988: 378). ‘Reason’ as a mental activity of reasoning will not suffice to turn *p* into a *belief that p*, but *without the work of ‘reason’*, there is *no possibility to have the belief that p*. In terms of an example, unless I infer that the house is on fire (on seeing smoke), I cannot believe that the house is on fire. Reason has to do its work for there to be beliefs. Reason has to make a causal inference. The inference in question is not simply a transition from one thought to another, but a causal inference which is the work of reason. However, all that reason as reasoning³⁸ does is *putting the inferences before the mind*. What happens *next* in terms of belief-formation is none of its doing (see Nuyen, 2008: 378). And what happens is that towards the inference, there arises this “je-ne-scai-quoi-non”-feeling, of which “’tis impossible to give any definition or description” (T: 629). Reason presents the possibilities to the mind, and then we are subject to a feeling that is part of what a belief is and results from past experience, associate mechanisms and the influence of nature. As Hume argues: “Nature, by an absolute and uncontrollable necessity has determin’d us to judge as well as to breathe and feel” (T: 183). More on nature’s role in belief formation in the next section.

Above I have argued for the view that reasoning plays a clear role in belief-formation. Having said that, it must be pointed out that in a number of passages in the *Treatise* Hume argues that the inference from impressions to ideas can be automatic, and escape our attention altogether (see Nuyen, 1988). Sometimes we do not catch ourselves making the inference. Thus, Hume makes the observation that “the past experience, on which all our judgments concerning cause and effect depend, may operate on our mind in such an insensible manner as never to be taken notice of, and may even in some measure be unknown to us” (T: 103). Hume writes that a non-swimmer does not stop at the water's edge and “reflects on any past experience, and calls to remembrance instances, that he has seen or

³⁸ Be or be it not also a ‘faculty.’

heard of, in order to discover the effects of water on animal bodies" (T: 103-4). Hume concludes that "custom operates before we have time for reflexion." (cited in Nuyen, 1988: 381).

Now, as Nuyen (1988: 382) argues, the fact that there are cases involving automatic decisions does not mean that reason is not important to belief-formation. Automatic inferences as the one in the swimming case show that some previously consciously drawn inferences have become habitual and thereby automatic. They do *not* show that (reason as) reasoning is superfluous. Indeed reasoning often is effortful for a long time; often we need to do a lot of reasoning before our actions can become habitual:

When I see an object moving in front of my car, I apply the brakes automatically. But this action, or rather reaction, was not automatic when I first started learning how to drive. It took a -lot of inferences from the general behaviour of the car to the effects of applying the brakes. What we seem to do automatically is actually the result of countless past experiences. A grown-up may stop at the water's edge without first making the necessary inferences, but a child who has less experience and has not formed the habit of reasoning under similar circumstances may have to check his or her steps, and make whatever inferences he or she can. The failure to do so has resulted in many drowning accidents. (Nuyen, 1988: 382)

To sum up, reason as reasoning plays an indispensable role in belief formation by presenting the mind with causal inferences from impressions to ideas which we may or may not believe; it is only after having been placed in a certain situation, or set of circumstances, that a belief-feeling arises, or fails to arise. A feeling that (a) is an essential feature of a belief and (b) does not seem to be under the control of reason as reasoning nor –most fundamentally- under the control of the agent. The relationship between reason and belief is one that consists in the fact that reason presents to the mind all the evidence there is. What happens next, when one

feels the force of assent or conviction, is not for reason or the agent to determine. Reason is impotent in this final stage (see Nuyen, 1988: 380-1), but often does and can do a lot of work before that.

10.4.2. Belief and the limits of reason

Here is all the LOGIC I think proper to employ in my reasoning; and perhaps even this was not very necessary, but might have been supply'd by the natural principles of our understanding. (T: 175)

In the previous section, I have explained the role reason as reasoning plays in Hume's account of belief. In this section, I focus on what beliefs have to do with the *limits* of reason.

Hume argues that “the understanding, when it acts alone, and according to its most general principles, entirely subverts itself, and leaves not the lowest degree of evidence in any proposition, either in philosophy or common life” (T: 267-8). It is not entirely clear to what *exactly* this statement means. That is, in the sense that I do not know what Hume intends to say when talking about the understanding's “most general principles”. In any case, as reflected by the quotation just presented, Hume has a clear and strong sympathy for Pyrrhonian scepticism. To understand this claim better, let us recall the following passage:

Matters of fact, which are the second objects of human reason, are not ascertained in the same manner; nor is our evidence of their truth, however great, of a like nature with the foregoing. The contrary of every matter of fact is still possible; because it can never imply a contradiction, and is conceived by the mind with the same facility and distinctness, as if ever so conformable to reality. *That the sun will not rise to-morrow* is no less intelligible a proposition, and implies no more contradiction, than the affirmation, *that it will rise*. We should in vain, therefore, attempt

to demonstrate its falsehood. Were it demonstratively false, it would imply a contradiction, and could never be distinctly conceived by the mind. (EHU: 25-6).

As Hume sees things, the less reflective people amongst us humans, people Hume calls the ‘vulgar’, will *not* normally get to the point of seriously believing that it is a real possibility that the sun does *not* rise tomorrow. Philosophers by contrast will do.³⁹ And as a result they are subject to a sceptical doubt Hume regards as a “malady” and one that “can never be radically cur’d” (T: 218). Despite it being a ‘malady’ (not in the least for Hume himself as a philosopher subject to sceptical doubt), it does not seem to be a normative option for Hume to advocate a philosophical view recommending that we abstain from/give up on refined philosophical reasoning. Hume asks the potentially descriptive and/or normative question: “Shall we, then, establish it for a general maxim, that no refin’d or elaborate reasoning is ever to be receiv’d?” (T: 268). Hume then answers his own question by saying:

Consider well the consequences of such a principle. By this means you cut off entirely all science and philosophy: You proceed upon one singular quality of the imagination, and by a parity of reason must embrace all of them: And you expressly contradict yourself; since this maxim must be built on the preceding reasoning, which will be allow’d to be sufficiently refin’d and metaphysical. What party, then, shall we choose among these difficulties? If we embrace this principle, and condemn all refin’d reasoning, we run into the most manifest absurdities. If we reject it in favour of these reasonings, we subvert entirely the human understanding (T: 268)...For my part, I know not what ought to be done in the present case. I can only observe what is commonly done; which is, that this difficulty is seldom or never thought of; and even where it has once been present to the mind, is quickly forgot, and leaves but a small impression

³⁹ That is, when exercising academic activity. Outside their office, dependent on their personality, they might be no more inclined than ‘the vulgar’ to take this possibility some sort of seriously. Hume would certainly want to say that philosophers no more than “the vulgar” are inclined to *believe* in the possibility that the sun may not rise tomorrow. For Hume, philosophers are just very normal beings subject to the same natural forces as ordinary people.

behind it. Very refin'd reflections have little or no influence upon us; and yet we do not, and cannot establish it for a rule, that they ought not to have any influence ...The *intense* view of these manifold contradictions and imperfections in human reason has so wrought upon me, and heated my brain, that I am ready to reject all belief and reasoning, and can look upon no opinion even as more probable or likely than another...Most fortunately it happens, that since reason is incapable of dispelling these clouds, nature herself suffices to that purpose, and cures me of this philosophical melancholy and delirium, either by relaxing this bent of mind, or by some avocation, and lively impression of my senses, which obliterate all these chimeras. I dine, I play a game of back-gammon, I converse, and am merry with my friends; and when after three or four hours' amusement, I wou'd return to these speculations, they appear so cold, and strain'd, and ridiculous, that I cannot find in my heart to enter into them any farther.

Here then I find myself absolutely and necessarily determin'd to live, and talk, and act like other people in the common affairs of life. But notwithstanding that my natural propensity, and the course of my animal spirits and passions reduce me to this indolent belief in the general maxims of the world, I still feel such remains of my former disposition, that I am ready to throw all my books and papers into the fire, and resolve never more to renounce the pleasures of life for the sake of reasoning and philosophy. For those are my sentiments in that splenetic humour, which governs me at present. I may, nay I must yield to the current of nature, in submitting to my senses and understanding; and in this blind submission I shew most perfectly my sceptical disposition and principles. *But does it follow, that I must strive against the current of nature, which leads me to indolence and pleasure; that I must seclude myself, in some measure, from the commerce and society of men, which is so agreeable; and that I must torture my brain with subtilities and sophistries, at the very time that I cannot satisfy myself concerning the reasonableness of so painful an application, nor have any tolerable prospect of arriving by its means at truth and certainty.* Under what obligation do I lie of making such an abuse of time? And to what end can it serve either for the service of mankind, or for my own private interest? No: If I must be a fool, as all those who reason or believe any thing *certainly* are, my follies shall at least be natural and agreeable. Where I strive against my inclination, I shall have a good reason for my resistance; and will no more be led a wandering into such dreary solitudes, and rough passages, as I have hitherto met with. (T: 268-70 cfT: 218; EHU: 161-2; DNR: 131-2)

So what Hume argues for here is that a good philosopher who exercises his imagination and understanding properly should become utterly confused, not knowing anymore what to believe.⁴⁰ However, according to Hume, even good philosophers will continue to believe, say, that the sun rises tomorrow (just as it did today). The epistemological force from Pyrrhonian skepticism will not get a foothold in the real world. The experiential evidence the skeptical philosopher has is too strong for that in order to prevent him from that belief. It is here that according to Hume, nature's innate constitution plays a role (see also T: 162, 187; EHU: 41-2). The philosophical confusion does not last as a violent phenomenon. It does not because we have plenty of counterevidence that together with some help of nature prevents us from being dominated by philosophical skepticism as a result of subjecting ourselves to forms of Pyrrhonian skepticism. Nature steps in and brings us back⁴¹ to a state that enables us to function as practical agents in a world about which the experienced philosopher knows that we cannot be sure about what appears to be. This is apparently a matter of physical necessity. (see T: 183, Lynch, 1996)

For Hume, nature's role as something that makes up for the situation where reason gives us 'rightly' the feeling to be totally lost is a *causal* one. In the italicized bit in the longer quotation above however, Hume seems to ask a normative question when he asks what stance to take towards the doubt arising from Pyrrhonism. And his answer seems to be that, while it should influence our lives, this extreme form of scepticism should not take over our lives. So what should we make of this normative claim that is apparently present in Hume's work? Considering Hume's philosophical project as a whole, it seems awkward to assume that for Hume there could be a justifying reason for continuing to believe that does *not* have its roots in some sort of naturalistic explanation. Possibly Hume would want to say that it is a

⁴⁰ My hunch is that we should say that for Hume Pyrrhonism is epistemologically true.

⁴¹ It seems also accurate to say that "nature usually takes care of the problem by not letting it arise" (Neto, 1991: 44).

fortunate coincidence that nature helps us out and that we should do everything not to interfere. Perhaps the (apparent) combination of a normative aspect and a causal one in Hume's work about Pyrrhonian skepticism is best understood as saying that as a result of natural forces we continue to go on trusting our judgments any-way and we have to live our life in some sort of doable way, so we have an obligation not to get in the grip of our philosophical reasoning that is allowed to take place in the closet and in a mitigated form should be part of the psychological dispositions of every human creature. We should not let our lives be ruined by extreme forms of skepticism, but Pyrrhonian skepticism as a modest form that should be adopted by all human beings so that our beliefs have an appropriate degree of warrant. Textual evidence for Hume's endorsement of mitigated Pyrrhonism –or at least textual evidence for 'normative endorsement about careful reasoning and belief-formation'- comes in the form of Hume's claim that: "tis frequently found, that observation is contrary to another, and that causes and effects follow not in the same order, of which we have had experience, we are oblig'd to vary our reasoning on account of this uncertainty, and take into consideration the contrariety of events" (T: 131). And in *EHU* he mentions that "a wise man (should proportion) his belief to the evidence (EHU: 110). Probably we can say that for Hume as a cognitive psychologist concerned with describing our psychology as well as for Hume the epistemologist concerned with the normative problem of warrant, the warrant for beliefs results from Pyrrhonian scepticism checked by the influence of our more permanent and constant natural belief-forming mechanisms (see Falkenstein, 1997: 31-2; see also Neto, 1991, esp. 48).

Above I have focused on scepticism with regard to matters of fact. And in § 4 and 6 I have argued that probably for Hume there is knowledge as *certainty* in the mathematical area. Interestingly however and relevant to our present discussion is the fact that there is some idea about distrust of reasoning faculties in Hume's work about mathematics. To understand this idea, let us recall that for Hume algebra and arithmetic are sciences that "preserve a perfect exactness and certainty" (T: 71). In these sciences, we can compare ideas

and “determine their relations without any possibility of error” (T: 71). As Lynch (1996) argues, this means that Hume takes it that when we perform arithmetic - during which we are surely reasoning - we can and do form judgments in which we are not only psychologically, but most probably also rightly confident. Let us also recall a quotation I presented earlier:

In all demonstrative sciences the rules are certain and infallible; but when we apply them, our fallible and uncertain faculties are very apt to depart from them, and fall into error. (T: 180)

The argument about sceptical distrust Hume has when it comes to the area of mathematics is summarized by Lynch in the following way:

In performing any set of calculations, no matter how simple, we are susceptible to error. This we know from past experience. We will, of course, be more confident in our conclusions if we recheck our calculations, or allow others to check them for us. But while these methods can increase the subjective probability that we are right, they can never increase it all the way to 1. For the possibility that we have made a mistake always remains. Therefore, we should never be completely certain of any of our beliefs, even of those concerning rudimentary mathematics, and hence “all knowledge degenerates into probability. (Lynch, 1996: 89-90)

How is it that Hume is providing such an argument while probably arguing for the claim that there *is* knowledge (in the form of *scientia*) in mathematics? There is reason to think that this is due to a particular sympathy Hume shared with the ancient sceptics. As Julia Annas and Jonathan Barnes argues:

The ancient sceptics did not [just] attack knowledge: they attacked belief. They argued that, under sceptical pressure, our beliefs turn out to be groundless and that we have no more reason to believe than to disbelieve. (Annas & Barnes⁴², cited in Lynch, 1996 : 95)

According to Lynch, Hume, like the ancient sceptics, was not content with showing that scepticism undermines knowledge while leaving justified belief intact (Lynch, 1996: 95).

10.5. Beliefs and representation

Hume hardly uses the term ‘belief’ in his *moral* work. In Book III of the *Treatise* (about morality) there are only eight passages that contain the term. In *EPM*, Hume uses the term only once. And in none of those passages, the term seems to carry any special importance. That seems something noteworthy, especially also given the way some meta-ethicists interpret Hume.

Recalling and elaborating on what I have said earlier, let me briefly mention a few things about beliefs and ‘representation’ in Hume’s work. I would like to do so because there is a highly dominant belief concept in the meta-ethical debate at large, one that holds that a moral belief is a mental state that *reflects the way the world is morally*. This concept is there and dominates, while many meta-ethicists ascribe a mistaken belief conception to Hume. When Hume is interpreted in a mistaken way, this mistake most of the time comes down to a mistake about Hume’s views as to the motivational power of beliefs. Having said that, the current meta-ethical debate gives the impression that there is quite a bit of room for improvement in terms of a correct interpretation of Hume’s ideas on belief. Given that Hume seems rather badly understood, given the nature of the dominant conception of belief in meta-ethics and given the fact that some meta-ethicists explicitly argue that according to

⁴² Annas, J. & Barnes, J. (1985). *The Modes of Scepticism*. Cambridge:Cambridge University Press.

Hume beliefs represent moral aspects of the world, it seems a good idea to put forward some explicit words on Hume and representation, even though I have already dealt with this topic rather extensively.⁴³

It is only very occasionally Hume talk fairly explicitly about ‘representation’. In his essay ‘Of the standard of taste’ and in regard to all ‘determinations of the understanding’ Hume argues: “But all determinations of the understanding are not right; because they have a reference to something beyond themselves, to wit, real matter of fact; and are not always conformable to that standard (ST: 7). Hume does say things like: “I say, then, that belief is nothing but a more vivid, lively, forcible, firm, steady conception of an *object*” (EHU: 49) my italics). But Hume’s attention seems to go to the manner of conception and the phenomenological experience rather than to the object. And if we then want to start talking about representation, we should keep in mind that this notion for Hume can never mean that we our lively belief ideas represent objects that are really there in the world. Hume’s philosophy is fully focused on the items of the mind called impressions and ideas with no commitment as to the cause of sensory impressions. Furthermore, we should keep in mind that for Hume the verb ‘represent’ and the noun ‘representation’ –as explained- typically are related to the theoretical thoughts that ‘ideas’ (including the lively ‘ideas’ called beliefs) are *copies of impressions*.

10.6. Beliefs and motivational power

As Nuyen (1988) points out, the power of association and transfer that so often takes place in our minds, is the work of nature which enables us to anticipate dangers and avoid them, without actually seeing them coming. Without it, "we should every moment of our lives be

⁴³ One claim that may come as a special surprise to meta-ethicists is Hume’s claim that: “*all our reasonings concerning causes and effects are deriv’d from nothing but custom; and that belief is more properly an act of the sensitive, than of the cogitative part of our natures.*” (T: 183; italics original)

subject to the greatest calamities" (T: 119, cited in Nuyen 1988). So having impressions and ideas with some sort of high degree of conviction component serves a real purpose. But if we were to experience only impressions and/or ideas with a high degree of conviction coefficient, we would "never enjoy a moment's peace and tranquility" (see T: 119). Having pretty liveliness ideas prevents us from having a (useless) restless psychology.

Out of an independent interest to present Hume's interesting and valuable philosophical ideas, but also very much in anticipation and support of my claim about there being a misreading in contemporary meta-ethics when it comes to Hume's concept of belief, let me say a few explicit words about what Hume seems to think about beliefs and motivation. Let us start with two quotations from secondary literature followed by a primary one from Hume himself.

We can imagine that the house is on fire. Being only in the imagination, this idea is, as Hume would say, cold and lifeless. By contrast, if we have impressions of smoke, i.e. seeing and smelling smoke, *the same idea* will arise in the mind through inference, but it is an idea with force and vividness, qualities that come from the idea's being associated with the sense impressions of smoke. If we then *believe* that the house is on fire, we will be moved to take actions *as if* we see that the house is actually on fire. (Nuyen, 1988: 376)

[A] careful and charitable interpretation [there is some significant variety in the descriptive list we have seen; see footnote 15 above] shows that, for Hume, a belief is a perception that has a certain feeling to the mind, which is the same as saying that it is a perception that is conceived in a certain manner. Furthermore, beliefs are those perceptions that most affect the will, and they are able to do this either by virtue of being impressions themselves or by virtue of their relations to the impressions or memories that give rise to them. (Gorman, 1993: 99)

But its true and proper name is belief, which is a term that everyone sufficiently understands in common life. And in philosophy we can go no farther than assert that it is something felt by the mind, which distinguishes the ideas of the judgment from the fictions of the imagination. It gives

them more force and influence; makes them appear of greater importance; infixes them in the mind; and renders them the governing principles of all our actions. (T: 629)

What does this mean in the light of the contemporary neo-Humean conviction that beliefs are motivationally inert? I think the following answers apply. First, there is *no* textual evidence that suggests that according to Hume beliefs *typically*, let alone always and necessarily, are motivationally inert. Secondly, there seems to be positive evidence that beliefs are motivationally powerful because of their intense and special phenomenological nature. And some sort of movement of the will may be implied for every belief. Before I move on, let us have a look at a remark from Michael Smith:

According to the standard picture of *Hume (1888)* – there are two main kinds of psychological state. On the one hand there are beliefs, states that purport to represent the way the world is. Since our beliefs purports to represent the world, they are assessable in terms of truth and falsehood, depending on whether or not they succeed in representing the world to be the way it really is. And on the other hand there are desires, states that represent how the world is to be. Desires are unlike beliefs in that they do not even purport to represent the way the world is. They are therefore not assessable in terms of truth and falsehood. Hume concludes that belief and desire are therefore distinct existences: that is, we can always pull belief and desire apart, at least modally. For any belief and desire pair we imagine, we can always imagine someone having the desire but lacking the belief, and vice versa. (Smith, 1994: 7)

Smith's belief concept here can be taken as representative of a dominant conception in meta-ethics. I think we should be more careful as to it being representative of a misreading of Hume, but that is a complex story we come to in the next chapter. Knowing what we know about Hume and taking Smith's belief concept as representative of a dominant meta-ethical concept, we (thirdly) can conclude that there is a world of difference between Hume's concept of belief and the concept of belief operative in neo-Humeanism.

Fourthly, approaching the concept of belief from as much of a neutral perspective as possible, while keeping the meta-ethical conception in the back of one's mind, it seems plausible to hold the view that at least *some* beliefs do not necessarily motivate, e.g. the belief that Paris is the capital of France. Fifthly, assuming for the sake of argument that someone has a bit of knowledge in his mind when it comes to Paris being the capital of France, unless such bits of knowledge have this special intense phenomenological nature that characterized Hume's beliefs, in likelihood we must conclude that what contemporary philosophers call 'beliefs' are not beliefs for Hume. Positively speaking we may have to conclude that for Hume there are such bits of knowledge called 'judgements' (recall that Hume does not have a very developed account of 'judgements', although there are a clear number of occasions where he uses the word- also occasions that point more to a mental *product* than a mental *act*). Or else or additionally, occasions of 'assent'. If someone has a bit of knowledge in his mind with the special experiential nature of belief, it seems plausible to assume that some sort of motivation will go with it. Suppose 'Paris is the capital of France' is a belief held by an Asian man who has never been to Europe and has just heard this from a trustworthy friend who never wants to go back to Paris, because it does not seem to be a pleasant city at all and despite that somehow gets to the point of having a lively conception of Paris in his mind with the special belief feeling Hume thinks we cannot really describe. The Asian man's will may then, despite the experience of his friend, being influenced by his belief, say by going to the library to find out more about what the city is like or by saving a particular amount of money every month that will finally enable him to buy a ticket to Paris.

11. Hume on human nature

To begin with vice and virtue, which are the most obvious causes of these passions; 'twou'd be entirely foreign to my present purpose to enter upon the controversy, which of late years has so much excited the curiosity of the publick, whether these moral distinctions be founded on

natural and original principles, or arise from interest and education. The examination of this I reserve for the following book; and in the mean time shall endeavour to show, that my system maintains its ground upon either of these hypotheses; which will be a strong proof of its solidity.
⁴⁴ (T: 295; italics original)

One way, a paradigmatic one, in which Hume is concerned with ‘human nature’ is by him being a cognitive psychologist: someone interested in describing our actual mental states and the cognitive mechanisms that bring them about. Be that the dominant way in which Hume is concerned with ‘human nature’, Hume also puts forward views about our innate constitution and our environmental contribution to who we are and what we do as practical agents. In this section I present ‘innate constitution’ and ‘environment’ as two other angles of Hume’s work as a philosopher interested in human nature.

11.1. Hume on human nature: innate constitution

It will not come as a surprise to many that the empiricist Hume believes that there are no such things as innate ideas. In book I of the *Treatise*, Hume clearly argues against their existence (see T: 158). Be that as it is, Hume’s work is not free from claims about ‘something given to us when we come to enter the world; something given to us by the world by the author of nature’. In *EPM* for example, Hume argues: “[N]o passion, when well represented, can be entirely indifferent to us; because there is none, of which every man has not, within him, at least the seeds and first principles” (EPM: 222). And in book II of the *Treatise*, Hume argues:

⁴⁴ In his essay ‘Of the middle station of life’, Hume says: “Nature must afford the richest Genius that comes from her Hands; Education and Example must cultivate it from the earliest Infancy; And Industry must concur to carry it to any Degree of Perfection” (MS: 6).

Beside good and evil, or in other words, pain and pleasure, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites. These passions, properly speaking, produce good and evil, and proceed not from them, like the other affections. (T: 439)⁴⁵

And Hume also says:

Now 'tis certain that, there are certain calm desires and tendencies, which, tho' they be real passions, produce little emotion in the mind, and are more known by their effects, than by the immediate feeling or sensation. These desires are of two kinds; either certain instincts originally implanted in our natures, such a benevolence and resentment, the love of life and kindness to children; *or* the general appetite to good, and aversion to evil, consider'd merely as such. (T: 417; my italics)

As to the passage directly above, in his (2007), Peter Kail argues that the emboldened 'or' above is an exclusive 'or'. If he is right about that, then this would mean that the 'or' functions to express the idea that the general appetites *but not* the instincts are acquired from encounters with particular instances of good and evil. This may very well be the correct interpretation. I find it plausible. Having said that, I am not entirely sure that this is the correct interpretation given that Hume somewhat later in his *Treatise* argues that: "The mind, by an original instinct, tends to unite itself with the good and to avoid the evil" (T: 438). The alternative to Kail's proposal would be to read Hume as saying that both the instincts and the general appetites mentioned are calm passions originally implanted in our natures. In this case, the 'or' would refer to a distinction between *specific* instincts and *general* appetites.

⁴⁵ On page 230 of *EPM*, Hume talks about the "natural sentiment of benevolence". I take it that this entails a claim about benevolence being given to us by the author of nature.

Note that at certain places in his work Hume relates views about affect to matters having to do with freedom. In his *Treatise* for example Hume states: “[’Tis] certain we can naturally no more change our own sentiments, than the motions of the heavens” (T: 517). And in his essay ‘The Sceptic’ Hume argues: The fabric and constitution of our mind no more depends on our choice, than that of our body (SC: 28).⁴⁶ If and when Hume talks about a relationship between passions and freedom, sometimes, if not always, these views are views about our innate affective constitution. Hume thinks that the mind is plastic (see SC: 31; *cf* CL: 1; *cf*EHU: 105) and that we can exercise some volitional influence on our passions, though he also seems to think that on a fundamental level the influence of passions on our minds is not within our volitional control.

Assuming for the sake of the argument that there are occasions where Hume talks about the relationship between affect and freedom where no view, at least no explicit view, is expressed about our natural constitution, then it still seems to be the case that also in these occasions Hume argues for some sort of lack of power for volitional control over emotions.

11.2. Hume on culture and education

That many of the natural virtues have this tendency to the good of society, no one can doubt of. Meekness, beneficence, charity, generosity, clemency, moderation, equity, bear the greatest figure among the moral qualities, and are commonly denominated the *social* virtues, to mark their tendency to the good of society. This goes so far, that some philosophers have represented all moral distinctions as the effect of artifice and education, when skilful politicians endeavour’d to restrain the turbulent passions of men, and make them operate to the public good, by the

⁴⁶ Paul Bloom and colleagues at the university of Yale have taken up the project of investigating the innate roots of morality by investigating prosocial behaviour and senses of justice in babies (Bloom 2013, 2010a, 2010b; Hamlyn, Wynn & Bloom, (2007); Hamlyn, Wynn & Bloom, (2010). For a quick introduction see (2010a). This is an article by Bloom in the New York Times The article includes a link to a video that explains in a nutshell the research from Bloom and his colleagues.

notions of honour and shame. This system, however, is not consistent with experience. For, *first*, there are other virtues and vices beside those which have this tendency to the public advantage and loss. *Secondly*, had not men a natural sentiment of approbation and blame, it cou'd never be excited by politicians; nor wou'd the words *laudable* and *praise-worthy*, *blameable* and *odious*, be any more intelligible, than if they were a language perfectly unknown to us, as we have already observ'd. But tho' this system be erroneous, it may teach us, that moral distinctions arise, in a great measure, from the tendency of qualities and characters to the interest of society, and that 'tis our concern for that interest, which makes us approve or disapprove of them. Now we have no such extensive concern for society but from sympathy; and consequently 'tis that principle, which takes us so far out of ourselves, as to give us the same pleasure or uneasiness in the characters of others which are useful or pernicious to society as if they had a tendency to our own advantage or loss. (T: 578-9)

In his *Treatise* Hume claims: "I am persuaded, that upon examination we shall find more than one half of those opinions, that prevail among mankind, to be owing to education" (T: 117). Hume ascribes an important role to education too when it comes to moral discriminations. However, just as the passage at the top of this section indicates, from the passage below it becomes clear that when it comes to moral discriminations an appeal to culture cannot do the whole trick.

From the apparent usefulness of the social virtues, it has readily been inferred by sceptics, both ancient and modern, that all moral distinctions arise from education, and were, at first, invented, and afterwards encouraged, by the art of politicians, in order to render men tractable, and subdue their natural ferocity and selfishness, which incapacitated them for society. This principle, indeed, of precept and education, must so far be owned to have a powerful influence, that it may frequently increase or diminish, beyond their natural standard, the sentiments of approbation or dislike; and may even, in particular instances, create, without any natural principle, a new sentiment of this kind; as is evident in all superstitious practices and observances: But that *all*

moral affection or dislike arises from this origin, will never surely be allowed by any judicious enquirer...

The social virtues must, therefore, be allowed to have a natural beauty and amiableness, which, at first, antecedent to all precept or education, recommends them to the esteem of uninstructed mankind, and engages their affections. And as the public utility of these virtues is the chief circumstance, whence they derive their merit, it follows, that the end, which they have a tendency to promote, must be some way agreeable to us, and take hold of some natural affection. It must please, either from considerations of self-interest, or from more generous motives and regards. (EPM: 214-5)

So what we see (at least indirectly) in this passage is that according to Hume our innate nature matters too in regard to our moral thoughts and deeds. Indeed, Hume sees it as something crucial and fundamental without which no cultural influence can operate. Be that as it is, for Hume justice is an “artificial” virtue meaning that it arises out of the efforts of human beings. Hume argues:

Unless we will allow that Nature has established a *Sophistry*, and rendered it necessary and unavoidable; we must allow that the Sense of Justice and Injustice is not derived from Nature, but arises artificially, tho' necessarily, from Education and human Conventions. Here is a Proposition which I think may be regarded as certain, *That it is only from the Selfishness and confined Generosity of Men, along with the scanty Provision Nature has made for his Wants, that Justice derives its Origin.* These Impressions, which give Rise to this Sense of Justice, are not natural to the Mind of Man, but arise from Artifice and human Conventions. Without such a Convention, no one would ever have dreamed that there was such a Virtue as Justice, or have been induced to conform his Actions to it. (LG: 10; cT: 483)

Let me make one final comment in regard to Hume, education and the practice of morality. As I have argued, the ‘common point of view’ plays an important role in Hume’s work. His thought seems to be that human beings do not do well when they favour their untutored

impulses over their tutored ones. And education for Hume is a tool to make sure that people get beyond their primitive impulses of passions that pull one away from an impartial judgement (*cf*RP: 11).⁴⁷

⁴⁷Let me finally say a few words about Hume's thoughts about malicious human beings. Some of Hume's ideas about malicious human beings map relatively well on contemporary scientific ideas about psychopathy. Consider the following passages from Hume's work:

“Let us suppose such a person ever so selfish; let private interest have engrossed every so much his attention; yet in instances, where that is not concerned, he must unavoidably feel *some* propensity to the good of mankind, and make it an object of choice, if everything else be equal. Would any man, who is walking along, tread as willingly on another's gouty toes, whom he has no quarrel with, as on the hard flint and pavement?” (EPM: 226)

“All mankind so far resemble the good principle, that, where interest or revenge or envy perverts not our disposition, we are always inclined, from our natural philanthropy, to give the preference to the happiness of society, and consequently to virtue above its opposite. Absolute, unprovoked, disinterested malice has never perhaps place in any human breast; or if it had, must there pervert all the sentiments of morals, as well as the feelings of humanity.” (EPM: 227)

“[W]here one is born of so perverse a frame of mind, of so callous and insensible a disposition, as to have no relish for virtue or humanity, no sympathy for his fellow creatures, no desire of esteem and applause, such a one must be allowed entirely incurable; nor is there any remedy in philosophy. [...] For my part, I know not how I should address myself to such a one, or by what arguments I should endeavour to reform him. Should I tell him of the inward satisfaction which results from laudable and humane actions, the delicate pleasure of disinterested love and friendship, the lasting enjoyment of a good name and an established character, he might still reply, that these were, perhaps, pleasures to such as were susceptible of them; but that, for his part, he finds himself of a quite different turn and disposition. I must repeat it, my philosophy affords no remedy in such a case; nor could I do anything but lament this person's unhappy conditions. But then I ask, if any other philosophy can afford a remedy.” (SC: 29)

My overall impression is that Hume was in doubt as to whether there actually exist, have ever existed and will ever exist such beings as described by him above. Explicit textual evidence for the view that Hume was in doubt about their actual existence comes from *EPM*, page 227 where Hume argues: “Absolute, unprovoked, disinterested malice has never *perhaps* place in any human breast; or *if it had*, must there pervert all the sentiments of morals, as well as the feelings of humanity” (my italics).

Chapter 3. On the risk of misreading Hume

1. Introduction

In the previous chapter, I discussed Hume's views on reason, belief and morality in detail. In this chapter I focus on the relationship between misreading Hume (especially on belief) and the theories of contemporary meta-ethical authors. I do this to explore unfruitful consequences of misinterpreting Hume for contemporary meta-ethical theories. While exploring the relationship in a variety of ways, I focus in particular on those theories and authors referred to as 'neo-Humean' and 'anti-Humean'. I proceed as follows.

In § 2, I briefly summarize Hume's views on 'reason' and 'belief' as I have presented them in the previous chapter and in such a way as is relevant to this chapter. In § 3, I focus on the nature of '*meta-ethical* neo-Humeanism'. (NB: in this chapter I focus exclusively on neo-Humeanism as it appears in *meta-ethics*, i.e. I leave aside neo-Humean theories as they appear in the philosophy of action and other areas.¹) In § 4, I explain the following three relationships. First, the connection between the *theoretical nature* of neo-Humeanism and *true* claims about the philosophy of Hume. Secondly, the relationship between the *theoretical nature* of neo-Humeanism and any claims about the philosophy of Hume, be they accurate or not. Thirdly, the relationship between *published claims of authors* defending or challenging meta-ethical neo-Humeanism *and these authors' claims about Hume being the historical father of neo-Humeanism*.

In § 5, I focus on drawbacks and intellectual hindrances that might currently be caused by misinterpreting Hume's views. In § 5.1, I focus on the problem of insufficient reflection in meta-ethics, i.e. reflection on concepts such as 'belief', 'emotion', 'sentiment', 'feeling' and 'reason'. In § 5.2, I focus on 'the interconnectivity problem'. That is, I approach the issue of a misinterpretation of Hume's views on 'belief' from the perspective of the

¹ On the relationship between neo-Humeanism in meta-ethics and neo-Humeanism in other areas see my footnote at the very end of this chapter.

interconnectivity of the three core meta-ethical debates: the one about the nature of and status of moral judgements (the debate about the correctness of cognitivism and non-cognitivism); the debate about the nature and status of moral objectivity; and the debate about moral motivation. In this section I offer hypotheses about possible far-reaching consequences for the debate about cognitivism and non-cognitivism and realism and anti-realism once it is (mistakenly) assumed that Hume held that beliefs are motivationally inert. In § 5.3, I focus on what perhaps is the worst possible intellectual upshot of misinterpreting Hume. I explore the possibility that meta-ethical authors X and Y, can be criticised on their own terms by arguments that derive from Hume. That is, in the sense that Hume's work surpasses the work of X and Y in theoretical substance when taking as a starting point certain intellectual aims X and Y have (e.g., providing a theory about moral motivation); aims Hume has written about in the form of substantive views. This phenomenon seems to exist. If it exists, then there seems to be reason to be concerned. And if and when the phenomenon exists *because relevant parts of Hume's work are misread*, then the strong recommendation to study these parts carefully before continuing contemporary discussion follows very quickly.

2. Hume on reason and belief: recalling some essential thoughts

Let me recall very briefly some important conclusions from the last chapter. For our purposes of this chapter, two sentences will suffice about Hume's view of '*reason*'. First, for Hume *reason alone* cannot make moral distinctions. Secondly, according to Hume, reason alone "can never be a motive to any action of the will" (T: 413).

I now summarize Hume's views on 'belief'. First in regard to the topic of motivation. In order to be accurate, it needs to be said that Hume does not talk very often *explicitly* about the relationship between *beliefs and motivation*. (NB: as to the latter term, Hume does

not use the term ‘motivation’ at all; we would need the term ‘will’ or else ‘action’ for Hume.) Whilst Hume rarely expresses views about the relation between belief and what we may call ‘motivation’, those that he does express imply strongly that beliefs have the intrinsic capacity to motivate. Let us for example recall the following claim:

But its true and proper name is belief, which is a term that everyone sufficiently understands in common life. And in philosophy we can go no farther than assert that it is something felt by the mind, which distinguishes the ideas of the judgment from the fictions of the imagination. It gives them more force and influence; makes them appear of greater importance; infixes them in the mind; and renders them the governing principles of all our actions. (T: 629)

I take it that overall, i.e. taking into account all Hume’s less and more explicit claims on ‘beliefs’, ‘impressions’ as well as the ‘will’, there is strong textual evidence for the view that for Hume at least some beliefs in contrast to what Hume calls ‘reason’ *can and do* motivate on their own. They do so because all beliefs have a phenomenological nature that has a particular impact on the mind. That is, for Hume beliefs are highly lively ideas. They are ideas with a strong ‘conviction-coefficient’ (recall Laird, 1993). Because of the special phenomenological nature that comes with this conviction coefficient –a nature that includes a special *feeling*- for Hume beliefs can and do regularly influence our will. Not all beliefs will influence our will, but in the absence of counter-acting beliefs they clearly have the possibility to do so (without any help from an additional passion) because of the phenomenological nature and impact of a belief on our minds that resembles the nature and impact of an impression.

Anticipating upcoming remarks about neo-Humeanism while considering Hume’s writings on beliefs comprehensively, I take it that the following conclusion should be drawn. In contrast to what some meta-ethicists think, Hume neither explicitly nor implicitly argued that ‘beliefs alone can never be a motive to any action of the will’. While Hume wanted to

say that not all beliefs affect the will, there is *no* textual evidence that supports the view that for Hume beliefs are *typically* or even *necessarily* motivationally inert.

Above, I have talked about *motivation*. Given meta-ethicists' *general*, i.e. Hume-independent and also '*Hume-related*' association of 'beliefs' with '*representation*' (see §§ 3, 5.1, 5.2 below), I now briefly summarize what I have said in chapter 2 about the relationship between beliefs and representation in Hume's philosophy. Most probably Hume's beliefs can be said to 'represent', i.e. to represent objects like trees and chairs. The following should be kept in mind however, especially in order not to associate Hume with correspondence theories about an extra-mental world. In regard to external objects, Hume's notion of 'representation' is *not* one that reflects a theoretical commitment to the existence of an extra-mental world. That does not mean that we can draw the theoretical conclusion that for Hume there are no extra-mental objects. We cannot. It only means that Hume (highly actively and intentionally) *avoids a theoretical commitment* to their existence. We cannot know more than our senses deliver us and therefore we need to be neutral on whether there actually are such (extra-mental) things as trees and chairs.

In regard to *moral* matters, the following remarks should be made about 'belief' and 'representation'. First, Hume uses the term 'belief' only once in *EPM* and in no more than eight passages in the part of the *Treatise* – book III – that is about morality. Secondly, the fact that *although* Hume seems to be a *truth-cognitivist* about morality and perhaps can even be referred to as a *realist* about morality (note esp. also Hume's own remark on this in *EPM*: 169-70), he might have been (highly) unsympathetic to something like the claim that '*true moral judgements represent moral aspects of the world*,' a dominant claim in meta-ethics (i.e. in the discussions about cognitivism and non-cognitivism, and between realism and anti-realism). This, both (A) because of his actively intended opposition to those of his contemporaries who claim that there are eternal and timeless truths of 'reason' and (B) because of his thoughts about the contribution of passions to moral discriminations. Probably, Hume would not have wanted to accept the dominant meta-ethical view that

'beliefs are states that represents moral aspects of the world'. Hume certainly would not accept this view as a defining view of 'belief'.

3. Meta-ethical neo-Humeanism

I now proceed by giving a summary of the position 'neo-Humeanism' as it exists in meta-ethics. In his 2003, Russ Shafer-Landau argues that neo-Humeanism² is widely accepted amongst philosophers (Shafer-Landau, 2003: 161; *cf* Rosati, 2006)³. Connie Rosati notes that "[a]ccording to the neo-Humean view, belief is insufficient for motivation, which always requires, in addition to belief, the presence of a desire or conative state. Moral motivation thus cannot arise from moral belief alone but must depend as well upon a preexisting desire or other conative or intrinsically motivating state" (Rosati, 2006).

Two ideas that permeate contemporary neo-Humeanism are the *neo-Humean split* and the *Inertia of belief thesis* (for the latter term, see Cohon, 2008). The former consists in a claim about a modal division between beliefs and desires. The latter consist in the claim that beliefs cannot motivate on their own. In terms of the *neo-Humean split*, typically if not standardly, neo-Humeanism in meta-ethics understands the strict distinction between beliefs and desires as a *functional* distinction: beliefs and desires are mental states with very different functional roles. The functional role of desires is to motivate. Here they stand in opposition to a negative feature neo-Humeanism attaches to beliefs: beliefs do not motivate.⁴

² Russ Shafer-Landau leaves out the prefix 'neo' and just calls the position I am talking about 'Humeanism'.

³ From informal correspondence it has become clear that this is a claim about meta-ethical philosophers rather than philosophers in general (across the board of disciplines) (Jansen/Shafer-Landau 30/10/2013).

⁴ The *neo-Humean split* reflects either the claim (i) that there are *no variations* in the 'belief-states' and the 'desire-states' *at all* or (ii) that there are *some variations* in either or both of the two types of states (most probably the 'desire-states'). If (ii), these minor differences are considered to be insignificant from the point of view of the alleged functional role of the two types of states.

The belief part of neo-Humeanism's strict distinction between beliefs and desires typically has one more essential feature in meta-ethics. This feature is that beliefs are representational states.

All things considered, the neo-Humeanist idea as it appears in meta-ethics is the following one: beliefs represent moral aspects of the world. That is their functional role. As mental states that represent moral aspects of the world they are motivationally inert. Desires by contrast are intrinsically motivational states and stand in opposition to the functional role of beliefs in that desires are not representational states.

When it comes to misinterpreting Hume, from a perspective of textual exegesis, the desire part of the *neo-Humean Split* is much less problematic than the belief part of the *neo-Humean Split*. Much less problematic, but *not entirely free from problems*. The desire part is *somewhat problematic* in the sense that for neo-Humeanists typically if not standardly it carries a high dosage of what is known as the 'direction of fit metaphor' (just as the belief part does in neo-Humeanism). This metaphor is a specific way to express the functional distinction between beliefs and desires that characterizes neo-Humeanism. It is one employed by many meta-ethicists.⁵ The idea of the directions of fit metaphor is to explain the nature of beliefs and desires, and the difference between them, in terms of their relationship with the world. To use John Milliken's words: "Put briefly and crudely, the thought is this: we want our beliefs to fit the world and the world to fit our desires" (Milliken, 2008). Michael Platt adds: "beliefs should be changed to fit with the world, not vice versa, while the world...should be changed to fit with our desires, not vice versa" (Platt, 1979: 257).

⁵ It is common to attribute the notion of directions of fit to Elizabeth Anscombe (1957, section 32), although it seems to have been John Searle who coined the term 'directions of fit' (1979; cf 1983). Searle (1979) applies the distinction to effect a taxonomy of speech acts rather than mental states. For this application, Searle uses the term 'words-to-world-direction' (for statements, predictions, etc.) and the term 'world-to-words-direction' (commands, promises, etc.). In his (1983) Searle uses the terms 'world-to-mind' and 'mind-to-world' to mark a version of the direction of fit metaphor that is concerned with mental states rather than speech acts (see Humberstone, 1992).

Returning to (any actual or possible claims about) the relationship between the desire part of the *neo-Humean split* and the work of the historical Hume on desire, the desire part of the *neo-Humean split* is problematic from an exegetical perspective in the sense that the direction of fit metaphor (or something sufficiently close to it) cannot be found in Hume's work. As I have explained in the previous chapter, generally, it has to be concluded that Hume is not very precise and clear about what a desire is. Granting that, the following remark can nevertheless be made. The direction of fit metaphor –indeed seemingly any functionalist account of desire- does not map in any sufficient sense onto what Hume argues for in relation to 'desires'. That obviously does not mean that for Hume desires do not motivate. It means that Hume does not think about them in functionalist terms.

4. On the relationship between neo-Humeanism and Hume

The relationship between neo-Humeanism and Hume is one that needs to be approached carefully. The first thing one can say about this relationship is that the neo-Humean doctrine in theory in all likelihood does *not* have a *necessary* relationship with some of the ideas Hume put forward. That is, neo-Humeanism as a contemporary position is *not* theoretically dependent on the historical Hume. As a contemporary position neo-Humeanism has a nature that does not include a necessary appeal to Hume's views *as he held them* (as would have to be judged by the standard of careful textual exegesis).

Secondly, in all likelihood neo-Humeanism is not theoretically dependent on any theoretical claim about what (with or without careful exegesis) is *taken to be* Hume's view, be such a claim reflective of Hume's view or not.

Thirdly, currently not every meta-ethicist who subscribes to neo-Humeanism subscribes to the claim that Hume is the intellectual father of neo-Humeanism in some important way(s) (nor does every meta-ethicist who wishes to challenge the neo-Humean

doctrine). This claim I spell out in the remainder of this section. I start with the following tentative belief/observation: the number of meta-ethicists who make an explicit appeal to Hume in their contemporary work is rather small. Having said that, there clearly are people who do so. Below I provide a list of such people. This list may be fairly or fully exhaustive or else may have room for being longer than it is (implying that I have failed to spot passages of authors that could be added to the list). Here then is the list:

According to the standard picture of *Hume (1888)* – there are two main kinds of psychological state. On the one hand there are beliefs, states that purport to represent the way the world is. Since our beliefs purport to represent the world, they are assessable in terms of truth and falsehood, depending on whether or not they succeed in representing the world to be the way it really is. And on the other hand there are desires, states that represent how the world is to be. Desires are unlike beliefs in that they do not even purport to represent the way the world is. They are therefore not assessable in terms of truth and falsehood. Hume concludes that belief and desire are therefore distinct existences: that is, we can always pull belief and desire apart, at least modally. For any belief and desire pair we imagine, we can always imagine someone having the desire but lacking the belief, and vice versa. (Smith, 1994: 7)⁶

Yet according to the most common view of motivation, taken from David Hume, our beliefs alone are unable to motivate. Desires are the ultimate source of motivation; beliefs play only the limited role of informing us about how best to satisfy the desires we already have. (Shafer-Landau 2003: 4)

If an agent is to be moved to action, then two requirements have to be fulfilled: first, the agent must possess beliefs about the way things actually are, about the actions possible given the way things are, and about the likely effects of those actions on how things are; and, second, the agent

⁶ Compare Smith (1991: 400): “According to the standard picture of human psychology- a picture we owe to David Hume-there are two main kinds of psychological state. On the one hand there are beliefs, states that purport to represent the way the world is. Since our beliefs purport to represent the world, they are subject to rational criticism: specifically they are assessable in terms of truth and falsehood. Desires are unlike beliefs in that they do not even purport to represent the way the world is”.

must have or form desires to change the way things are by resorting to this or that course of action. The beliefs tell the agent about how things are and about how they can be altered; the desires attract the agent to how things are not but can be made to be.

This rough sketch of beliefs and desires is widely endorsed in contemporary philosophy; it derives in many ways from the seminal work of the eighteenth century Scottish philosopher David Hume. The striking thing about it, from the point of view of desire, is that it characterizes desire by the job desire does in collaborating with belief and thereby generating action: it characterizes desire by function, not by the presence of any particular feeling. (Pettit, 1998)

But now things get complicated, and this drags us back to the metaphysical issues from above. The specific question here is: *What sort of mental state is expressed by our moral judgements?* So far we have considered only two options: a representing belief state, or a desiring attitudinative state. The former is a picture of the world, whilst the other is a want to change the world in some way. Crucially it is assumed that this is a mutually exclusive division. A mental state can be one or the other but not both. This classic conception of matters is due to the eighteenth-century Scottish Enlightenment philosopher David Hume. Hume himself thought that moral judgements were, in our terminology, desire-like states. (Fisher & Kirchin, 2006: 17-18)

The picture that I have just sketched, according to which some mental states are like maps of the world, and other mental states are like goals about which destination we are headed for, is sometimes called the *Humean Theory of Motivation*. It is called that because, in the eighteenth century, David Hume drew this picture in a very compelling way.⁷ (Schroeder, 2010: 11)

What strikes these quotations is the fact that rather than (just) some modest claim about the essence of (neo-)Humeanism, all authors above make a hard claim –to me they really

⁷ In footnote 9 on page 231 he adds: “Though this view is often called the Humean Theory of Motivation, not everyone agreed that Hume really believed it”. I am not sure whether such a sentence is of much help as a cautionary note when the main text states what it states.

read as truths beyond dispute- about what Hume thought.⁸ And most of them are inaccurate so we have to conclude on basis of the previous chapter. Only the last sentence of the passage by Fisher and Kirchin could reasonably be perceived as (very roughly speaking) sufficiently accurate.

Let us now suppose for the sake of the argument that we find *no more* hardcore claims about Hume in print than those listed above. One then might want to say that given the quantity of published misreferrals, there is no reason to be overly concerned. About these five passages, one might want to say, *just* these five passages, that present Hume incorrectly, one should not make too much fuss. I would want to disagree with such a response. The reason is that those passages come from widely read books.⁹ And perhaps not unimportantly, these passages are written by established scholars. The first and probably also the second are facts that should be given their due attention. It seems dangerous to have many frequently read meta-ethical books (some might want to say that some of these books exhaust the class of 'dominant books') on the market telling one that Hume thought something that he did not think (at all). Being an unexperienced second year or third year undergraduate interested in meta-ethics, one would need a very autonomous and critical mind to not have at least some sort of *prima facie* conviction that Hume indeed claimed what the (established) authors argue Hume claimed.

I have just explained why we should take the list of five seriously, even if no more hard-core claims about Hume can be found in print. There is a different related point that should be made. The fact that not very many meta-ethicists invoke Hume's text in their printed publications –which I take to be the case- by no means implies that other meta-ethicists, be they neo-Humeanists, anti-Humeanists or others, do not have the same mistaken views about Hume as advocated by the authors of the list I have provided above. Let me add

⁸Note that in four out of the five passages cited, the author(s) provide(s) no reference *at all* to back up the hard claim about Hume. In one case (the first passage from Smith) a book *year* is offered. In none of the fives passages, the author(s) cite(s) page numbers to back up their claim.

⁹ Rather than from a meta-ethical text book, the passage by Philip Pettit comes from the *Routledge Encyclopedia of Philosophy*.

two remarks to that. First, whether or not a smaller or bigger contribution to that is made by (the) explicit hard and published claims about Hume (cited) most probably differs for individual cases. Secondly, I take it that there is/are some good reason(s) to not make hard claims about Hume if one does not have the exegetical competence to do so. Such objective good reason could be intellectual humility/modesty or self-interest for example. In any case, such reasons and/or else simple common sense might (have) cause(d) a significant number of meta-ethicists to refrain from publishing hard claims about Hume.

5. Intellectual drawbacks and hindrances as a result of misreading Hume

5.1. The conceptual poverty problem

It is easy for a false hypothesis to maintain some appearance of truth, while it keeps wholly in generals, makes use of undefined terms, and employs comparisons, instead of instances. This is particularly remarkable in that philosophy, which ascribes the discernment of all moral distinctions to reason alone without the concurrence of sentiment. (Hume, EPM: 287)

In this fifth section, I focus on possible negative consequences of misreading Hume. That is, possible negative intellectual consequences for contemporary meta-ethics.

Meta-ethics seems to be an intellectual domain that belongs paradigmatically to the ‘analytic’ philosophical tradition. Given that, one would expect collective intellectual activity that includes the pursuit of conceptual accuracy and precision. However, in my considered opinion, some practices in meta-ethics when it comes to conceptual specification are extremely loose. That is, not much is happening in term of precision of and reflection on many concepts one could refer to as ‘mental states/items/faculties’.¹⁰ That is, not much

¹⁰ Generally, there are two big meta-ethical boxes: one encompassing ‘reason’ and ‘belief’ and one encompassing ‘desire, emotion, sentiment, attitude and feeling’, while there is very little reflection on differences between the things in those big boxes. Having said that, I point out that in meta-ethics, there is quite a bit of reflection on the concept of ‘desire’, although –as far as I can see - not as much and not as rich as in the ‘philosophy of desire’.

reflection takes place in regard to the concept 'belief', 'reason', 'sentiment', 'emotion' and 'feeling'. There seems to be an unfortunate cultural trend here. Assuming for the sake of argument that this is true, I do not want to go so far as to conclude that this definitely is caused by (even just partially) a misreading of Hume. But I certainly cannot exclude that this problem has some relationship with a misreading of Hume. Misreading Hume on belief seems to cause and nourish the fact that meta-ethicists think a lot in terms of a dichotomy between beliefs and desires, taking it as a theoretically important one. This, then, in combination with other facts, that could paradigmatically include the fact that there is a long-lasting intense camp battle between cognitivists and non-cognitivists, may cause the conceptual looseness that looks seems rather remarkable for an analytic discipline as meta-ethics wants to be. Because there is a cultural trend in dichotomy thinking with reason and belief on the one side and sentiments, desires, feelings, emotions on the other side (paradigmatically reflected by the debate about cognitivism and non-cognitivism) there is not only no trigger for conceptual reflection; there is even a positive motivation in the form of a dialectical atmosphere not to reflect.

In the remainder of this section I give one example of conceptual looseness and the problems that come with it. Below I distinguish seven different guises in which the notion 'reason' can be found in the meta-ethical debate, but that do not always seem to be clearly recognized as different guises and sometimes seem to be conflated. For the first four conceptions I present, I have invented related terms to make clear about what conception of 'reason' as I see it in the meta-ethical debate I am talking. The latter three conceptions carry terms one actually finds regularly in the meta-ethical debate.

The first conception of 'reason' we can find in the meta-ethical debate is 'morally authoritative reason'. Morally authoritative reason we can take broadly here in terms of moral principles as well as underlying justifications. Morally authoritative reason comes in all kinds of substantive forms.

Secondly, epistemic intellectual reason. Epistemic intellectual reason covers the epistemological part that either detects, is constitutive of, or plays some other significant role in terms of knowledge of substantive authoritative reason. Epistemic intellectual reason typically comes either in the form of intellectual intuition or else in the form of reasoning processes.

Thirdly, reason close-to-logos. This conception refers to some kind of all-encompassing faculty of knowledge that is rather ambiguous between the role of intellect and the role of morality as an authoritative power. Because it is all-encompassing it has some similarity with the ancient Greek conception of reason as a power to make sense of the world. It might be somewhat more narrow; the focus might be more on knowledge than on 'making sense of the world'.

Fourthly, there is a reason notion that relates to rationality; therefore let us call it reason-rationality. This conception refers to an optimal option under a set of constraints.

Fifthly, there is a reason as something that counts in favour of some judgement sensitive attitude. Scanlon argues that "[t]he question 'What is a reason?' is misleading insofar it suggests that reasons are a special ontological class. What is special about reasons is not the ontological category of things that can be reasons, but rather the status of being a reason, that is to say, of counting in favour of some judgement sensitive attitude (Scanlon, 1998: 56).

Sixthly, there is a normative/justificatory/justifying reason, generally understood as a thing that provides normative grounds for action (and belief). At places in *The Moral Problem* Smith engages heavily with the notion of normative reasons (see esp. Smith 1994, chapter 5). Perhaps it is correctly said that this sixth conception is a mixture of conception one and five while it always has a strong ontological (normative) nature. Conception five seems to have less of a normative nature and seems to be less normatively demanding than conception six.

Seventhly, in the meta-ethical debate one finds the concept of a motivating reason. This thing equals a motive for action. It signifies why a particular action was performed.

In what follows I concentrate on the first, second and third conception mentioned, but let me begin by admitting that it is not easy to pin down what 'reason' is. Having said that, in meta-ethics there seems to be room for a lot more conceptual clarity.

I believe that when talking about 'reason' in the context of the debate about cognitivism and non-cognitivism we meta-ethicists should be a lot clearer on the relationship between a psychological and a factualist claim.¹¹ As to the latter, at the very least it should become clear whether or not we are talking about a mental process/activity of reasoning, or an outcome of that: a judgment. And then a claim about its relationship with affect should be made. If the concept of 'reason' does not allow any role for affect at all, then that is a very good thing to know for our inquiries having to do with the nature and status of moral objectivity. Having said that, we may just have caught a paradigmatic piece of philosophical mythology that can be demolished by a sharp sentimentalist account and/or some bit of current or future empirical science.

Now, why exactly is it moral rationalists and realists should be interested in offering their opponents lots of precision when it comes to our notion of 'reason' in the meta-ethical debate, especially as being ambiguous between an authoritative institution of morality and an intellectual notion and in terms of a detailed specification of the latter? A first answer and a good starting point for others seems to be that the debate about the status of moral objectivity has an awful lot to do with the role sentiment plays in the metaphysics and epistemology of moral distinctions and that philosophical and scientific literature suggests that it is hard to prevent sentiments from coming onto the stage here in matters metaphysics and epistemology. Therefore, as theorists in the meta-ethical debate, we cannot afford being

¹¹ In the main text I have opted for much more clarity about a psychological and a factualist claim as part of the 'cognitivist' position. This claim concerns individual authors publishing individual papers as well as cognitivism as some sort of general overarching position that is opposed to 'non-cognitivism'. Having said that, what meta-ethics may first and foremost need is the awareness that cognitivism as some sort of general position is probably as much an involuntary blur as a voluntary conjunction of the following three types of claims: (i) some sort of robust metaphysical claim about *moral objectivity* (some sort of *factualist* claim), (ii) a claim about moral judgments being essentially *representational*, (iii) a claim about moral judgements being essentially *intellectual*. These are clearly three different claims and their connection (if figuring in a conjunction) should be maximally clearly put onto the table rather than that being done so rather loosely or indeed staying largely under the surface.

conceptually liberal. General (and multi-interpretable) use of terms makes it very difficult to challenge an (influential global) position as the moral rationalist one¹². General and multi-interpretable terms prevent people such as Hume who are critical towards something called ‘reason’, make good on their overarching worries. Such terms also prevent empirical scientists (either with a similar critical attitude or not) and whoever else to be usefully

¹² Meta-ethics has several intellectual realms in which one finds a bewildering variety of different positions about which an intelligent outsider can reasonably ask what they have in common. ‘Moral realism’ is such a realm. For ‘moral rationalism’, the variety is not quite as big as in the realm of moral realism, but it is still very significant. And the variety may be more nasty than in the domain of realism. In the domain of realism participants in the debate seem to be aware of the extreme variety and see it largely as a term of art for some sort of substantive position. I am inclined to think that meta-ethical participants see ‘moral rationalism’ far less as a term of art. Perhaps even as a clearly defined position that is opposed to ‘moral sentimentalism’. But a look at the literature clearly shows that there are very different *descriptions of ‘moral rationalism’ (as some sort of general position)* operative in the meta-ethical debate as well as wildly different *substantive views*. Below I offer some different descriptions of the general position ‘moral rationalism’ as one finds it in contemporary meta-ethical literature. I start with a statement from Shaun Nichols:

“Over the last 20 years, a number of central figures in moral philosophy have defended some version of moral rationalism, the idea that morality is based on reason or rationality ...According to rationalism, morality is based on reason or rationality rather than the emotions or cultural idiosyncrasies, and this has seemed to many to be the best way of securing a kind of objectivism about moral claims.” (Nichols, 2002: 285)

A very different description comes from James Doyle (2000: 1). He describes moral rationalism as “the view that moral constraints are rational constraints”. Then there is Mark van Rooijen’s description. He argues: “Metaethical rationalism can be roughly characterized as the idea that the requirements of ethics are requirements of practical reason” (Van Rooijen, 2010: 495).

Probably if one takes all substantive views about ‘moral rationalism’ (as self-ascribed by the author and ascribed by colleagues) together and tries to make a view out of it as to what the moral rationalist position amounts to, one gets something of the kind Russ-Shafer Landau offers. In this (2006), Shafer-Landau argues that moral rationalism is an umbrella term for a variety of positions. Shafer-Landau suggests that the central claims that have been associated with ‘moral rationalism’ are:

- (1) The metaphysical thesis: basic moral requirements are constituted by the deliverances of sound practical reason
- (2) The epistemological thesis: humankind’s basic moral requirements are knowable a priori
- (3) The normative thesis: moral requirements entail excellent reasons for action

I point out that Shafer-Landau argues that *not all moral rationalists endorse all of these claims*. Let me finally note a description from Terry Horgan and Mark Timmons. It needs to be noted however that Horgan and Timmons make clear that the rationalism about morality they talk about is a rationalism that concerns *the psychology of moral judgement*.

“According to rationalism regarding the psychology of moral judgment, people’s moral judgments are generally the result of a process of reasoning that relies on moral principles or rules. By contrast, intuitionist models of moral judgment hold that people generally come to have moral judgments about particular cases on the basis of gut-level, emotion-driven intuition, and do so without reliance on reasoning and hence without reliance on moral principles.” (Horgan & Timmons, 2007; 279)

critical in terms of the challenge/refutation of a particular claim about 'reason' made by moral realists and moral rationalists. And the upshot of that may be that we miss out on some thoughts from people like Hume, or empirical scientists, or others, that plausibly reflect the idea that there is a significant role passions play when it comes to the nature of moral distinctions, whatever that role exactly may be.

Let us go back to Hume to illustrate the point I have just been making. As we have seen in chapter 2, Hume has a concern about reason –as an intellectual notion- being incapable of terminating a thinking process (consisting in 'interposition of ideas'). And as we know as well, Hume also thinks that only impressions – recall that for Hume passions are one type of impression- can help reason out here. That is, by being a crucial component for that which can terminate a reasoning process: a judgement.

As we have seen in the previous chapter, there are various complexities as to how Hume's general notion of 'reason' figures in his moral philosophy. One complexity consists in the fact that in his moral philosophy (in contrast to other areas of his work), Hume's stance towards 'reason' is largely polemical. That is, unlike what is the case for other areas of his work, when Hume writes about reason in his moral philosophy, Hume, in an oppositional way, responds often to the views of his moral rationalist contemporaries' who advocate the existence of timeless and eternal truths of reason. Another complexity - a related one however- comes from the fact that Hume, just as in other areas of his work, in his moral philosophy is often a 'cognitive psychologist'. As such he takes as a starting point for description actual moral judgements made by human beings. These sorts of (subjective) moral evaluations however are not the primary objects of attention of his rationalist contemporaries whose views Hume rejects and tries to reject by means of his own sentimental theory of morality.

Granting and admitting that there are problematic aspects in Hume's philosophy as laid out in a detailed way in the previous chapter and as summarized above, I would say that given (a) the prima facie implausibility of theories that consist in a claim about the existence

of mind-independent (timeless and eternal) truth of reason and given (b) the power of Hume's idea that a process of interposition of ideas without help of impressions can never result in some sort of evaluative distinction, it is extremely relevant that any contemporary moral realist and rationalist makes clear the exact relationship between an authoritative and intellectual notion of reason in his theory. Furthermore, moral realists and rationalists should spell out into detail any conceptual claims they want to defend about an intellectual notion of reason. That is, in a maximally informative way that includes the absence of any ambiguities about reason as a reasoning process and reason as judgment.

Below I present a quotation from Rachel Cohon. As I hope to have made clear in the previous chapter, I do not completely agree with Cohon's interpretation of Hume. That is, I think Hume's concept of 'reason' is (indeed- in line with what Cohon argues for) largely a concept of the process of reasoning, but this, according to me, does not exhaust Hume's concept of 'reason'. According to me, Hume's concept of reason cannot completely be captured by the process/activity view of Cohon. For the sake of clarity, I would like to recall this little, but significant disagreement between Cohon and me before presenting the, following statement from Cohon I take to be useful as a stimulus for conceptual specification of the contemporary moral rationalist position.

While Hume's conception of reason as reasoning may not satisfy us today, it is not unwarranted or tendentiously narrow (though it may still be too narrow), and it raises what remains a very good question: If there is something more to reason than [the process of reasoning], what is it, and how is that further factor related to the activity of inference? This sort of question must have concerned Kant, who was at pains to demonstrate the unity of theoretical and practical reason. If the essence of immorality lies in an action or trait's unreasonableness, then we must be made to see why this is rightly called by the same name that is applied to the outcome of incorrect inference. (Cohon, 2008 : 94)

I think the general methodological suggestion of Cohon's message above is a valuable one. In connection with what I have said earlier, I think it does not do sufficient justice to Hume's ideas on 'reason'. But that is no problem for now. Namely, this critical note from me that comes from an exegetical point of view does not prevent Cohon's passage from being independently plausible as a methodological suggestion for (improvement of) the concept of 'reason'.

Generally speaking, my concern as expressed in this section comes down to the following. I think that in meta-ethics there is some sort of phenomenon of conceptual underdetermination in terms of the concept 'reason' that clouds what deep down is theoretical vulnerability, a vulnerability that comes to the surface (and perhaps only then) with better (more detailed) specification of the popular realist and rationalist notion 'reason'. This needs to happen so that sentimentalists and scientists have a maximally well-specified potential object of refutation in terms of the psychological notion of reason. Importantly, this on its turn may very well have an effect on the normative claims of realists and rationalists. That is, once some ambiguous notion of 'reason' is supposed to play a constitutive role in moral discriminations and once there are arguments to believe that there are limits to an intellectual power of reason, an appeal to sentiments may be needed to make up for intellectual reason's limitations. This is indeed what seems to happen in Hume's philosophy (albeit partly in an undesirable ambiguous implicit way). This fact then that sentiments enter the constitutive scene of moral discriminations (however much we might wish something else) may then imply there being limits to the status of 'authoritative moral reason', whatever those limits might be.

5.2. The interconnection problem

There are close connections between three of the core debates on meta-ethics, i.e. the debate about the nature and status of moral objectivity, the debate about the nature and status of

moral judgements and the debate about moral motivation. Not only are there theoretically intelligible connections between those debates, many meta-ethicists also reason in an interconnected way about the debates mentioned. Before I give an example of this, let me first mention something about the debate between cognitivism and non-cognitivism.

If one asks an advanced undergraduate student who has just completed a course in meta-ethics to describe in rough terms the debate between moral cognitivists and non-cognitivists most probably he or she will *not* offer a story about moral knowledge (which one would probably expect given the terms), but will tell (perhaps not exclusively, but probably as a first thing) a particular story about the nature of moral judgements. Most probably, the student will mention that according to cognitivists, moral judgements are beliefs, mental states that purport to reflect the way the world is morally. And she will note that while cognitivist-type moral judgements that should be called 'beliefs' are apt for truth and falsity, non-cognitivist-type moral judgments are not. Probably the student will mention that (paradigmatic) non-cognitivist-type moral judgements are not beliefs and that they are not because they are not in the business of stating something true or false about the moral world and thereby they are not mental states that represent anything that is part of the world. And the student most probably will add in the same breath that this is implied by the passion-like nature of non-cognitivist moral judgements.

When it comes to problematic belief-concepts, the problem with the cognitivism versus non-cognitivism debate seems to be that the belief-concept that is operative in this debate is really dogmatic. In my perception, in the cognitivism versus non-cognitivism debate proponents as well as opponents allow cognitivists to work *en masse* with the concept of a moral belief mentioned above. That is, it is standardly regarded as a 'mental state that purports to reflect the way the world is morally'. Both parties never seem to question this concept. And that seems a problem as well as a rather awkward phenomenon. It is a problem given that the term 'moral belief' may very well deserve a more suitable concept, one that reflects better what a moral belief is and/or helps us better to understand the problems of

meta-ethics. It seems rather awkward because one of the core tasks and obligations of philosophers seems to be critical towards concepts. Very critical, if intellectual purposes demand. But there does not seem to be any significant critical reflection¹³ on the concept 'belief' in the debate about cognitivism and non-cognitivism. It is a standard concept and one employed dogmatically. And that seems to be a real problem in terms of intellectual progress, especially so given neo-Humeanism's belief concept. Here my suggestion is that if the debate about the truth of neo-Humeanism is run relatively independently from the debate between cognitivists and non-cognitivists, there is a danger that two operative beliefs concepts find each other and feed each other.¹⁴

For the sake of the argument having considered the debate about the truth of neo-Humeanism as independent from the debate about cognitivism and non-cognitivism, I now point out that there is an indirect connection between these two debates. To see this, let us assume that neo-Humeanism reflects some sort of theory about the relationship moral judgement and moral motivation. Not any type of moral judgement however. The relationship between moral judgement and motivation concerns a particular type of moral judgement called 'belief'. Since it is assumed that beliefs are 'cognitive' states (see § 5.1. above) and since it is assumed that such belief-type cognitive states contrast with non-cognitivist moral judgement that have a sentimentalist component in them, it is assumed that there arises a puzzle:

¹³ It seems fair to make an exception for Terry Horgan and Mark Timmons (2000). 'Nondescriptivist Cognitivism: Framework for a New Metaethic', *Philosophical Papers*, 29, pp. 121–153; cf Timmons, M. (1999). *Morality Without Foundations*. New York: Oxford University Press.

¹⁴ Here is another possible effect of the dogmatic belief conception in contemporary meta-ethics. Once (i) it is assumed that the only mental state that can reflect moral knowledge is a 'belief' that is supposed to be 'a mental state that reflects the way the world is morally' and once (ii) some meta-ethical *psychological* conception of 'reason' fully or at least primarily comes in a *non-emotional* guise, one probably sneakily smuggles in notions of representation/detection and pushes aside notions like 'projection'. This may well result in preventing ourselves from the opportunity to see how authoritative conceptions of 'reason' are crucially dependent on emotion/affect; something that could be crucial to the legitimacy of objectivity pretensions of 'morally authoritative reason'.

The real puzzle as to how moral judgments can motivate arises for those who maintain that moral judgments express *moral beliefs*, for the connection between belief, a *cognitive* state, and motivation is uncertain. (Rosati, 2006: § 3.1)

Let us take up this quotation in a wider literal context so that it becomes clear that and how there are direct links between the debate about neo-Humeanism and the debate between cognitivists and non-cognitivists.¹⁵

[O]ne way in which moral judgments could motivate, and, indeed, motivate on their own, would be if moral judgments were not representational after all. Suppose moral judgments did not ascribe properties and express moral beliefs about what things have those properties. Suppose instead, as *moral noncognitivism* maintains, that moral judgments express desires or other *conative* states—what philosophers sometimes call ‘pro-attitudes’. Then it would be clear how moral judgments connect to motivation. They simply express a motivating state that the individual already has; to make a moral judgment is already to be motivated, at least to some degree. The real puzzle as to how moral judgments can motivate arises for those who maintain that moral judgments express *moral beliefs*, for the connection between belief, a *cognitive* state, and motivation is uncertain. (Rosati, 2006: § 3.1)

Let us now have a look at the following argument:

1. Necessarily, if one sincerely judges an action right, then one is motivated to some extent to act in accordance with that judgement (*Motivational Judgement Internalism*)
2. When taken by themselves, beliefs neither motivate nor generate any motivationally efficacious states (*Motivational Humeanism*)
3. Therefore moral judgements are not beliefs (*Moral Non-cognitivism*) (2003: 121)

¹⁵ NB: the quotation below may also make plausible my claim that if the debate about the truth of neo-Humeanism is run relatively independently from the debate between cognitivists and non-cognitivists, there is a danger that two operative beliefs concepts find each other and feed each other.

4. Beliefs are our mental vehicles for representing facts
5. Therefore moral judgements are not essentially fact-stating
6. Moral realism is false^{16 17 18} (Shafer-Landau, 2003: 121; *cf* 22, 120)

Let us assume that there is indeed such a train of thought in meta-ethics (possibly up until (3) and possibly *only* up until (3) the train of thought can be seen as a *cultural trend*). This means that the idea that beliefs do not motivate can have extremely far reaching consequences. If that is true, then it is not clear yet that this is primarily or exclusively due to an incorrect reading of Hume. But the train of thought pointed *may* to some extent be caused by a misreading of Hume's work on belief and reason.

5.3. Intellectually surpassed by the misread philosopher?

¹⁶ Note that Shafer-Landau, as someone who presents this argument, believes that there is good reason to be suspicious of thesis 1 and 2 of the above argument. Shafer-Landau tries to show that it is false to think that beliefs can never be sufficient for motivation and that it is at best questionable that desires are necessary for motivation.

¹⁷ In a footnote, Shafer-Landau explains that no one lays out the argument quite as baldly as he does above, but that it nevertheless does seem to be "an accurate representation of an argumentative strategy" endorsed by:

- Stevenson, C.L. (1937). 'The emotive meaning of ethical terms'. In Stevenson, C.L. (1963). *Facts and Values*. New Haven: Yale University Press
- Aiken, H. (1950). 'Evaluation and obligation'. *Journal of Philosophy*, 47, pp. 5-22.
- Hare, R.M. (1952). *The Language of Morals*. Oxford: Oxford University Press, pp. 79-9.
- Nowell-Smith, P.H. (1954). *Ethics*. London: Penguin Books, pp. 36-43.
- Harman, G. (1975). 'Moral relativism defended'. *Philosophical Review*, 85, pp. 3-22.
- Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin, pp. 27-42.

Shafer-Landau argues finally that "Simon Blackburn also advances this argument, but with some reservations" (Shafer-Landau, 2003:121). See Blackburn, S. (1984). *Spreading the Word*. Oxford: Oxford University Press, pp. 188-9.

¹⁸ Note that that the first three steps of the presented argument/train of thought are explicitly laid out in the way I do above: they reflect a quotation. Step 4-6 follow from running main text of Shafer-Landau's book (text on the same page as step 1-3).

In this section I explore the possibility that Hume as an author misread by contemporary meta-ethicists intellectually surpasses his misreaders in terms of the misreaders' own intellectual goals. In order to make a claim as to whether or not this happens, we must make a comparison between some aspects of Hume's philosophy and the intellectual quality of meta-ethical theories aiming at a particular intellectual aim X. This is difficult, partly because it is not always very clear what the intellectual aim of meta-ethical theories and/or their authors is. What does neo-Humeanism in meta-ethics aim at? Is it:

- The explanation of moral motivation and/or action.
- A mildly comprehensive theory about objective justification of our beliefs and actions (with or without special attention to something (whatever it may be) having to do with 'folk psychological practices').¹⁹

¹⁹ Some philosophers, Michael Smith for example, connect their defence of or challenge to neo-Humeanism to something they call folk psychology (see e.g. Smith, 1987; 1988, 1994; Pettit, 1987). In his (1987) Smith claims that philosophers have the task to formulate a philosophical conception of folk psychological states (1987a; cf 1987b; 1994). His (1994) sheds a little more light on what this means. Namely, in that work Smith argues that there are two distinctive features of morality "that are manifest in ordinary moral practice as it is engaged in by ordinary folk" (1994: 5). The first distinctive feature concerns the objectivity of moral judgement. The second distinctive feature is the practicality of moral judgement. The first amounts to the idea that "[m]oral judgements of the form 'It is right that I Φ expresses a subject's beliefs about an objective matter of fact, a fact about what it is right for her to do". The second reflects the idea that "If someone judges that it is right that she Φ s then *ceteris paribus*, she is motivated to X" (Smith, 1994: 12; more on these features in § 6). Smith argues that it is "[t]he philosopher's task is to make sense of a practice having these features" (1994: 5).

Although Smith states *something* as to what his conception of folk psychology refers to, it is rather unclear to me what *exactly* he means thereby. The explanation in his (1994) is more informative than the one he gives in his (1987) in that it becomes clear from the former work that sense must be made of the objectivity and practicality of moral judgement that are supposed to be features of ordinary practice. But a detailed specification of the concept 'folk psychological states' is not available. The same counts for the task ascribed to the philosopher 'to make sense of ordinary moral practice'.

For a helpful article about ambiguity of the concept folk psychology, see Stich & Ravenscroft (1994). Stich and Ravenscroft distinguish between external and internal accounts of folk psychology. On an external reading folk psychology "ain't in the head". (Stich & Ravenscroft, 1994: 460). External accounts either collect or systematize the intuitively recognizable generalizations of folk psychology, while internal accounts focus on the cognitive mechanism that underlies our ability to have those intuitions, to predict behaviour and so on.

Apart from making the distinction mentioned, Stich and Ravenscroft helpfully present five ways in which folk psychology can be an internally represented theory of human psychology exploited in e.g. the explanation or prediction of behaviour.

- A heavily comprehensive theory about objective justifications of our beliefs and actions *and explanation* of our actions (with or without special attention to something having to do with ‘folk psychological practices’).
- A comprehensive justificatory and explanatory theory about ‘reasons for action²⁰’ in the moral domain.

What plays a role in the uncertainty –or at least my uncertainty- as to what exactly neo-Humeanism aims at, is the theory’s, or at least, the authors’ attachment to the *neo-Humean Split* as something of utmost importance that needs to be upheld. What comes with that is the concept of a belief. In meta-ethics belief is the bearer of objectivity. So that intellectual concern enters the scene. One question I have is what remains of neo-Humeanism and its purposes once the strict distinction between beliefs and desires is not anymore taken as one that is essential input for neo-Humeanism. I do not know the answer.

The (presumed) fact that it is not very clear what the intellectual aim is of meta-ethical neo-Humeanism and or its authors, makes it difficult to say whether or not (core) aspects of Hume’s philosophy surpass the philosophy of meta-ethical authors defending (or

²⁰ In meta-ethics, neo-Humeanism is characterized either as a theory about ‘*moral motivation*’ or as a theory about ‘*reasons for action*’. Tim Schroeder describes the latter in the following way:

“[In meta-ethics there exists a vibrant debate] over the relation of desires to reasons to act. According to one tradition, typically called ‘Humean’ or ‘Neo-Humean’, the existence of reasons to act depends on the existence of desires possessed by the agent who would act. Thus, my reason to drink hot chocolate depends on my desires, and likewise my reason to help a stranger depends on my desires, according to the Humean.”
(Schroeder, 2009)

The former conception might well result from a mistaken reading of Hume on beliefs. At least the explanatory (if not also some justificatory) seeds for the latter meta-ethical conception might lie in Donald Davidson’s (1963), a work that has been very influential in especially the philosophy of action. An alternative or further cause for the ‘reasons for action-conceptions’ of neo-Humeanism in meta-ethics might be Bernard Williams’s thoughts on ‘internal reasons’ as advocated in his (1981). In any case, the neo-Humeanist ‘reasons for action’-conception as it appears in contemporary meta-ethics seems to be strongly associated with the debate about internal and external reasons for action that –as a contemporary meta-ethical debate- is still very much about and has its roots in the philosophy of Bernard Williams (for Williams’ position see also his 1995 a, 1995b, 2001; for critical engagement with Williams’ position see e.g. McDowell, 1995).

else challenging) meta-ethical neo-Humeanism in the sense of the Hume's philosophy being more argumentatively rich than the neo-Humean doctrine, while neo-Humeanist authors might well adopt an unnecessarily simplistic conceptual dictionary, and therefore possibly end up with suboptimal theory construction, because of a misreading of Hume.²¹

After this engagement with the 'surpassing-problem' in relation to *neo-Humeanism*, in the remainder of this final section I engage with the same problem while addressing *more broad meta-ethical aims*. That is, I now assume that three primary aim of meta-ethics and its authors are (1) offering the best possible account of the nature and status of moral judgements; (2) offering the best possible account of the nature and status of moral objectivity; (3) offering the best possible account about the relationship between moral judgements and motivation. It is difficult to say in what way and to what extent we can justifiedly say that Hume shared these assumed meta-ethical core aims. Here, let me say first that I think it uncomfortable and anachronistic to somehow squeeze Hume's holistic and rich philosophy into these three goals. Having said that, I would also want to say that one cannot reasonably deny that Hume wanted to say something about all three. Assuming for the sake of argument and reality that Hume not only says, but indeed intended to say something about these three (broad) meta-ethical goals (and intentionally), what I think should be avoided in contemporary meta-ethics is the following. A situation characterized by the fact that meta-ethicists misread Hume as a result of which there is some sort of obsessive engagement with a distinction between beliefs and desires that prevents intellectual progress and/or an insight into the theoretical philosophy of Hume that has the intellectual substance

²¹ The concern I have raised will also apply to and may especially also apply to that part of the meta-ethical literature that is about 'besires' (see e.g. Zangwill, 2008). Besires are conceived of as unitary states that have both the representational characteristics of beliefs and the motivational characteristics of desires (Smith, 1994, p. 119). Matthew Bedke (2009: 207; cf. Bedke, UM) points out that the most forceful arguments against besire theory depend upon the separability of cognitive and non-cognitive functional roles (cf. Smith 1994, chapter 4). He also points out that these arguments fail to refute the claim that these separable roles are actually realized by one and the same mental state token in our moral judgments. Dependent on what the (different?) aims of besire theorists are, Bedke may be right on the first. Bedke seems anyhow right on the second. Whether or not Bedke is right on the first (and if so probably, therefore on the second), one may still have the concern –and I have it- that Hume's original philosophy of belief can silence much if not all of the contemporary discussion on 'besires'.

to account for a satisfactory theory in regard to one or more of the three broad contemporary meta-ethical intellectual goals of I have mentioned above.²²

²² In this chapter I have been concerned with neo-Humeanism in *meta-ethics* and have not been concerned with neo-Humeanism in other intellectual areas. This partly for reasons of space. The space argument becomes more pertinent given that an enormous amount of complexity follows from properly descriptively and evaluatively comparing and contrasting *meta-ethical* Humeanism with neo-Humeanism in other areas, e.g. the philosophy of action. Below I very briefly compare and contrast an aspect of neo-Humeanism in the philosophy of action with meta-ethical neo-Humeanism.

It sounds plausible to me to say that there is such a thing as a ‘traditional’ Humean picture in which desires initiate action and are coupled with means-end beliefs to yield action (see Shafer-Landau, 2003). Compare:

“One debate over the status of explanations of human action turns on how we should interpret the following principle of folk psychology (Roth, 2012):

[L] If any person, agent, individual, wants some outcome, d, and believes that an action, a, is a means to attain d under the circumstances, then x does a” (Rosenberg, 2012, cited in Roth, 2012; my italics).

The ‘traditional Humean picture’ may follow from an instrumentalist reading of Hume (possibly since a few decades somehow accompanied by some influence from Davidson’s (1963)). It looks like this ‘traditional picture’ maps incredibly accurately onto a big part of our daily practical life. Imagine e.g. a hot summer evening when you are craving for an ice-cold bottle of beer that lies in the fridge. You will be concerned with how to get the beer from the fridge into your hands. You may consider getting up from your comfy garden chair and walk to the fridge. You believe that this is a way to achieve your goal. Or else you might believe that begging your partner will do the trick. Life really seems to be full of such examples of *–intentional-psychological* belief-desire combinations. Also, from an *externalist* perspective it seems rather plausible to assume that your action can be explained by a combination of beliefs and desires. Therefore there is (*prima facie*) reason to think that neo-Humeanism (in its mentioned ‘traditional form’) is not misplaced (and probably even really useful) for the philosophy of action and some wider area of practical philosophy (similar claims can be made about ‘folk psychology’).

So in regard to the abovementioned example (supposed to represent a non-moral subclass of practical action) it seems highly plausible to see the ‘traditional neo-Humeanist picture’ as applicable. But things seem to be different for moral matters. While a large part of our daily practical life seems to work in the belief-desire way reflective of the example given, it does not seem to be a plausible way to sketch our moral life, unless one has a very broad conception of what our moral life amounts to (a picture that encompasses everything or more or less everything practical). In regard to a psychological internal point of view, I would argue that a paradigmatic and substantial part of our *moral* practical life is not (very) well captured by ‘moral instrumentalism’. What one desires or aims at is generally *not* an *intentional guiding force* when it comes to the area of morality. When it comes to morality, we typically do not intend to fulfill our desires. Our daily life practices and experiences of morality most of the time do not start with aims and desires for which we need beliefs as means to realize them. In a particular situation that demands our moral attention we generally start with some sort of moral conviction (or else plain doubt and deliberation) and act in the light of what (after deliberation) we think should happen. Typically it does not seem to be the case that a conative state starts off some kind of ‘instrumentalism about morality’ in our minds. From an externalist point of view it seems plausible enough to say that a typical moral case was performed because the agent had a desire and a belief. But I bet, Hume’s extensive philosophy can do a lot more plausible explanatory work here than any account of beliefs or desires in meta-ethics.

Chapter 4: Kant's moral philosophy and the limits of moral objectivity

1. Introduction

This chapter is about Kant's philosophy. Kant aimed at a justification for moral discriminations completely free from anything empirical. Said slightly differently, Kant's *Formula of Universal Law (FUL)* is a strikingly *a priori* form of moral justification. In this chapter I discuss *FUL* and other aspects of Kant's moral philosophy into detail. There seems something special and important about *FUL* from the perspective of the problem of the status of objective morality. That is, what seems to characterize *FUL* is that it is a mind-independent justification. While that seems true, at the same time *FUL* seems to be *grounded in* the reasoning capacities of practical agents. Generally, such a subjective constructivist aspect avoids the problem of how authoritative bits of moral objectivity gain their place in the universe. Another thing to mention about the alleged mind-dependent aspect of *FUL*, is that it does not seem to be constructively dependent on sentiments in any sense.

There are two reasons for focusing on Kant's moral philosophy given an interest in the status of moral objectivity. First, given the alleged completely empirical-free nature of *FUL* (a. mind-independent, b. grounded in pure intellect), an investigation of *FUL* might tell something about the possibility for *FUL* and/or about a more general possibility for certain or all objective moral justifications to have a really hard objectivity status. Secondly (alternatively or additionally), an investigation of Kant's moral philosophy might open up possibilities to *put limits on the objectivity status* of *FUL* and/or other sorts of alleged objective justifications because of some philosophical-cum-empirical issues having to do with freedom and/or embodied agency. These issues are pertinent to Kant's philosophy, largely as *underexplained (mysterious) views* about the relationship between some sort of intelligible aspect of the self and a natural aspect of the self that both seem to be part of Kant's conception of a human being. Issues about intelligible and natural aspects of the self also

have a strong relationship with *FUL*. In the guise of pertinent but underexplained views, Kant's ideas about intelligible and natural selves may be of help to determine criteria for overambitious claims about moral objectivity. A focus on Kant's views might draw our attention to issues having to do with limited judgemental freedom (as a result of some form of embodied agency). Such issues, on their turn, might stand in some relationship with a less than die-hard factual status of moral objectivity. Indeed this is a view I will propose in the conclusion.

While I believe Kant's work can do a lot in terms of drawing our attention to issues having to do with limited judgemental freedom (as a result of some form of embodied agency) as of relevance to the problem of moral objectivity, in this study, an investigation of Kant's work will also turn out to be of use in terms of comprehending the strength of Hume's claim that reason alone cannot make moral distinctions.

This chapter is one that is exclusively concerned with Kant exegesis. It is in the next chapter and the general conclusion I connect Kant's philosophy to my meta-ethical project.

This chapter is set up as follows. In § 2, I give a brief introduction to Kant's philosophical project as exemplified by his first and second *Critique*. In § 3.1, I briefly focus on a concept central to Kant's *theoretical* work. That is, the concept 'noumenon'. My discussion includes a discussion of a problematic aspect of Kant's writings about the noumenon. In § 3.2, I discuss Kant's account of the 'homo noumenon', a concept important to Kant's *practical* work. I claim that the alleged problematic aspect of the 'noumenon' is in a comparable way present in Kant's concept of the 'homo noumenon'.

In § 4, I focus on 'the *a priori*' and 'the empirical' in Kant's practical philosophy. In § 4.1, I focus on the relationship between 'the *a priori*' and 'the empirical' as reflecting a sharp distinction between the two with a superior role for 'the *a priori*'. In § 4.2, I explain and illustrate the importance of the empirical in Kant's practical philosophy by discussing one predominant way in which the empirical figures in a non-inferior way in Kant's practical philosophy.

In § 5, I focus on Kant's notion of 'Vernunft' (typically translated as 'reason'). In § 6, I offer a first discussion of Kant's *FUL*. In this section, my focus will be primarily on the relationship between *FUL* and *reasoning*. In § 7, I focus on Kant's ideas about freedom and embodied agency. In § 8, I try to make comprehensive sense of *FUL* by combining information we have gained thus far and adding several different types of thought to it based on some striking remarks Kant makes in this theoretical and practical work. In § 9, I offer some final Kant exegesis. I focus on his account of the mind of the evil person and present his account of conscience.

2. An introduction to Kant's work

2.1. Kant's philosophical project: comparing the first and second *Critique*

There are striking differences between the nature of Kant's project as he pursues it in his *Critique of Pure Reason* compared to the *Critique of Practical Reason*. In the preface to the first edition of *KrV* Kant describes his project as follows:

[By a critique of pure reason] I ...mean....a critique....of the faculty of reason in general, in respect of all knowledge after which it may strive *independently of all experience*. It will therefore decide as to the possibility or impossibility of metaphysics in general, and determine its sources, its extent, and its limits- all in accordance with principles. (CpR-KS: 9/ A: 9)

Slightly later in the same preface Kant explains his project by saying that the question around which his project revolves is the question "what and how much can the understanding and reason know apart from all experience?" (CpR-KS: 12/ A: 11-12).

What becomes clear from these Kantian remarks is that Kant's project as laid out in the first *Critique* has a very important role for the *limits* of both reason and metaphysics.¹ This focus on *limits* in the first *Critique* has a strong relationship with Kant being all in favour of doing philosophy in a non-preposterous and careful way. This partly, but not exclusively, because of the need to ground any philosophical conclusions about knowledge on extremely stable foundations (see e.g. B: 13, 21). For the remaining one seemingly can say that Kant had an aversion to preposterous philosophical claims (i.e. 'don't pretend that you can claim more than you can reasonably claim'); dogmatic philosophy ('look for the pathway of most resistance; actively expose yourself to problems that do not suit you') and insufficiently careful philosophical inquiry.

Kant does not see it as important to focus essentially on the *limits* of metaphysics and reason in his *practical* philosophy. In the context of explaining and justifying the title of his second *Critique* Kant argues:

Why this critique is called simply *Critique of Practical Reason* and not *Critique of Pure Practical Reason*, though the parallelism between it and the critique of speculative reason seems to demand the latter title, will be sufficiently shown in the treatise itself. Its task is merely to show that there is a pure practical reason, and, in order to do this, it critically examines reason's entire practical faculty. If it succeeds in this task, there is no need to examine the pure faculty itself to see whether it, like speculative reason, presumptuously overreaches itself. For if pure reason is actually practical, it will show its reality and that of its concepts in actions, and all disputations which aim to prove its impossibility will be in vain. (CprR-LWB: 118 / KpV: 3)

There clearly is much less of a focus on the limits of reason in Kant's practical work, although incidentally one finds some remarks that point into that direction. As to there being remarks (albeit incidental ones) about 'reason' and 'limits' in Kant's practical work, that

¹ see also e.g. A:238 / B:297.

seems *fortunate* assuming that Kant believed that ‘reason’ is in the end one and the same thing in the theoretical and practical domain, which he did (see G: 391).

As indicated, the focus on the limits of reason that essentially characterizes Kant’s project in the first *Critique* has a strong relationship with Kant being sympathetic to pathways of philosophical enquiry that are conducted in a secure way and rest on extremely stable foundations (see e.g. B: 13, 21). This caution, just as the caution with ‘limits’, I take it, is less present in his practical work. And that seems to be well-reflected by his often ambiguous and puzzling remarks about the intelligible (noumenal) nature of human agents with their capacity for ‘Vernunft’.

3. Kant’s account of the noumenon

3.1. The concept of the noumenon in the first *Critique*

Kant’s account of the ‘noumenon’ is puzzling, at least to me. What is important about this for this chapter is that I think some problematic puzzling aspect of this concept as we find it in Kant’s theoretical work is also reflected by his account of the ‘homo noumenon’ as we see it in Kant’s practical work. Before I comment on this puzzling aspect of the ‘homo noumenon’, indeed before I in any way discuss Kant’s concept of the ‘homo noumenon’, I first discuss Kant’s notion of the ‘noumenon’ as it appears in his *Critique of Pure Reason*. Below I present a long quotation from the section of noumena and phaenomena in Kant’s first *Critique*. This I do because I would like Kant to speak for himself on this difficult topic (in a sufficiently extensive way), also largely so as to see clearly and have the textual evidence for an ambiguity in Kant’s concept of the noumenon. This then is the longer quotation:

Appearances, so far as they are thought as objects according to the unity of the categories, are called phaenomena. But if I postulate things which are mere objects of understanding, and which,

nevertheless, can be given as such to an intuition, although not to one that is sensible – given therefore coram *intuit intellectuali* – such things would be entitled *noumena (intelligibilia)*.

Now, we must bear in mind that the concept of appearances, as limited by the Transcendental Aesthetic, already of itself establishes the objective reality of *noumena* and justifies the division of objects into *phaenomena* and *noumena*, and so of the world into a world of the senses and a world of the understanding (*mundus sensibilis et intelligibilis*), and indeed in such manner that the distinction does not refer merely to the logical form of one and the same thing, according to which as it is indistinct or distinct, but to the difference in the manner in which the two worlds can be first given to our knowledge, and in conformity with this difference, to the manner in which they are in themselves generically distinct from one another. For if the senses represent to us something merely as it appears this something must also in itself be a thing, and an object of a non-sensible intuition, that is, of the understanding. In other words, a [kind of] knowledge must be possible, in which there is no sensibility, and which alone has reality that is absolutely objective. Through it objects will be presented *as they are*, whereas in the empirical employment of our understanding things will be known only *as they appear*. If this be so, it would seem to follow that we cannot assert, what we have hitherto maintained, that the pure modes of knowledge yielded by our understanding are never anything more than principles of the exposition of appearance, and that even in their *a priori* application they relate only to the formal possibility of experience. On the contrary, we should have to recognise that in addition to the empirical employment of the categories, which is limited to sensible conditions, there is likewise a pure and yet objectively valid employment. For a field quite different from that of the senses would here lie open to us, a world which is thought as it were in the spirit (or even perhaps intuited), and which would therefore be for the understanding a far nobler, not a less noble, object of contemplation...

The cause of our not being satisfied with the substrate of sensibility, and of our therefore adding to the phenomena noumena which only the pure understanding can think, is simply as follows. The sensibility (and its field, that of the appearances) is itself limited by the understanding in such fashion that it does not have to do with things in themselves but only with the mode in which, owing to our subjective constitution, they appear. The Transcendental Aesthetic, in all its teachings, has led to this conclusion; and the same conclusion also, of course,

follows from the concept of an appearance in general; namely, that something which is not in itself appearance must correspond to it. For appearance can be nothing by itself, outside our mode of representation. Unless, therefore, we are to move constantly in a circle, the word appearance must be recognized already indicating a relation to something, the immediate representation of which is, indeed, sensible, but which, even apart from the constitution of our sensibility (upon which the form of our intuition is grounded), must be something in itself, that is, an object independent sensibility.

There thus results the concept of a *noumenon*. It is not indeed in any way positive, and is not a determinate knowledge of anything, but signifies only the thought of something in general, in which I abstract from everything that belongs to the form of sensible intuition. But in order that a noumenon may signify a true object, distinguishable from all phenomena, it is not enough that I *free* my thought from all conditions of sensible intuition; I must likewise have ground for assuming another kind of intuition, different from the sensible, in which such an object may be given. For otherwise my thought, while indeed without contradictions, is none the less empty. We have not, indeed, been able to prove that sensible intuition is the only possible intuition, but only that it is so for us. But neither have we been able to prove that another kind of intuition is possible. Consequently, although our thought can abstract from all sensibility, it is still an open question whether the notion of a noumenon be not a mere form of a concept, and whether, when this separation has been made, any object whatsoever is left. (CpR-KS: 265-271 / A: 162-5)²

I have underlined Kantian claims that I take to be of special importance for understanding Kant's notion of the noumenon. Or rather, or in addition, they are underlined to show an important ambiguity in Kant's writings about the noumenon. In line with popular academic thought, Kant gives the impression that a noumenon is a thing in itself. However, when Kant talks about the noumenon, at times, he also gives the impression that it –probably partially rather than exclusively – is a concept that is about a particular kind of intuitive perception: an intellectual rather than a sensuous one. I take it to be an unsettled scholarly issue as to how to balance the 'thing in itself' aspect' and the 'particular kind of intuitive perception'

² This passage appears in only the first edition of the first *Critique*.

aspect in one concept. That is, if they have to be merged in order to accurately reflect Kant's views.

Kant prominently associates the term 'intellektuelle Anschauung' with the concept of a 'noumenon'. There is something important about the two German terms 'intellektuelle Anschauung' and 'sinnliche Anschauung', (potentially generally and as Kant uses them). They are ambiguous between (a) 'a particular kind of intuitive perception by means of which an object of thought is generated' and (b) 'a particular object of thought generated by a particular form of intuitive perception'. This ambiguity, together with two other types of influences might explain the fact that there is ambiguity in Kant's concept of the 'noumenon'. The first further influence would be Plato's writings about Ideas, writings that have a relationship with things as they really are and with an intellectual realm. The second influence would be Kant's aim to not make any overambitious claims about what we can know. I assume that because Kant supposes that there is some sort of necessary metaphysical link between purely intellectual forms of intuitive perception and the existence of things in themselves, while at the same time he has as a highest philosophical priority not to make overambitious philosophical claims about what we can know, a fuzzy concept of a noumenon arises in Kant's work; a fuzzy concept that, when used by Kant, *sometimes* is bound up with explicit Kantian claims and often implicitly seems to be bound up with the idea that there is a noumenal world of things as they are in themselves.

3.2 The *homo noumenon*

I think problematic aspects about the concept of a noumenon as they appear in Kant's theoretical philosophy are visible in those parts of his practical philosophy where Kant talks about the *homo noumenon*. For Kant the *homo noumenon* stands in contrast to the *homo phenomenon*. That could be so in three different senses that are all intelligible from a

perspective of exegesis. First, in the sense that they are two *perspectives* a practical agent can take towards himself (see e.g. G: 456-8; KpV: 114). Secondly, in the sense that the two terms reflect concepts that provide a different outsider *perspective* on what a human being is. Thirdly, in the sense that they refer to *two different aspects* of the human being.

As Kant regards the human being, 'from one perspective' and/or 'partly', he is/can regard himself as a '*homo phaenomenon*' (see MS: 293, 295, 418, 420, 423, 434). In some way the human being is/can be regarded as a physical being (MS: 430) that belongs to the natural world; a natural world where appearances are the most we can have knowledge of. To the 'homo phaenomon' Kant typically ascribes a capacity for 'Verstand', some form of intellect that is not the purest of pure and can be ascribed to a physical being. The 'homo noumenon' (for the term see MS: 239, 295, 335, 418, 423, 430, 434, 439; Ende: 334; Handschriftlicher Nachlass XXIII: 257) is a 'Vernunftwesen': a being of pure reason. Importantly, this 'Vernunftwesen' needs to be distinguished from a 'vernünftiges (Natur)Wesen' (see e.g. MS: 418). The human being is a 'homo noumenon' insofar he is a *completely* intellectual being (and as such he is unaffected by anything empirical). The creature Kant calls a 'vernünftiges (Natur)Wesen' is one that has empirical aspects, but makes efficacious use of his intellectual aspects. Note that 'Verstand' and 'Vernunft' as Kant uses them are things we could both refer to by the term 'reason' (although in translations, 'Verstand' is typically translated by 'understanding').

Let me now return to the concept of the 'homo noumenon'. In my opinion, there are similar ambiguities to this notion as used in his practical work as there are to the notion 'noumenon' in the first *Critique*. As Kant sees things, the human being who typically recognizes all of nature just through his sensibility, *recognizes* himself through mere apperception in actions and inner ends that cannot fall under the impressions of sense (see Eisler, 1930; entry 'Mensch'). For Kant the story does not seem to end with *the human being recognizing himself as such*. Kant seems to think that there actually *is* such an intra-human entity that is purely intellectual, although we human beings as members of the world of

sensibility cannot *know* its existence. For Kant the *homo noumenon* has an intrinsic capacity to be free (see e.g. MS: 418). The idea seems to be that there is a free God-like intellectual part in our soul that we are in touch with somehow and can exercise, although as members of the world of sense we cannot know that there is such a part of us, except *perhaps* – I tentatively suggest- *through the moral law*. At least, Kant gives the impression, it is *through the moral law we really know that we are free* (KpV: 4). Having said that, I would suggest that Kant wants to deny that thereby we can also be said to *know* ourselves as a *homo noumenon*. So for Kant we can *think of* ourselves as a *homo noumenon*. We even seem to ‘*recognize*’ ourselves as such through some sort of *supersensible experience* that although supersensible in some way is nevertheless part of us comprehensive human being with two intra-personal components. And in that supersensible experience we have encounters with the moral law through which we *know* that we are *free*. But this for Kant does not lead to the conclusion that we can *know* ourselves as a *homo noumenon*.

Issues about knowledge of the *homo noumenon* are epistemological issues. As to the metaphysical issue whether there *is* a purely intellectual self, this problem is one that is difficult in a similar way and to a similar degree as the question whether there are noumenal objects as things in themselves. Given the textual support for and the plausibility of the ‘perspective’ interpretation of the *homo noumenon*, there are reasons to deny that for Kant there is such a thing as a purely intellectual self. Having said that, there is also support that for Kant there is a purely intellectual self. A type of support that is on the boundary between direct and indirect support comes from Kant’s remark that there is a ‘real self’ (ein eigentliches Selbst) (see G: 457, 458, 461). Having said that, in his first *Critique* Kant explicitly denies that there is such a thing (KrV: 339). Related, in his first *Critique* Kant denies that there is a real ‘ich’ (a real ‘I’). Kant denies it in the sense that he advocates the view that the ‘I’ rather than a thing in itself is a ‘vehicle for all our thought’. Whether or not there is a real self for Kant, it seems beyond reasonable dispute that *if* Kant wants to make

the metaphysical claim that there is an essential self, this self for Kant is of an *intellectual* kind.

When I read Kant's practical work, I sense a substantial implicit commitment towards there being an essential self of an intellectual kind, an essential non-animalistic self that can set laws for himself. As such it seems to be metaphysically connected to a strong-willed autonomous self who can be in control over his inclinations. This interpretation of mine is hardly compatible with Kant's denial in the first *Critique* that there is a real self, but it also does not seem good to let Kant's remark in his theoretical work simply overrule all he implicitly and explicitly says in his practical work that points to the existence of an essential self. I think the best one can do is trying to render the two mildly compatible by recognizing the different contexts of his practical work and his theoretical work when he implicitly and explicitly talks about essential and real selves and 'I's.

My suggestion would be to see the term 'eigentlich' as in 'eigentliches Selbst' as having a somewhat different meaning in his theoretical and practical work. For his practical work I would prefer the term 'essential'. For his theoretical work, I would propose that the term 'eigentlich' is synonymous with the German term 'wirklich', a term probably best translated in English by 'genuine'. What we would end up with is the claim that in his theoretical work Kant denies that there is 'ein wirkliches Selbst'. The term 'wirklich' here then would designate that the self is not a thing in itself, no simple substance or power, but just some union of consciousness that must accompany all we can think of. A vehicle of all concepts that has merely a formal nature. A presupposition of all thought (Eisler, 1930: entry 'Ich'). Again, a full-fledged compatibility between my reading of Kant's practical work and Kant's remarks in his theoretical work seems impossible, but if my interpretation of Kant's work about essential selves in his practical philosophy needs to be upheld (for which there is explicit textual evidence), then possibly what I have proposed is amongst the best possible ways to render all Kant's says about eigentliches selbst and 'ich's' somewhat compatible.

4. On the relationship between morality and ‘the empirical’ in Kant’s work

According to popular opinion, Kant’s practical philosophy is devoid of any sympathy for the empirical. In some important sense this is true. In an essential way for Kant there is an *extremely strict division* between the ‘*a priori*’ and the ‘empirical’ in the sense that empirical moral principles are inferior to ones arrived at *a priori* (see esp. G: 388). Having said that, it is important to note that quite a bit of what Kant says about ‘the empirical’ in his practical philosophy does *not* reflect a Kantian view about the *inferiority* of the empirical to the pure.

Below in § 4.1, I comment first on the strict division between ‘the *a priori*’ and ‘the empirical’ as it permeates Kant’s work. In § 4.2, I explain a predominant way in which there is an important role for ‘the empirical’ in Kant’s practical philosophy. This way does not exhaust the empirical input in Kant’s practical philosophy. Having said that, with an understanding of this predominant bit in which the empirical plays a non-inferior role in Kant’s practical philosophy we get a long way as to a positive understanding of the non-inferior role of the empirical in Kant’s practical work. Importantly, what comes with this understanding is a non-exhaustive but sufficient bit of counterevidence to the view that Kant’s philosophy lacks a connection to daily life and the related view that for Kant as a moral philosopher only an empty *a priori* formalism mattered.

4.1. A strict division

Kant argued that only those sciences whose certainty is apodictic can be rightfully called real sciences (MAN: 468; see Louden, 2000: 20). By apodictic Kant means that their claims are “connected to consciousness of their necessity” (KrV: 54; Handschriftlicher Nachlass

XVIII: 290) and that their claims are arrived at by means of pure reason (see Prol: 275, 284)³. As Kant sees things, ethics and physics pass the litmus test of apodicticity: chemistry and biology (not to mention the biological sciences) all fail to pass this test (G: 388-9; MAN: 468, 471; *cf* KrV B: 152; A: 343; A: 848 / B: 876; cited in Louden, 2000: 20). According to Kant, apodictic certainty required by true science cannot be achieved via empirical principles.

Kant believes that moral discriminations must “have their source completely a priori in pure, but practical reason” (G: 408; see also 389, 452; KrV: B4, B: 124; A: 112: A547 / B575; cited in Louden, 2000: 20). Kant’s point is that a totally pure ethics would contain no empirical content whatsoever. As Kant sees things, in constructing scientific as well as ethical theories, there exists an “indispensable duty” to expound the pure part separately and entirely unmixed [*ganz unbemengt*] with the empirical part,” in order that the “apodeictic certainty sought by reason” can be achieved (MAN: 469; *cf* G: 389, see Louden, 2000: 10). The principle of action must be “free from all influence by contingent grounds, the only kind that experience can supply” (G: 426; see Louden, 2000: 10).

What I have explained above counts as a first motivation for Kant to argue for a sharp division between the *a priori* and ‘the empirical’. That is, Kant is interested in offering something that is completely necessary and certain. Granting that as a first motivation, there is a second motivation. Here I can only mention it in a nutshell. I will unpack it throughout this chapter. The second motivation for a distinction between ‘the *a priori*’ and ‘the empirical’ is Kant’s belief that a practical agent with some sort of core identity of Vernunft should be in control over his sentimentalist-animalistic nature.^{4 5}

³ Compare his expression in the *Groundwork*: “A categorical imperative, which declares an action to be objectively necessary in itself without reference to some purpose—that is, even without any further end—ranks as an *apodeictic* practical principle”(G: 415; *cf* e.g. B:63).

⁴ In his essay ‘A conjectural beginning of human life’ (‘*Mutmaßlicher Anfang der Menschengeschichte*’), Kant speculates about the very beginnings of human life. Kant suggests that a crucial turning point in human development occurred when our distant ancestors first became aware of their capacities to make free choices. At some point in the distant past the human being discovered in himself a faculty of choosing for himself a way of living and not being bound to a single one as other animals are. At this juncture the human being “stood, as it were, on the brink of an abyss; for instead of the single objects of desire to which instinct had up to now directed him, there opened up an infinity of them” (Anfang 8:112; cited in/see Louden, 2011: xxii). With this capacity the human being can create its

I take it to be important to be clear on how there is a difference between (i) the *a priori* as referring to Kant's *Formula of Universal Law* as a *highest principle of morality with just 'form'* and (ii) the *a priori* as obligatory (forbidden) or permissible substantive moral principles that pass the test of *FUL*. As explained, Kant wanted his highest principle of morality to be necessary and thereby free from all contingent elements. This thought is reflected by Kant's *Formula of Universal Law*. *FUL* states:

Act only in accordance with that maxim through which you can at the same time will that it become a universal law. (G-Paton: 88 / G: 421)

In the way presented, *FUL* is completely devoid of substantive (material, empirical) content. At this high level of abstraction of Kant's categorical imperative, a level that reflects Kant's outcomes of a quest for a necessary and maximally empirically free principle, no substantive/empirical/material content is allowed in by Kant.

In an important sense *FUL* is a (prospective) daily life test for action that determines which maxims are forbidden and permissible (see e.g. G: 402-3, 422; see also § 6 below). *FUL* is a test for which maxims can go through as practical principles and which ones cannot. *FUL* is a practical test for an agent in daily life faced with choices for prospective actions. A test that (i) *starts* with subjective principles for action, principles I as a practical agent want to act

own ends, will and character either in line or out of line with nature's destiny for the human species: self-realization through self-control by means of his capacity to be a 'Vernunftwesen' (being of reason). See also Kant's essay *Idee*.

⁵ Both motivations I have just ascribed to Kant can be found in the following passage from Kant's *Collins lectures* where he argues:

"Ethics can propound laws of morality that are lenient, and adjusted to the weakness of human nature. It can make itself comfortable to the human being, so that it demands of people only so much as they can perform. But on the other hand ethics can also be rigorous and demand the highest moral perfection. The moral law... must not be lenient and accommodate itself to human weakness; for it contains the norm of moral perfection. But the norm must be exact and rigorous – geometry, for example, lays down rules that are strict: it pays no heed to whether a human being can observe them in practice or not; the center point of a circle, for example, is too thick to be a mathematical point. Now since ethics also proposes rules, which are meant to be the guideline for our actions, *they must not be adjusted to human capacity, but have to show what is morally necessary*. An indulgent ethics is the ruin of moral perfection." (*Collins lectures*, 27:301; cited in Louden 2000: 8-9; my italics; cf KpV: 21)

on. A test that furthermore (ii) *consists in FUL* as a principle with just 'form'. And a test that (iii) *delivers* substantive practical principles for obligatory, (forbidden) and permissible actions. These principles are *free from contingent* content. That however does *not* seem to mean that *thereby* the 'pure' part of Kant's philosophy is 'not totally empty of content' or 'not merely formal' as Robert Louden claims below. There seems to be something wrong about Robert Louden's claim in italics in the quotation below. At best, this claim seems to be partly accurate. At worst this claim seems to be fully mistaken.

Although Kant occasionally seems to equate the pure part of moral philosophy with the form of knowledge and the empirical part with the matter of knowledge obtained by the senses (e.g. G: 400; KpV: 25-27)⁶, *it is a mistake to view the pure part as totally empty of content or "merely formal."* The pure part of ethics, on Kant's view, is not "completely separated from reality," and practical reason on his view does not involve "the complete abstraction from all content."...⁷ However, any information or content gleaned from the pure part of moral philosophy will always concern universal and necessary aspects (aspects which on Kant's view are more than merely human) of moral reality rather than particular and contingent ones. (Louden, 2000: 5)

I *agree* with Louden that there is a non-empirical part to Kant's ethics and that that is a part that "is not completely separated from reality". My claim would be –Louden will agree– that this part concerns moral principles that are necessary rather than contingent on one's motivational make-up. They also are universally binding –in a cosmopolitan way– across the entire range of beings with a capacity for reason.

Louden seems to assume that the 'pure' part of Kant's ethics is not necessarily empty. And he thereby presumably also wants to say that the '*a priori*' part of ethics is not

⁶I have adjusted this reference slightly to the referencing format I use. Works and page numbers are the same ones as used by Louden.

⁷ Footnote 8 omitted.

necessarily empty. Below I show why and how we should be rather reluctant to assume the former and very tentative in assuming the latter.

In contemporary moral philosophy discussions about ‘the *a priori*’ are typically either discussions about the *nature of justification for epistemic beliefs* or about *a priori knowledge*. Below I very briefly say something about Kant’s views on *a priori knowledge* as expressed in his theoretical work. I do so in order to require useful intellectual material for the question whether or not and if so how the substantive moral outcomes that pass the test of *FUL* can be said to be (a) *a priori* and/or (b) *pure*.

In the introduction to his first *Critique* Kant describes *a priori* knowledge as knowledge “absolutely independent of all experience” (K-S: 43/ B: 28). This Kantian understanding of the *a priori* seems rather narrow, both from the perspective of textual exegesis and independently. It looks like if Kant’s description were correct, all *a priori* knowledge would have to rest on innate ideas. That is a pretty strong position from an independent perspective and not quite what one wants from a perspective of Kant exegesis, either, for a variety of reasons. A more plausible suggestion is that *a priori* knowledge (and that will count for justification as well) must be independent of experience *beyond that needed to acquire the concepts required to understand the proposition at issue* (Russell, 2010: § 1).

One reason why from a perspective of Kant exegesis the above suggestion seems better than Kant’s claim that *a priori* knowledge is knowledge “absolutely independent of all experience” is the following one. Kant himself allows for empirically affected forms of *apriority*. Assuming that and if we express that claim with an appeal to the concept of ‘the pure’, then we can say that Kant wants to say that *a priori* knowledge is *not necessarily pure*. In the introduction of his first *Critique* Kant explicitly allows for empirically affected forms of a priority, while at the same time not allowing for ‘the pure’ to be mixed with anything empirical. Kant argues that “*a priori* modes of knowledge are entitled pure when there is no admixture of anything empirical” (K-S: 43 / B: 28). This remark pointing to a subclass of the

pure within the class of ‘the a priori’ explains itself and shows itself as relevant once Kant explains by example. Kant illustrates his abstract remarks here by saying that while ‘every alteration has its cause’ is an *a priori* proposition, it is not a pure proposition because alteration is a concept that can be derived only from experience (K-S: 43 / B: 28).

Useful explicit descriptions on the conceptual nature of ‘the *a priori*’ and ‘the pure’ in Kant’s *practical* philosophy seem to be absent. Therefore, on basis of the above remarks from Kant as they appear in his theoretical work, let us assume that in order for the substantive moral outcomes that pass the test of *FUL* and in order to count as *a priori* they must not depend on experience any more than the proposition ‘every alteration has its cause’ depends on experience. Now, it seems correct to say that substantive principles cannot be there as such, without them being subject to the test of *FUL*. There are no such things as independent substantive moral principles that have not been subject to *FUL* (and survived its test). Granting that, while going one step down the ladder, *unless* there are subjective principles for action –principles I want/intend to act on without having bothered about *FUL*- that should be tested, there is *no input* for *FUL*. That seems to imply that the outcomes of *FUL* somehow rely on subjective principles for action. Having said that, it sounds awkward to say that they rely on them. But in some sense they do, albeit clearly not in the sense that the subjective principles of action have any decisive influence on the outcomes of *FUL*. In any case, in order to arrive at a successful answer to the question whether or not the outcomes of *FUL* count as *a priori* one would seemingly need to take into account the nature of acquisition of the concepts of the subjective principles for action *plus* the reasoning procedure that is inherent to *FUL* (see § 6), because it is *FUL* that is crucial to the acquisition of pieces of moral *knowledge*. Here then we face Kant’s account about the intelligible nature of human beings that stands in some relationship with the human being in the empirical world. Since Kant leaves the relationship between the intelligible and natural nature of human beings underexplained, it is a tricky and hard thing to say whether the substantive principles passing *FUL*’s test count as *a priori*.

As to the *a priori* nature of *FUL*, there are further issues that require attention. As we will see in § 6, there are two variants of *FUL*: the *Contradiction in Conception test (CCT)* and the *Contradiction in the Will test (CWT)*. In his *Groundwork* Kant argues:

We must be able to will that a maxim of our action becomes a universal law: this is the canon of moral appraisal of action in general. Some actions are so constituted that their maxims cannot even be *thought* without contradiction as a universal law of nature...In the case of others that inner possibility is indeed not to be found, but it is still impossible to *will* that their maxim be raised to the universalizability of a law of nature because such a will would contradict itself. (G-Paton: 91 / G: 424)

When it comes to *CWT* and *from a perspective of textual exegesis*, there is a problem when as interpreters of Kant's texts we have to make a judgement about its *a priori* nature. This is tricky and hard given that Kant leaves the relationship between the intelligible and empirical nature of human beings underexplained.

Now for *CCT*. From an independent (non-exegetical) perspective *CCT* seems most *plausibly* considered as *a priori* if we take as defining the idea that the *a priori* must be independent of experience beyond that needed to acquire the concepts required to understand the proposition at issue (Russell, 2010). That is, this idea seems to be a better accurate reflection of reality of how knowledge can be *a priori* than Kant's own explicit claim that the *a priori* is "absolutely independent of all experience" (K-S: 43 / B: 28). Having said that, note that I assume Kant would have been sympathetic towards Russell's proposal (which I follow) too.⁸ Note also that I assume that in the (slightly more liberal) Russellian-

⁸ However, things become rather complicated once we connect his phrase to passages in Kant's work that indicate that (a) the intelligible world where there is freedom and where an intelligible self operates and (b) the concept of experience are *incompatible*. Generally, across his theoretical and practical philosophy, in Kant's work, the concept of and mental possibility for 'experience' applies to the empirical world, a world that is one in which the *homo noumenon* does *not* operate and in which freedom exists. And it very much looks like, for Kant 'experience' cannot apply to any other areas than the empirical world. One way to render Kant's remarks about the incompatibility between 'the noumenal' (in a broader sense) and 'the empirical world' somewhat compatible, would be to opt for

way proposed, *FUL's CCT* is a serious candidate, and probably the only one to counter Hume's claim that reason alone cannot generate moral discriminations. More about this in the next chapter.

Above I have addressed the question whether we can or cannot say that the substantive moral principles that pass *FUL's* test count as *a priori*. I take it that this question is relevantly different from the question whether or not we can say that the substantive moral principles that survive the test of *FUL* can be said to be 'pure'. In regard to this latter question, I would like to say the following. First, I have a modest preference for thinking that Kant's intention was only to refer to *FUL* –and not its substantive outcomes– as 'pure'. Kant searched for a highest moral principle, a principle free from all non-contingent content and thereby 'pure'. He found it in *FUL*.

For the remaining, the question whether or not we can say that the substantive moral principles that survive the test of *FUL* can be said to be 'pure' is one that can and should be answered from two different perspectives. A split should be made between (i) what is plausible from a perspective of Kant-exegesis and (ii) what is plausible from an independent (non-exegetical perspective). For both, the contradiction in conception variant of *FUL* seems to be 'pure' (see § 6 and chapter 5) (if we take the Russellian definition of the *a priori* I proposed above). From an exegetical perspective, complications arise as to *CWT* for reasons of an underdetermined relationship between the intelligible and natural self. From an independent (non-exegetical perspective), *CWT* seems to be too much dependent on experience to be 'pure' (and *a priori*).

the existence of 'supersensuous (extra-sensible) experience' as something applicable to the mental life of the intelligible aspect of the self.

4.3. 'Practical anthropology'

Kant's project in practical philosophy is one that explicitly seeks both to construct the foundational principles of theory from non-empirical sources and to bring in 'the empirical'⁹ for purposes of application to human life (Louden 2000: 11). In his *Groundwork* Kant argues that "moral philosophy canhave an empirical part, since ...[it] has to formulate its laws for the will of man so far as affected by nature-the ... laws being ... in accordance with which everything ought to happen ...". (G-Paton: 56 / G: 387). Kant then goes on to argue that "the empirical part [of moral philosophy] might be called specifically *practical anthropology*, while the rational part might properly be called *morals* (G-Paton: 56 / G: 388). At several places in his work Kant gives descriptions of what 'practical anthropology' (or 'moral anthropology', terms which Kant uses interchangeably)¹⁰ amounts to. In his *Metaphysics of Morals* Kant states:

The counterpart [*das Gegenstück*] of a metaphysics of morals, the other member of the division of practical philosophy as a whole, would be moral anthropology, which, however, would deal only with the subjective conditions in human nature that hinder human beings or help them in *the carrying out [die Ausfuhrung]* of the laws of the first part [namely, the metaphysics of morals]. It would deal with the development, spreading, and strengthening of moral principles (in education in schools and in popular instruction), and with similar teachings and precepts based on experience" [*auf Erfahrung grundende Lehren und Vorschriften*]. (MS : 217; cited in Louden, 2000:14)

⁹ Louden uses the term 'empirical content' throughout his work. I prefer to stick to the broadest possible intuition of the reader by using the term 'the empirical' since there is a variety of ways in which the empirical enters in Kant's philosophy.

¹⁰ Kant uses a variety of labels when he refers to the empirical. Besides 'practical anthropology', Kant also talks about 'moral anthropology', just 'anthropology', 'empirical psychology', the 'second part' of ethics, 'empirical ethics', 'applied moral philosophy' (Louden, 2000: 13, *cf* 27).

As Robert Louden makes clear, similar descriptions occur throughout his various lectures on ethics. In the *Praktische Philosophie Powalski* lectures, for instance, Kant argues:

One must not merely study the object (that is, moral conduct), but also the subject (that is, the human being). This is necessary because one must see what sorts of hindrances to virtue are present in the human being. The first part of ethics contains the criteria of discrimination of that which is practically good and evil. . . . The second [part of ethics] contains the rules and means of execution—the means by which it is possible for a will to act according to rules. This second part is the most difficult, because one must study the human being. (*Powalski* lectures: 97-98; cited in Louden, 2000: 14)

And in the *Moralphilosophie Collins* lectures we are told that consideration of rules "is useless [*unnutz*] if one cannot make human beings willing to follow them" (27: 244). If one chooses unwisely to pursue practical philosophy "without anthropology, or without knowledge of the subject, then it is merely speculative, or an idea; the human being must therefore at least be studied later on [*hernach*]" (*Collins* lectures: 244; cf. *Moral Mrongovius*: 1398; cited in Louden 2000: 14).

In summary, one can say that there is an essential empirical component to Kant's practical philosophy Kant calls 'practical anthropology'. This component consists in (a) *information* necessary and (b) *methods* useful for *achieving implementation of action* in concordance with the highest principle of morality in the lives of practical agents. An essential empirical component that by Kant is *not* considered as something inferior to the pure, but rather as something complementary to it.

5. 'Vernunft'

5.1. Introductory comments

In this section I focus on Kant's concept of reason: 'Vernunft'. Before I focus on Kant's conception of 'Vernunft' however, I first say something on the topic from a wider (i.e. not specifically Kantian) perspective.

According to the frequently used German-English/English-German *Collins dictionary* 'Verstand' is the right German term for the English term/concept 'reason' understood as a mental *faculty*. It furthermore tells us that 'reason' as *common sense* is what the German term 'Vernunft' reflects. It also tells us that 'to reason' as a verb meaning 'to think logically' is represented by the term 'vernünftig oder logisch denken'. And finally it tells us that reason as an ability to think logically can be represented by the term 'logisches Denkvermögen'.

Overall, it seems that across German dictionaries we find different senses of 'Vernunft'. That leaves us with a variety of questions. What exactly does the concept 'Vernunft' mean? How many meanings actually are there if more than one? In what way does the concept include a normative aspect? Is this exclusively an aspect having to do with getting things intellectually right or is there also a normative component in the sense of there being a requirement to act in line with what intellectual reason judges? In what way is it a psychological-intellectual concept in the sense of it being a capacity? Note that these questions/concerns about the contemporary concept map well onto Kant's philosophy about which more now.¹¹

¹¹ So, from a contemporary perspective we will not have an easy time with the term and concept 'Vernunft', including when one wants to use it as some sort of means to get a grip on Kant's own concept of 'Vernunft'. If one adds to these troubles certain etymological considerations about the German concept 'Vernunft', things become even more complex. That is, both for the concept 'Vernunft' in its contemporary German meaning and for Kant's conception. As to both, but especially the latter, amongst other things since the concept *might* have been significantly under development when Kant wrote his philosophy. As to 'might', I do not have the evidence that this indeed was the case, but it seems a fact that the rather closely related concept 'Verstand' *was* under development then (see Drosdowski & Greber, 1963). Even if the concepts 'Vernunft' (and Verstand) were under development then, this *could* but *need not* mean that this developments somehow had an impact on Kant's work.

My investigations in the form of an exploration of both frequently used 'normal' German dictionaries and etymological dictionaries as well as correspondence with an expert on Kant and German etymology have shown the

5.2 Kant's concept of 'Vernunft'

Kant typically speaks about 'Vernunft' as a 'Vermögen' (see e.g. B: 237; KU: 167; KpV: 119). Kant used the term *Vermögen* in an 18th century Latin meaning of '*facultas*'.¹² Because it is the case that the Latin term *facultas* grounds Kant's concept, the English translation 'faculty' for Kant's term 'Vernunft' can sensibly be used. From a neutral (i.e. Kant-independent) semantical perspective, the term 'faculty' could reasonably be used for some kind of overarching mental or soul-like thing (with certain functional capacities/powers). That fits well with a possible interpretation of Hume's concept of reason as explained in chapter 2. In this sense, the translation 'faculty' fits also quite well with one Kantian use of the term: Vernunft as a power of the soul. (For translations of 'Vermögen' into the term faculty see e.g. K-S: 301 / B:237 and LWB: 223 / KpV: 119).

One could also quite sensibly translate Kant's term 'Vermögen' by the term 'capacity'. And it is indeed done in the literature (see e.g. KU-GM: 55). From a perspective of Kant exegesis, 'capacity' as a translation of 'Vermögen' is nevertheless something tricky because from an *independent* semantical perspective the term 'capacity' seems to be ambiguous between *predispositions* (nature has given us) and *actualized manifestations of them* (that can fail to manifest themselves over an entire human life span).¹³ Both seem to play a role in

following results as to the development of the concept 'Vernunft': 'Vernunft' is a noun derived from the verb 'vernehmen'. Initially the meaning of 'Vernunft' was 'wahrnehmen'/erfassen. Importantly, this initial meaning seem to have been purely descriptive. By the time of the influential German philosopher Christian Wolf the term Vernunft had normative meaning. Taking into account information from all non-etymological and etymological dictionaries, the assumption seemingly has to be that over time, one or more normative element have sunk into the concept and the intellectual aspects seem to have changed., 'Vernunft' seems to have become exclusively or else partly an *appropriate* (or 'healthy'). functioning/exercise of an *intellectual 'faculty' or 'capacity'*. Apart from that, it may also have a normative *action* component in the sense of there being a requirement of 'congruence in line with what some intellectual Vernunft prescribes'. The reason why this is important is because there is some evidence, albeit limited, to think that for Kant such an aspect was part of the concept 'Vernunft'. What speaks in favour of that is his work as a whole, but also the following somewhat ambiguous explicit claim he makes in the introduction of his *Critique of Practical reason*. There Kant argues: "For if pure reason is actually practical, it will show its reality and that of its concepts in actions, and all disputations which aim to prove its impossibility will be in vain" (CpR-LWB: 118 / KpV: 3).

¹² I am grateful to Michael Franz for this information.

¹³ Thanks to John Fischer here.

Kant's work and Kant seems to stress (more specific variations of these) two different things in different parts of his theoretical and practical work. Note that one can see both meanings of 'capacity' I have just given reflected in Kant's idea of the human species having pre-dispositions (*Anlagen*), things that are present in an *animal rationabile* who can develop himself by his own powers into an *animal rationale*: a being who uses his faculty of intellect optimally by acting in line with the right principles (see Antr: 321; MS: 434).

There is more to be said on Kant's notion of 'Vermögen' as translated by the English word 'capacity'. Capacities are often thought of as things that are *quantitatively extendable*. Kant's notion of Vermögen must *not* be understood as such. Kant's notion of Vernunft as a 'Vermögen' is *not* something that is *quantitatively extendable*. It is an all or nothing, black or white affair and does *not* come in degrees.¹⁴ This would both *reasonably* and *at most* mean that the *use* of Vernunft (*usus rationis*) comes in degrees.

I have just argued that the 18th century Latin meaning of the term 'Vernunft' seems to require that Vernunft is an all or nothing, black or white affair and *does not come in degrees*. Below I present my view on Vernunft being (and not being) a black and white affair in Kant's work. I cannot do that however before having explicitly mentioned the fact that besides a component of the soul there is a *normative* component to Kant's conception of 'Vernunft'.

My preferred interpretation of Kant's concept of Vernunft is a rather comprehensive one and looks like an extensive version of a description Kant gives in his essay 'Idee'. There he presents a description of the concept that covers a variety of uses throughout his work that I include in my description. In 'Idee' Kant argues that "Vernunft in a creature is a 'Vermögen' to extend/further (*erweitern*) the rules and purposes of the use of all his powers far beyond the instincts of nature and does not know any boundaries of her plans (*Entwürfe*)" (Idee: 18-9).

¹⁴ Thanks to Michael Franz here.

The first thing to say as to my preferred interpretation of Kant's concept of Vernunft is that I *do* believe that we *must not* say that Vernunft as a 'Vermögen'= 'capacity' is quantitatively extendable. But from here onwards I hold a very specific view about Vernunft being and not being something black or white, *and not grey*. I would suggest that for Kant Vernunft is a capacity of the soul ('Vermögen der Seele') that reflects *a potentiality the boundaries of which are fixed* (fulfills the 'not quantitatively extendable requirement'). Next then, I would argue that one can actualize that capacity of the soul to a greater or lesser extent. I would then hold that Vernunft is intrinsically connected to its 'potential objects of acquired knowledge' meaning potential objects that reflect an efficacious use of the power of Vernunft as a potentiality of the soul.

Before I move on with a discussion of Kant's *FUL* let me make one further remark about Kant's views about 'Vernunft'.¹⁵ According to Kant, practical Vernunft has primacy over its theoretical brother. By primacy between two or more things connected by reason, Kant understands the prerogative of one by virtue of which it is the prime ground of determination of the combination with all the others. Kant notes that in a narrower practical sense 'primacy' refers to the prerogative of the interest of one so far as the interest of the others is subordinated to it and it is not itself inferior to any other (KpV: 119). According to Kant, if practical reason (Vernunft) may not assume and think as given anything further than what speculative reason (Vernunft) affords from its own insight, the latter has primacy. But assuming that the former has of itself original *a priori* principles with which certain theoretical positions are inseparably bound but which are beyond any possible insight of the speculative reason (although not contradictory to it), then according to Kant we need to ask ourselves which interest is superior: practical or theoretical? (see KpV:120) And then it turns out to be the case that the answer is: 'practical'. Why? Because "every interest is ultimately

¹⁵ There is more that can be said about Kant's concept of Vernunft. It is beyond the purposes of this chapter to provide a fuller account and a textually more thoroughly supported suggestion of what the concept of Vernunft amounts to for Kant. That would really require a study in itself.

practical, even that of speculative reason (Vernunft) being only conditional and reaching perfection only in practical use” (CprR-LWB: 225 / KpV: 121).

6. The *Formula of Universal Law*

Kant’s doctrine of *a priori* knowledge rests mainly on the assumption that mind- or reason, as he calls it-functions actively in accordance with principles which it can know and understand. He holds that such rational principles can be manifested, not only in thinking as such (which is studied in logic), but also in scientific knowledge and moral action. We can separate out these rational principles, and we can understand how they are necessary for any rational being so far as he seeks to think rationally about the world and to act rationally in the world. (Paton, 1948: 14)

I take this to be a helpful description as to how to understand the role of something psycho-intellectual and something normative in Kant’s practical philosophy. One thing that characterizes this passage is that it does *not* contain any big mysteries that somehow have a relationship with the noumenal. On a related note, the passage reads as perfectly compatible with daily life.

As to the connection between Kant’s moral philosophy and daily life, regularly Kant connects his moral philosophy to the morality of the ordinary person. At several places in his work he expresses the idea that the task of the philosopher is to set out clearly, correctly and precisely what already is “inherent in the structure of every man’s reason” (MM: 376; see also A: 730 / B:758, A: 807 / B: 835, A:831 / B:859; G: /389, 8/397, 17/402, 20-24/403-5, 34/411; KpV 8n, 30, 36, 87, 91-2, 126; MM 206; mostly cited in: Sullivan, 1989: 4-5).¹⁶ Kant believed that everyone must have a fundamentally correct pre-philosophical understanding of morality, even if that understanding lacks clarity and adequate expression (Sullivan, 1989: 4).

¹⁶ Kant also wrote: “Who would want to introduce a new principle of morality and, as it were, be its inventor, as if the world had hitherto been ignorant of what duty is or had been thoroughly wrong about it?” (KpV: 8n, cited in: Sullivan, 1989: 5).

Generally, whatever Kant argues about freedom and morality seems to have a connection to ordinary moral life. Having said that, in Kant's work these views about the relationship between Kant's moral principles and views about freedom on the one hand and ordinary life on the other hand get clouded by and seem to be in some sort of tension with Kant's underexplained and sometimes rather mysterious and awkward remarks about noumenal selves. In § 8, I put forward a proposal as to how to render the noumenal self and ordinary life aspect of Kant's practical philosophy at least somewhat compatible.

Below I investigate the relationship between *FUL* and something psycho-intellectual that characterizes 'Vernunft'. As a starting point I take the connection of Kant's morality to ordinary life that is clearly visible in Kant's practical philosophy, especially his *Groundwork*. But I do this while suggesting that we keep in the back of our minds Kant's rather puzzling remarks about the *homo noumenon* as explained in section 3 and about which more in the next two sections.

Kant's 'highest principle of morality' in fact turns out to be different principles or at least one principle that comes in different guises. *FUL* is one (guise) of Kant's highest principle(s) of morality. For reasons of purpose and space, in this chapter, I focus almost exclusively on this version (I point to some others in the next chapter). I assume it is *FUL* that is most helpful in connection to the aim of understanding the limits of the power of moral objectivity. Let us recall what *FUL* states:

Act only in accordance with that maxim through which you can at the same time will that it become a universal law (G-Paton: 88 / G: 421).

As indicated earlier, there are two applications of *FUL*. The *Contradiction in Conception*-test (CCT) and the *Contradiction in the Will*-test (CWT). In his *Groundwork* Kant argues:

We must be able to will that a maxim of our action becomes a universal law: this is the canon of moral appraisal of action in general. Some actions are so constituted that their maxims cannot

even be *thought* without contradiction as a universal law of nature...In the case of others that inner possibility is indeed not to be found, but it is still impossible to *will* that their maxim be raised to the universalizability of a law of nature because such a will would contradict itself. (G-Paton: 91 / G: 424)

It is *CCTI* will spell out here, because it is *CCT* that is first and foremost important with an eye on Hume's philosophy and the disagreement between him and Kant as to the involvement of passions in moral discriminations. In order to understand *CCT* let us consider the following helpful remarks from Onora O'Neill:

A maxim of promising falsely commits an agent to supporting means to promising falsely, hence to maintaining enough public trust for promises to gain acceptance. But willing false promising 'as a universal law' (*per impossible*) commits an agent to willing the consequence of universal false promising, which include the destruction of trust, hence it is incompatible with willing any reliable means to false promising-for oneself or for others. (O'Neill, 2004: 99)

So this is why Kant thinks that practical agents cannot will false promising.

As we can see from the explanation of O'Neill, a reasoning procedure is inherent to *FUL*. In a spelled-out version of *FUL*, it is part of *FUL*'s metaphysical nature. Assuming that, the question is whether this reasoning procedure must be necessarily seen as a theoretical one that can be disconnected from the supersensuous nature and/or whether it has necessarily a relationship with the minds of intelligible agents.

Formalizing O'Neill's explanation of *FUL* (as a test for *prospective* action), *FUL* can be constructed as a practical syllogism (that in a deformed way takes place in the mind of practical agents).

- *Major Premise*: Act only in accordance with that maxim through which you can at the same time will that it become a universal law" (*FUL*) (G: 421)

- *Minor Premise:* A subjective principle (maxim) reflecting an intended action/an action under consideration; for example, ‘I want to break a promise’.

(Implied reasoning process: if everyone did that, there would be no background of trust which is necessary for promises to be broken [recall the explanation from O’Neill above])

- *Conclusion:* I should not break a promise¹⁷

Generally, what should be concluded about *FUL* is that the theoretical reasoning procedure implied by *FUL* as a test for the moral acceptability of maxims is also the epistemological means in practical agents for acknowledgement of *FUL*’s judgemental outcomes. And most probably *FUL* is also a constructive constitutive means in practical agents for rightful judgements about obligatory, permissive and forbidden actions. Making a reference to the strong connection between Kant’s moral philosophy and ordinary life, it seems as if we have to say that human practical agents can follow in their thought the reasoning procedure that is implied by *FUL* when it tests for the moral acceptability of certain subjective principles of actions an agent has. And this capacity practical agents have, has an epistemological character, but seemingly also a constructive constitutive one in terms of the identification of moral objectivity.

¹⁷ In the following passage Kant paradigmatically presents an example of a deformed version of a practical syllogism. Although deformed, it still emphasizes the abstract reasoning process constitutive of *FUL*. Kant argues:

“What form of a maxim makes it suitable for universal law-giving and what form does not do so can be distinguished without instruction by the most common understanding. I have, for example, made it my maxim to increase my property by every safe means. Now I have in my possession a deposit, the owner of which has died without leaving any record of it. Naturally, this case falls under my maxim. Now I want to know whether this maxim can hold as a universal practical law. I apply it, therefore, to the present case and ask if it could take the form of a law, and consequently whether I could, by my maxim, make the law that every man is allowed to deny that a deposit has been made when no one can prove the contrary. I immediately realize that taking such a principle as a law would annihilate itself, because its result would be that no one would make a deposit.” (CprR-LWB: 138-9 / KpV: 27)

7. *FUL*, Kant and the embodied nature of the rational judge

While Kant is very much a daily life/commonsensical philosopher in the ways mentioned, there is this aspect of his philosophy that is rather peculiar: the split between ‘*the homo phaenomenon*’ as part of the empirical world and the ‘*homo noumenon*’ as part of the intelligible world. Insofar as human beings are a part of nature, they have empirical characters, "the causality of which must stand under empirical laws" (KrV: A: 546 / B: 574). And in respect to this empirical character there is "no freedom" (A: 550 / B: 578) (cited in/see Louden, 2000: 17). On the other hand, as rational beings who have the power to determine their own actions through efforts of will, human beings also follow "a rule and order altogether different from the order of nature" (A: 550 / B: 578). In respect to this intelligible character the agent "stands under no conditions of time, for time is only the condition of appearances, not of things in themselves" (A: 539 / B: 567) (cited in/see Louden 2000: 17). And when Kant says no conditions of time, he really means no conditions of time *at all*. Kant even argues:

To look for the temporal origin of free acts as such (as though they were natural effects) is ... a contradiction: hence it is also a contradiction to seek the temporal origin of the human being's moral character . . . since this character signifies the ground of the *exercise* of freedom; which (like the determining ground of the free will generally) must be sought in representations of reason alone. (Rel: 40; cited in Louden, 2000: 17)

On basis of his theoretical philosophy we have to conclude that for Kant there is some sort of hard gap between the sensible world where a natural being with ‘Verstand’ and ‘Vernunft’ operates (on the one hand) and the supersensible world (on the other hand). That is, in the sense that ‘Verstand’ and theoretical reason can never have knowledge of noumenal objects. Or at least, on basis of the nature of ‘Verstand’ and theoretical reason (Vernunft) one cannot make rightful claims about knowledge of objects beyond those claims that concern

appearances (recall the problems about the concept 'noumenon' as explained in § 3). Anticipating some comments in the next paragraph, it furthermore is the case that neither 'Verstand' nor anything else that is part of the natural world, can influence the existence of noumenal objects. This is another way in which there is some sort of gulf between the world of sense and the noumenal world.

For Kant, the supersensible world is the world where a practical sense of freedom operates. In connection with this idea of the existence of freedom in the noumenal world, Kant says some striking things about the influence of the supersensible world on the natural world. He argues:

Now although there is an incalculable gulf fixed between the domain of the concept of nature, as the sensible, and the domain of the concept of freedom, as the supersensible, so that from the former to the latter (thus by means of the theoretical use of reason) no transition is possible, just as if there were so many different worlds, the first of which can have no influence on the second; yet the latter should have an influence on the former, namely the concept of freedom should make the end that is imposed by its laws real in the sensible world; and nature must consequently also be able to be conceived in such a way that the lawfulness of its form is at least in agreement with the possibility of the ends that are to be realized in it in accordance with the laws of freedom. –Thus there must still be a ground of the unity of the supersensible that grounds nature with that which the concept of freedom contains practically, the concept of which, even it does not suffice for cognition of it either theoretically or practically, and thus has no proper domain of its own, nevertheless makes possible the transition from the manner of thinking in accordance with the principles of the one to that in accordance with the principles of the other. (CPJ-GM: 63 / KU: 176; cfMS: 418)

Kant argues that the intelligible world *should* causally affect the world of sense. Having said that, Kant does not tell us how that would work. This is a problematic aspect of Kant's philosophy. Be that as it is, it is not the case that Kant without qualification 'just states

rather remarkable/awkward views on the relationship between morality, freedom and the empirical world'. Below I explain how to understand this.

I have already argued that for Kant there is an important connection between his moral philosophy and daily life. As mentioned, this comes very much in the guise of Kant believing that *FUL* is accessible to the ordinary practical agent. Indeed, *FUL* seems to permeate the lives of ordinary people. Now, when it comes to Kant's views about morality, freedom, intelligible and natural selves, there is a particular aspect to our daily lives – or perhaps it is better called a commonsensical aspect – to Kant's philosophy that must be noted. That is, this aspect must be noted *while* it can be granted that Kant has underexplained, rather mysterious if not also awkward views about the intelligible and the natural world. The point is this. While stating rather puzzling things about intelligible selves and while not explaining how the intelligible world can have an influence on the natural world, Kant believed that there are certain aspects to a human being which make it possible for him to be affected by the principles of duty. Furthermore, this is something that his intelligible self recognizes and that might be part of the intelligible world in some way, a mind-*independent* way. Said slightly differently, Kant was aware that there are aspects to a human being due to which it is possible for him to be affected by *legislations* 'coming from the noumenal world'. Note that the interesting thing about the term 'legislation' here is that it potentially fits something mind-*independent*, also in Kant's work, but that for Kant it often if not always signifies a law that is determined by the noumenal self of some sort of comprehensive creature called human being. I come back to that in the next section. For now, I mention that Kant calls these aspects (because of which it is possible for a practical agent to be affected by legislations coming from the noumenal world) (in a natural mind-dependent and/or extra-natural mind-dependent way) 'subjective conditions of susceptibility for the notion of duty'. In his *MM* Kant argues:

These are such moral qualities as, when a man does not possess them, he is not bound to acquire them. They are: the moral feeling, conscience, love of one's neighbor, and respect for ourselves (self-esteem). There is no obligation to have these, since they are subjective conditions of susceptibility for the notion of duty, not objective conditions of morality. They are all sensitive and antecedent, but natural capacities of mind (*praedispositio*) to be affected by notions of duty; capacities which it cannot be regarded as a duty to have, but which every man has, and by virtue of which he can be brought about under obligation. Consciousness of them is not of an empirical origin, but can only follow on that of a moral law, as an effect of the same on the mind....(MM-Gregor: 59 / MM: 399)

Because of thoughts of the kind expressed in the passage above, there is some reason to think that Kant is a *weak naturalist*. Someone who endorses the idea that "empirical facts about human nature, though they cannot in themselves establish or justify normative moral principles, also cannot contradict such principles" (Louden, 2000: 8).

Having remarked on Kant's reference to conditions for receptivity of duty and having said that therefore there is *some* reason to regard Kant as a naturalist, Kant himself does *not* categorize the receptivity conditions as naturalistic. He does not qualify them as unnaturalistic either however. What he does state is that *consciousness* of these conditions is not of an empirical origin, but that does not seem to tell us anything about the nature of these conditions themselves. In short, he leaves their nature unclear and that is unfortunate. Nevertheless, he seems to have been aware that it is not the case that some principle from the intelligible world can be acknowledged by an agent and move an agent without there being something in his psychology that makes this possible.

8. A comprehensive understanding of *FUL*

In this section I build to a comprehensive understanding of *FUL*. We have seen that *FUL* states: Act only in accordance with that maxim through which you can at the same time will that it become a universal law (G: 421). In the upcoming §§ 8.1-8.5, I discuss several elements that are meant to enlighten the nature of *FUL* in a holistic comprehensive way. That is, as a moral principle that has a positive relationship with the idea of ‘necessity’; with something psycho-intellectual and with freedom to judge and act as well as a two-fold antagonistic relationship with ‘the empirical’.

In § 8.1, I discuss the element of universality. In § 8.2, I discuss Kant’s attitude of ‘the further away from the empirical, the more lofty’. In § 8.3, I focus on the relationship between ‘Vernunft’ and logic. In § 8.4, I discuss the relationship between ‘moral *laws*, practical agents and independence from the empirical’. In § 8.5, I discuss Kant’s thoughts about there being a power of the soul that can be a master over one’s sentiments. I start every section by one or more quotations from Kant’s work characteristic of the element that I then discuss.

8.1. ‘Universality’

Experiences teaches us that a thing is so and so, but not that it cannot be otherwise (K-S: 43 / B: 28; *cf*B: 30/ K-S: 45)

Necessity and strict [generality]¹⁸ are thus sure criteria of a priori knowledge, and are inseparable from one another (K-S: 44 / B: 29)

¹⁸ The original translation by Norman Kemp-Smith says ‘universality’. I address the reasons for substitution directly in the main text of this section.

These two quotations reflect claims Kant makes in the introduction of his first *Critique*. In order to accurately reflect what Kant meant by ‘universal’ in his first *Critique* we should resort to two German terms: ‘Notwendigkeit’ (necessity) and ‘strenge Allgemeinheit’ (strict generality). Kant appeals to both as necessary conditions and essential characteristics of synthetic *a priori* judgements. Kant claims that ‘Notwendigkeit’ and ‘strenge Allgemeinheit’ somehow cannot be metaphysically separated, but nevertheless he seems to think they are conceptually different. Both notions are a response to a concern Kant has about an inductive form of generality as it figures in Hume’s philosophy. It is a response in the sense that for Kant such a form of generality can never give us knowledge because it cannot appeal to more than possible experiences and that is too little for Kant. It doesn’t meet Kant’s criterion for knowledge. Rather than inductive generality, there must be some sort of higher criterion, in order for our scientific judgements to count as knowledge. This criterion is that they must be able to count as knowledge without being dependent on possible experience. If our scientific judgements do, they are both necessary and meet the criterion of ‘strenge Allgemeinheit’ (strict generality). And as such they function as a requirement for what Kant is after: certain knowledge. Note that for Kant this knowledge necessarily comes in the form of synthetic judgements. Analytic judgements do not pass the ‘candidate for a *priori* knowledge test’ because they do not add something new to what we already have when operating with a certain concept).

In the practical area, Kant sometimes points to reasons for it being important that practical rules are necessary in a way that resembles his talk about necessity (rather than some weaker form of generality) in the first *Critique*. In *KpV* Kant argues:

It is certainly undeniable that every volition must have an object and therefore a material; but the material cannot be supposed for this reason to be the determining cause of the choice, the dependence of the faculty of desire on the existence of some thing would have to be made basic to volition, and this dependence would have to be thought out in empirical conditions and

therefore never could be a foundation of a necessary and universal rule. (CprR-LWB: 145 / KpV: 32; *cf* CprR-LWB: 140, 147-8)

In passages of the above type that occur in his practical work, Kant approaches *FUL* from the perspective of the importance of necessity as a maximally hard form of generality.

8.2. The further away from the empirical, the more lofty

As indicated, when we get to the realm of freedom in Kant's practical philosophy, there is this puzzling mysterious talk about intelligible selves operating in a supersensible realm. The reason we might get this puzzling talk from Kant might be well reflected by the following passage from the introduction of his first *Critique*. Kant argues:

It is precisely...in a realm beyond the world of senses, where experience can yield neither guidance nor correction, that our reason carries on those enquiries which owing to their importance we consider to be far more excellent [vorzüglich], and in their purpose far more lofty [erhaben], than all that the understanding can learn in the field of experience. (K-S: 45-46, A: 18; *cf*e.g. A: 250, B: 30-1)

It is not entirely clear to me whether this is a statement that carries any normative force. It seems primarily, if not exclusively descriptive. Having said that, on basis of his work as a whole, it seems rather beyond dispute that –besides a clear attachment to daily life- there some sort of thought in Kant that 'the further away from the empirical, the more excellent/lofty' things are.

8.3 Reason in logic

In the earliest times to which the history of human reason extends, mathematics, among that wonderful people, the Greeks, had already entered upon the sure path of science. But it must not be supposed that it was as easy for mathematics as it was for logic – in which reason has to deal with itself alone – to light upon, or rather to construct for itself, that royal road. (KS-19 / B: 9; *cf*B: 15)

That logic should have been thus successful is an advantage which it owes entirely to its limitations, whereby it is justified in abstracting – indeed, it is under obligation to do so- from all objects of knowledge and their differences, leaving the understanding nothing to deal with save itself and its form. (K-S: 18/B: 8; *cf*A: 10)

The very foundation of all laws of reason seems to be the law of non-contradiction. It is this logical law that is core to *FUL*. What we have then in regard to the nature of *FUL* is an intellectual reasoning process focused on consistency in thought via the law of contradiction. What more could one go for when being an author as Kant was who tried to get as far as possible away from the empirical while somehow also trying to keep a connection to ordinary people?

8.4. Moral laws, moral agents and independence from the empirical

8.4.1. The supersensuous nature of rational beings

[T]he moral law...does provide a fact absolutely inexplicable from any data of the world of sense or from the whole compass of the theoretical use of reason, and this fact points to a pure intelligible world- indeed, it defines it positively and enables us to know something of it, namely, a law. (CprR-LWB: 153 / KpV: 43)

The supersensuous nature of [rational] beings...is their existence according to laws which are independent of all empirical conditions and which therefore belong to the autonomy of pure reason. And since the laws, according to which the existence of things depends on cognition, are practical, supersensuous nature, so far as we can form a conception of it, is nothing else than nature under the autonomy of the¹⁹ pure practical reason. The law of this autonomy is the moral law, and it, therefore, is the fundamental law of supersensuous nature and of a pure world of the understanding.... (CprR: 153-4 / KpV: 43-4; my italics)

Since the material of the practical law, i.e. an object of the maxim, cannot be given empirically, and since a free will must be independent of all empirical conditions (i.e., those belonging to the world of sense) and yet be determinable, a free will must find its ground of determination in the law, but independently of the law. But besides the latter, there is nothing in a law except the legislative form. Therefore, the legislative form, in so far as it is contained in the maxim, is the only thing which can constitute a determining ground of the free will...Thus freedom and unconditional practical law reciprocally imply each other. (CprR-LWB: 140 / KpV: 29)

As we already knew and as becomes explicitly clear from the second quotation above, as intelligible (or supersensuous) beings, human beings have a nature that is free from all

¹⁹ I doubt that the compound term 'the pure practical reason' here is a correct translation of the term 'Vernunft' in this sentence. I think the article 'the' should have been left out. Compatible with being generally very happy with Lewis White Beck's translations of *KpV*, I am in a similar way concerned about some further translations of the German term 'Vernunft' in White-Beck's translation of *KpV*. (pages: 167, 172, 176, 180, 187, 209, 219, 222). If the translations are indeed suboptimal in terms of inappropriate copying of articles where they should be left out, a recognition of this may also be somehow important for creating a good meta-ethical practice as to the use and reflection on the multi-interpretable term 'reason'. Removing inappropriate copying of articles from material published in English may help Kant's own ideas about Vernunft to be better communicated. This could have a beneficial effect on meta-ethical practice. Either on its own or else in combination with the fact that it rules out any associations with concepts of 'reason' that need an article. Here I am thinking about normative/justifying/Scanlon-type of reasons. Reasons that broadly speaking can be referred to as 'things that speak in favour of something else'. A third 'in between alternative' in which a removal of suboptimal translations characterized by inappropriate copying of articles, could help meta-ethical practice, would be the following one. A correct translation could prevent some sort of *ambiguous* English sentence, where 'reason' can be read in a variety of ways that resemble some of the seven types I distinguished in chapter 3. If, in such circumstances, it can be prevented that the reader chooses an interpretation that is sensible from a neutral perspective, while not being sensible from an exegetical perspective, a conflation of 'reason' as normative authoritative institution (conception 1 as mentioned in chapter 3), 'reason' as something intellectual (conception 2) and justificatory, normative (different from conception 1)/Scanlonian types of reasons *at least* will not be *encouraged*.. Thinking that it actively *discourages* seems to me wishful thinking rather than a realistic thought.

empirical conditions. In this second quotation Kant says that as intelligible beings, humans have an existence according to laws which are independent of all empirical conditions. This is not a particularly clear remark, to say the least. And in combination with the first and third quotation of this subsection and his practical philosophy at large, as a reader one faces ambiguities and other unclarities as to the conceptual nature of ‘laws’ in Kant’s work and any relationship between mind-dependent and independent laws if both are part of Kant’s work.

8.5. *Ratio subiectivae sumta* as a master over one’s sentiments

It is nothing else than personality, i.e. the freedom and independence from the mechanism of nature regarded as a capacity of a being which is subject to special laws (pure practical laws given by its own reason), so that the person as belonging to the world of sense is subject to his own personality in so far as he belongs to the intelligible world. (CprR: 193 / KpV: 86)

What we see in this remark is another way in which Kant’s philosophy distances itself from the empirical (compare Kant’s quest for something absolutely necessary as addressed in § 8.1). Here it comes in the form of there being a power of the soul in a human practical agent that can be a master over all untutored and somewhat reflective motives that lead him away from acting in line with his duties as set by Kant’s categorical imperatives. I assume that this power of the soul for Kant belongs to same faculty as the power of the soul that can follow *FUL* in thought and sets laws for himself on basis of *FUL*. The *ratio subiectivae sumta* that thinks and sets laws for himself is the same *ratio subiectivae sumta* that can exercise control over an agent’s motivations and actions. That is, some sort of essential self.²⁰

²⁰ Note that when it comes to the aspect of ‘being a master over inclinations’ what matters rather than practical agents having no sentiments at the time of acting, is that a practical agent’s actions *proceed from the right motives*. Actions that proceed from the right motive are those that proceed from respect for the moral law. For Kant these are the sole morally worthy forms of actions. It is this type of actions Kant regards as *autonomous* actions. By contrast, *heteronomous* actions have motives other than the motive to perform one’s duty because it is a moral duty. Such motives can have some sort of reflective justification (say, an agent deliberately acting out of self-interest), but can

8.6. Summary

In the previous sections I have offered several ingredients that reflect and/or explain in a holistic comprehensive way the nature of *FUL*. Let me make one further remark. There might be a tension between a daily life aspect there is to Kant's philosophy and the motivation for *FUL* I proposed in § 8.2 on behalf of Kant. In § 7 I explained that Kant's philosophy has a close connection to daily life. There is textual evidence that one way in which this is the case is in the sense that Kant thinks that philosophers only have the task of illuminating a principle that is already present daily life. That seems difficult to reconcile with my proposed motivation for *FUL* 'the further way from the empirical, the more lofty'. My proposal would be to be faithful to the first idea by ascribing some sort of fairly simple golden rule principle to Kant when he mentions that philosophers only have a task to make more clear what is already very much part of our daily life, while for the remaining putting a firm emphasis on Kant's belief in necessity and a defeat of the empirical.

In regard to the (possible) tension pointed to, it should also be noted that some parts of Kant's work (also *within his practical work*) seem to have arisen from polemical motives in the author, while this is unlikely to be the case for others parts (see also Sullivan, 1989: 3, 4). I take it that occasions where Kant's writings are, or are likely to be polemical, are those that have some sort of close relationship with Kant's inclination 'to get away as far from the empirical as possible'. That is, in two ways. First in the sense of Kant's search for and emphasis on the importance of a hard form of generality: necessity (= a polemical response to Hume). Secondly, in the way of a polemical Kantian response to the moral sense theories of his contemporaries.

also proceed from primitive inclinations (say, an unreflective form of sympathy that strikes an agent's mind and leads him to act on it).

9. Kant on conscience and other aspects of the mind of the scoundrel

9.1. Kant's (naïve) optimism

A large part of Kant's account of evil takes place in *Religion within the boundaries of mere reason*. Unfortunately one is –or at least I am– looking in vain for a satisfactory explanation of the fundamental roots of evil behaviour. For Kant, evil behaviour is wilful acting out of line with what the categorical imperatives prescribe. As to 'wilful', Kant believes –quite plausibly– that unless the motives from which actions proceed are ones that can be chosen freely, practical agents cannot be held accountable for their actions. This counts both for good and bad actions. (see e.g. Rel: 21). As a result of this assumption, Kant avoids two commitments. First, he avoids a commitment to the idea that morally blameworthy actions are somehow an unavoidable upshot of our constitutional make-up. Secondly, Kant avoids a commitment to the idea that the performance of morally praiseworthy actions is necessarily supported by our naturalistic constitutional make-up. The latter seems problematic to me. However much reflective control a human being might have over his egoistic inclinations that pull him away from acting in line with Kant's categorical imperatives, I take it that without there being certain conditions related to her constitutional make-up, a human being would not be able to be motivated by Kant's categorical imperatives (or any other moral justifications). I take it that at least in the presence of extremely strong egoistic conative drive in the motivational profile of an adult human being (equally strong or stronger than the egoistic drives he was born with), while certain elements supportive of the performance of other-regarding actions were not in the constitutional motivational profile of this practical agent (elements such as the pre-dispositions to develop an other-regarding conscience and the natural disposition to experience empathic distress), a human being could not be motivated by *FUL* or any other Kantian categorical imperative. More about this in the next chapter. In the remainder of this section, I present and briefly discuss a passage from Kant's *Groundwork* about the minds of evil people.

In his *Groundwork* Kant presents the following view of evil people:

There is no one, not even the most hardened scoundrel—provided only he is accustomed to use reason in other ways—who, when presented with examples of honesty in purpose, of faithfulness to good maxims, of sympathy, and of kindness towards all (even when these are bound up with great sacrifices of advantage and comfort), does not wish that he too might be a man of like spirit. He is unable to realize such an aim in his own person—though only on account of his desires and impulses; but yet at the same time he wishes to be free from these inclinations, which are a burden to himself. By such a wish he shows that having a will free from sensuous impulses he transfers himself in thought into an order of things quite different from that of his desires in the field of sensibility; for from the fulfillment of this wish he can expect no gratification of his sensuous desires and consequently no state which would satisfy any of his actual or even conceivable inclinations (since by such an expectation the very Idea which elicited the wish would be deprived of its superiority); all he can expect is a greater inner worth of his own person. This better person he believes himself to be when he transfers himself to the standpoint of a member of the intelligible world. He is involuntarily constrained to do so by the Idea of freedom—that is, of not being dependent on *determination* by causes in the sensible world; and from this standpoint he is conscious of possessing a good will which, on his own admission, constitutes the law for the bad will belonging to him as a member of the sensible world—a law of whose authority he is aware even in transgressing it. The moral ‘I ought’ is thus an ‘I will’ for man as a member of the intelligible world; and it is conceived by him as an ‘I ought’ only in so far he considers himself at the same time to be a member of the sensible world. (G-Paton: 122-3 / G: 454-5)

This passage has some clear scent of overly optimistic, naïve and also rather premature psychology. In the abovementioned quotation Kant argues that “the most hardened scoundrel”, although unable to realize the aim of being a moral being, nevertheless wishes to be one. At least from a descriptive perspective of actual operative psychology, this seems to be a too ambitious claim in the form of a too ambitious generalization for those who are

slightly less hardened evil people. Furthermore, it seems to be an overambitious claim when it comes to some sort of deep self of the most hardened scoundrel. There may well be human beings (albeit perhaps only very few), whose deepest/essential self does not wish to be a moral being. Some or all psychopaths for example.

9.2. Kant's account of conscience

In the next chapter, I say something about the concept of conscience as relevant to an argument about the limits of objectivity and as grounded in research about psychopathy. As we have seen in chapter 1, it is generally agreed that psychopaths do not have a conscience. Kant might have wanted to disagree with that claim. Kant states that every man, as a moral being, has a conscience inherent in him. Kant however might have wanted to exclude psychopaths from the class of creatures who could ever become a human being. He might have seen them as brute beasts who somehow permanently lack the subjective conditions for susceptibility to duty as presented in § 7. That is not clear. He might also have wanted to include some psychopaths and exclude others. An interesting question here seems to be whether Kant would have made a different judgement about the presence of a conscience for two different types of psychopaths Ben Karpman (1949) distinguishes on basis of his observations as a clinical psychiatrist. A type that has quite a reflective nature and one that has the nature of rather brute beasts.

In a footnote in *MM* Kant regards 'conscience' as essentially being about some sort of intrapersonal relationship. He argues:

The man who accuses and judges himself in conscience must think of himself as a twofold personage, a doubled self who, on the one hand, has to stand in fear and trembling at the bar of the tribunal which is yet entrusted to him, but who, on the other hand, must himself administer the office of judge which he holds by inborn authority. And this requires clarification, if reason is

not to fall into self-contradiction. I, the prosecutor and yet the accused as well, am the same *man* (*numero idem*). But man as the subject of the moral legislation which proceeds from the concept of freedom and in which he is subject to a law that he himself gives (*homo noumenon*) is to be considered different (*specie diversus*) from man as a member of the sensible world who is endowed with reason. But it is only from the viewpoint of practical knowledge that he is to be regarded in this way, since there is no theoretical knowledge of the causal relation of the intelligible to the sensible; and this specific difference is that of the human faculties (the higher and the lower) which characterize man. The first is the prosecutor against whom the accused is granted a legal adviser (defence counsel). When the proceedings are concluded the inner judge, as the person vested with authority, pronounces the sentence of happiness or misery, as the moral consequences of the deed. Our reason cannot pursue further his authority (as ruler of the world) in this function; we can only reverence his unconditioned *iubeo* or *veto*. (MM- Gregor 104²¹ / MM: 439; italics original; *cfe.g.* MM: 400-1 /MM-Gregor: 60-62²²).

What we see here is that Kant thinks of the nature of conscience as an intrapersonal relationship in a practical agent between a prosecutor and an accused. Saying that that there is a role for a prosecutor and an accused *within* the same man, seems to be a correct way of phrasing what Kant thought. Be that as it seems to be, what is not so clear is how we should conceive of the role of the prosecutor in terms of the agent being subject to a voice and the agent somehow creating this voice. It seems as if Kant somehow sees a passive role for the practical agent as one who is subject to the voice of the prosecutor. At the same time however, one gets the impression that is too much of a passive account to ascribe to Kant. There might have to be a more active role for the practical agent as a prosecutor.

²¹ In Gregor's perception of the academy version, the original German passage is to be found on page 438. In my perception it is on page 439.

²² In Gregor's perception of the academy version, the original German passage is to be found on pages 438. In my perception it is on page 399-401.

10. Looking ahead

In the next chapter I make use of the very different elements I have addressed thus far in my thesis to provide one argument and a building block for two further lines of reasoning in favour of there being a non-realist component to moral objectivity. The Kant exegesis I have provided in this chapter should do the job of illustrating the first argument and helping us to see why it is important to focus on the human nature when being concerned with the problem of the status of moral objectivity.

Chapter 5: 'Building towards three anti-realist criteria'

1. Introduction

In this final chapter I present one fully developed criterion for there being a non-realist component to the nature of moral objectivity (if there indeed is such a thing as moral objectivity; a problem that is beyond the purposes of this study). This criterion arises from Hume's philosophy. Besides presenting this fully developed criterion, in this fifth chapter I also present a *comprehensive* argumentative foundation for two further criteria for there being a non-realist component to moral objectivity. These two criteria I expect to arise from future research that builds on two qualitatively different lines of argumentative reasoning I develop into detail at the very end of this study, i.e. in its overall conclusion.

The comprehensive argumentative foundation I present in this chapter consists in the combination of two ideas. The first idea reflects an empirical fact about psychopaths. The second arises from a hypothesis about the ultimate cause of psychopathy –nature rather than environment-, while using this thought for a conclusion about the mind of non-psychopathic 'typical' practical agents. Concretely, the two ideas are:

- (a) Within the class of human practical agents there is a class of agents (psychopaths) who do not take the interests of others seriously in practical deliberation and judgement.

- (b) Without there being certain elements in her constitutional make-up, elements given for free by the author of nature, (whoever or whatever that is) a human practical agent will not be able to have psychological experiences that include sincere moral reasoning and moral decision-making.

So what I offer in this chapter is the abovementioned argumentative foundation. In the general conclusion then I offer two lines of reasoning that are two different argumentative branches for two different criteria for there being a non-realist component to moral objectivity. These two different argumentative branches have in common that they both build on the comprehensive argumentative foundation I present in this chapter. This comprehensive argumentative foundation on its turn builds on thoughts about psychopathy as presented in chapter 1.

By the very end of this study I will have offered two detailed lines of argument that both have the potential to be criteria for there being a non-realist component to the nature of moral objectivity. My assumption is that future research by others and/or me that builds on this study delivers two criteria that follow from the lines of argument I offer in this study. Having just said that I assume that there arise two criteria from future research, future research might show that there do *not* arise two criteria. There is an issue with the first line of argument in the sense that it relies on a controversial largely empirical assumption I endorse. That is, the assumption that at least about some psychopaths it can accurately be said that they have (what by some philosophical criterion can qualify as) a genuine evaluative normative outlook on the world. In regard to the first line of argument I offer, I assume that the criterion for there being a non-realist component to moral objectivity does not get off the ground once this assumption is false. I explain this further in the general conclusion.

For the second line of argument it is the case that future research needs to indicate how *exactly* the second line of argument constitutes an argument in favour of there being a non-realist component to the nature of moral objectivity. What delays a full-fledged development of this criterion are complex issues about necessity, contingency and possible worlds of human nature. These issues need attention in future research before any definite conclusions can be drawn about the exact nature of the second argument as reflective of a criterion for a non-realist component to moral objectivity.

In the general conclusion I lay out the two lines of argument into detail (up until the end point of this study). This I follow up by outlining the framework for philosophical and empirical research I take to be needed before definite conclusions can be drawn about whether and how –in full detail- the two lines of argument as they arise from the argumentative foundation I present at the end of this chapter constitute an argument in favour of there being a non-realist component to moral objectivity.

This chapter is set up as follows. In § 2 I present the full-fledged first criterion for there being a non-realist component to the limits of moral objectivity. As indicated, this criterion is based in Hume's philosophy. More specifically, it is based in his philosophy of reason. Strictly speaking this criterion arises from Hume's philosophy only, but it is well-illustrated by means of considering it in the context of Kant's philosophy. This I will do. In § 2.1, I do some preliminary work for understanding the nature of the criterion by recalling essential thoughts about Kant's and Hume's conception of 'reason' as presented in chapter 2 and 4. Subsequently in § 2.2, I connect Hume's conclusions about the limited powers of reason to Kant's philosophy.

In § 3, I recall some essential thoughts about psychopathy. That is, I recall thoughts about the nature of the syndrome as it presents itself in the mind and behaviour of psychopaths while I also refer back to scientific thoughts about the *etiology* of psychopathy.

In § 4, I present the concept of 'pre-intentional emotions' as developed by Ratcliffe. 'Pre-intentional emotions' determine what kind of intentional state it is possible to have. They provide us a range of possibilities while excluding another range. They are emotions that are characterized by a person's sense of what is possible (Ratcliffe, 2010).

In § 5, I concentrate on theoretical explanations and examples of (a) cases in which practical agents can easily understand other people's minds and (b) cases in which practical agents might have real difficulties understanding other people's minds. As to (b), after having focused on the case of depression for introductory purposes, I point out several ways in

which human beings without psychopathy are likely to have difficulties understanding the minds of psychopaths.

In § 6, I develop two conceptual suggestions as to an ultimate etiological explanation of why it is psychopaths do not take the interests of others seriously in practical deliberation and judgement. The first represents a comprehensive conceptual idea that appeals to (a) a lack of pre-dispositions for empathic distress; (b) an absence of pre-dispositions for the development of a conscience and (c) an absence of pre-dispositions for the development of a particular dialectical self. This comprehensive first conceptual suggestion assumes that psychopaths have an animate perception of other human beings. The second conceptual suggestion reflects the idea that amongst the class of psychopaths, there might be some who have an inanimate perception of other human beings due to a particular perception of themselves.

Section 7 is an intermezzo section. I apply Ratcliffe's concept of pre-intentional emotions to *FUL* in such a way that I deny Kant's claim that his ultimate moral principle is completely free from empirical influences.

In § 8, I intertwine Ratcliffe's research and empirical research about psychopathy by presenting the argumentative foundation consisting in the combination of the two ideas mentioned.

2. Hume (and Kant): the first criterion

2.1 Recalling Kant and Hume on 'reason'

In chapter 4 I have argued that there is a subjective component –a *capacity of the soul*– as well as an objective aspect to Kant's concept of 'reason' ('Vernunft'). I have provided the following suggestion as to how the subjective and objective component should be conjoined in one comprehensive Kantian concept.

‘Vernunft’ as a capacity of the soul reflects a *potentiality of the soul* the boundaries of which are *fixed*. (Recall that because of the 18th century meaning of the Latin term ‘*facultas*’, it should be assumed that for Kant a ‘capacity’ is *not* something that is quantitatively extendable.) I have proposed that for Kant this capacity of the soul can be actualized to a greater or lesser extent. I have also proposed that ‘Vernunft’ as a power of the soul is intrinsically connected to objective principles in the sense of the latter being the former’s ‘potential objects of acquired knowledge’. This means that these ‘potential objects of acquired knowledge’, if indeed acquired, reflect an efficacious use of the normatively rational potential of Vernunft as a power of the soul.

In chapter 4 I also focused extensively on Kant’s *FUL*. As explained, for Kant there is a strong relationship between *FUL* and *reasoning*. I have explained that *FUL* can be constructed as a practical syllogism. It can be done so from a non-subjective theoretical perspective. However, it could be that we even have to say that a (non-natural!) *subjective* reasoning procedure is *part of FUL’s essential nature*. That is, for Kant *FUL* might not be able to exist without it being intrinsically connected to a test for the universalizability of maxims as conducted as a reasoning procedure in the supersensuous intellectual mind of practical agents. I am not entirely sure however as to whether or not we *can* or even *should* say this.

I now summarize Hume’s conception of ‘reason’. In chapter 2 I argued that for Hume ‘reason’ is an exclusively psychological concept. This in contrast to what is the case for Kant. As an exclusively psychological concept, for Hume ‘reason’ is essentially a *reasoning process/activity*. Possibly (sometimes) for Hume, this reasoning process is part of a *faculty view* of reason. Here then, I have claimed, the notion of contradiction becomes important for both the area of mathematics and the areas of matters of fact. I have claimed that Hume wants to argue that it is only when our faculty of reason *cannot* conceive of contrary instances that it can give us *a priori* knowledge as it can in mathematics. When reason *does* give us a *priori knowledge* the knowledge results from *a priori reasoning processes* that are

terminated by our faculty of reason by means of an *a priori act of judgement* resulting in an *a priori product of judgement* reflecting *a priori knowledge*.

2.2. Hume's philosophy and the first criterion

Tacitly, there seem to be three different claims in Hume's work when it comes to matters reason, passion and morality. First, the claim that reason alone cannot generate moral discriminations. Secondly, the claim that passions are a necessary component of moral discriminations. Thirdly, the claim that "reason alone can never be a motive to any action of the will" (T: 413). Hume subscribes to all three and Hume himself does not separate them clearly. And neither are they typically separated in scholarly literature. I believe however that they should be. It is helpful in terms of Humean exegesis. Also, it is crucial when comparing Hume's moral philosophy with accounts from others, paradigmatically Kant. It is also a helpful tool when the aim is to make progress with the problem of the nature and status of moral discriminations. In this chapter, I provide support for the view that setting apart the first two Humean claims from the third is important in regard to Kant's philosophy. What results from my presentation of this argument is a paradigmatic *illustration* of the first criteria for the limits of moral objectivity. As to 'illustration', *strictly speaking*, an appeal to Kant's philosophy is not needed for an identification of the first criterion. However, the argumentative force of the first criterion is revealed in a very clear way once it is considered in the context of Kant's philosophy. Furthermore, an appeal to Kant's philosophy in interaction with Hume's conception of reason (in the way I intend to present in this section) informs us about a high intellectual cost that seems to come with meta-ethical theories that challenge the first and second Humean claim identified above. That is, in case such theories do fundamentally rely on a Humean psychological conception of reason as reasoning.

Kant intended *FUL* to be an *a priori* justification. As explained in the previous chapter, there are two reasoning variants of *FUL*: *CCT* and *CWT* (G: 424). I take it that *CCT* as one reasoning variant of *FUL* in some way responds successfully to Hume's concern that reason alone cannot deliver moral discriminations, a claim of which the other side of the coin (at least for Hume) is that reason needs help from impressions. Namely, what characterizes *FUL* is its intrinsic reliance on an *a priori* reasoning procedure with an essential role for the principle of non-contradiction (the mother of all reason). Because of this, I take it, *FUL's CCT* is not in need of any impressions (i.e. neither external sensory nor passion-type of impressions) to terminate a reasoning process, a concern Hume had about 'reason'. My suggestion is that in the context of *FUL*, the principle of non-contradiction can do the same job of terminating a mental reasoning process as passions do for Hume. I suggest that *FUL's CCT* can function as a stopping tool for a reasoning process and is thereby a candidate for preventing that a practical agent with his capacity for linking thoughts step by step stands "like the schoolman's ass, irresolute and undetermined" (EPM: 172). This claim from Hume seems to be about motivation rather than moral judgement. What Hume seemingly wants to argue here is that unless there are passions, the will does not know what to will. However, given Hume's conception of inferential reasoning, I think we can apply the presented Humean phrase to passion-type of impressions and *judgement* too as something Hume would have been sympathetic to in regard to his moral philosophy. Perhaps he can even be said to argue for it. Hume clearly thought that only passions can determine a reasoning process, but this claim is more explicit in his philosophy about causality/matters of fact than in his moral philosophy. Having said that, I think we can apply it intelligibly and indeed helpfully (even if just for purposes of Hume exegesis) to his moral philosophy.

So *FUL's CCT* seems to be able to substitute the role of impressions Hume thinks are necessary to terminate a reasoning process. If this counts as a *success* for Kant and/or *FUL's CCT*, the *problem* with the mentioned Kantian reasoning variant of *FUL* seems to be that outside the context of Kant exegesis (i.e. a context where an appeal to intelligible agents is

made) it runs the risk of being no more than a piece of philosophical curiosity having to do with non-contradiction (*cf* O'Neill, 2004). Phrased from the perspective of *FUL's CCT's* success in regard to Hume's claim, the issue is the following. The (alleged) fact that there is a philosophical device to stop reason's thinking processes is one thing. That device having authoritative force as a moral truth and/or binding force as a moral command is quite another.

The above seems problematic about *FUL*. However, it possibly is only problematic if *FUL* is considered *outside the exegetical context* in which it for Kant's functions; a context that appeals to intelligible beings in a supersensuous realm. As long as *FUL's CCT* functions *within* this context, it might be able to escape the problem for one or two related, but different reasons I explain below.

As briefly indicated in chapter 4, it might not be accurate to say that there is a single 'highest principle of morality' for Kant. Whether or not it is *inaccurate*, there is anyhow something *misleading* about the phrase. In short, the issue is that either Kant's 'highest principle of morality' is one principle that comes in a variety of guises or else there are a few different (related) 'highest principles of morality' Kant offers in his work, one of which is *FUL*. Roughly speaking there are four other principles/guises of Kant's highest principle: the *Formula of the End in Itself (FEI)*; the *Formula of Autonomy (FA)*; the *Formula of the Law of Nature (FLN)* and the *Formula of the Kingdom of Ends (FKE)*.¹ In scholarly literature it is

¹ Below I present Kant's descriptions of the five (versions of the) categorical imperative(s):

FUL: "Act only on that maxim through which you can at the same time will that it should become a universal law". (G-Paton: 88 / G: 421; italics original).

FEI: "Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end". (G-Paton: 96 / G: 4:429; italics original)

FA: "Never to choose except in such a way that in the same volition the maxims of your choice are also present as universal law". (G-Paton: 108 / G 4:440)

FLN: "Act as if the maxim of your action were to become through your will a universal law of nature" (G-Paton: 89 / G: 421; italics original)

FKE: "[N]ever to perform an action except on a maxim such as can also be a universal law, and consequently such that the will can regard itself as at the same time making universal law by means of its maxim". (G-Paton:

typically assumed that all of these different categorical imperatives or rather ‘versions of one and the same categorical imperative’ are *somehow* related. The question is ‘just’ how. The puzzle about the relationship between Kant’s *FUL*, *FEI*, *FA*, *FLN* and *FKE*, arises especially in regard to the relationship between *FUL* and *FEI*, since these two give the impression to be the ones most important to Kant while their content looks very different. For the remainder of this section, an understanding of *FUL* and *FEI* will be relevant. Therefore recall that *FUL* states:

Act only on that maxim through which you can at the same time will that it should become a universal law. (G-Paton: 88 / G: 421; italics original)

FEI states:

Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end”. (G-Paton: 96/ G: 4:429; italics original)

I take it that ‘Vernunft’ as a concept is central to all categorical imperatives. I also take it that ‘Vernunft’ has more implicit and explicit meanings in the different (versions of the) categorical imperatives. Here my assumption is that there is some comprehensive Kantian conception of ‘Vernunft’ that has different aspects to it that for Kant have some sort of metaphysical relationship. The difference I have in mind here concerns an intellectual power and an essential self. I assume that Vernunft more explicitly and implicitly for different aspects and different *CI*’s underlies all categorical imperatives. In *FUL*, e.g. it is primarily (or perhaps better ‘in the foreground’) an intellectual power of the soul that –I suggest- in *FUL*’s *CCT* gets an impulse of Kant’s strong sympathy for the *a priori* and his antipathy towards the

101 / G: 434; italics original)

I borrow the abbreviations for the five (versions of the) categorical imperative(s) from O’Neill (2004).

empirical. In *FEI* e.g. it is primarily an *essential self*. The essential self, I take it, is on some ultimate level also inherent to *FUL* but stays very much implicitly in the background there.

Let me get back to my claim that as long as *FUL's CCT* functions *within* a Kantian exegetical context where an appeal is made to intelligible agents in an intelligible realm, *FUL's CCT* perhaps can be said to rightfully escape the problem of being no moral justification *while being* a successful response to Hume's concern about the limitations of reason. *FUL* might do so because *FUL* is *inextricably bound up with* the concept of Vernunft *as seated in an practical agent*. As such both Vernunft and *FUL* seem inextricably bound up with an essential intellectual self. Assuming this, the fact that there is an intellectual essential self to a practical agent that deserves respect may prevent *FUL's CCT* from being no moral justification, while having the feature of being a successful device to answer a Humean concern about the limitations of reason. Alternatively or additionally, because Vernunft is the connector for all (versions) of Kant's categorical imperatives, *FUL's CCT* may be able to make use of the argumentative force of one or more of the other (versions of) Kant's highest principles of morality (assuming for the sake of the argument that at least some have indeed argumentative force).

3. Recalling psychopathy

In chapter 1 I have discussed psychopathy. Much of the discussion has revolved around the scientific hypotheses and claims of Robert Hare. I have intended to make clear that the discussion about the nature of psychopathy is not settled. While the dominant conception of psychopathy holds that psychopathy is a 'personality syndrome', there are other conceptions as well in scientific literature (that however are often compatible with the conception of psychopathy as a personality syndrome). Furthermore, assuming that it is a personality syndrome, there is room for reflection as to what character traits and behaviours belong to it.

In chapter 1 I have explained that the dominant measurement tool for psychopathy is the PCL-R. I have explained that it consists of a semi-structured interview accompanied by case history information. Recall that the semi-structured interview consists of 20 items that can each score a '0', '1' or '2', giving a maximum score of 40. I have also explained that a psychopathy diagnosis requires minimally a score of 30 out of 40. *Technically*, this would be possible by scoring low on several items of the PCL-R that concern traits and behaviours that are generally considered to be core to psychopathy. I yet have to come across a case however, where someone has been diagnosed as a psychopath and scored '0' on the item 'lack of empathy', something which is generally considered to be an essential feature of psychopathy.² Recall here that there is scientific consensus in psychological literature and clinical practice that something that is core to psychopathy are one or more affective atypicalities. From the perspective of empirical literature, a lack of empathy is the first candidate for being an essential affective atypicality. Other affective candidates are a lack of remorse and guilt and a lack of conscience.³ Let us also recall that other (non-affective) features that are often considered to be a core characteristic of psychopathy are a sense of grandiosity, superficial charm and a disposition to lie frequently.

We have seen that the classification of psychopathy, unlike the diagnoses of conduct disorder or antisocial personality disorder, identifies a relatively homogeneous population. There is a unitary disorder upon which a causal account can be developed. And on the basis of current research, there is reason to think that there is a genetic and not a social ultimate cause to the core of the syndrome (*cf* Blair et al., 2006). Clearly there is a lot of unclarity as to what it means that there is a strong genetic component to psychopathy, but nevertheless

² That is, even wider than Robert Hare's conception of psychopathy and even wider than either the PCL-R concept or measurement tool which -unfortunately- are not so clearly separated.

³ Note that the former reflects an item of the PCL-R, while the latter does not. Having said that, it is a widely accepted thought that psychopaths lack a conscience. Indicative is the title of Hare's seminal work: *Without conscience, the disturbing world of the psychopaths amongst us*.

there seems to be such a thing. Blair et al. (2005: 45) claim that “genetic abnormalities give rise to a specific deficit in neurotransmitter function and neuro-development such that the emotional responsiveness of individuals with psychopathy is muted”.

4. Pre-intentional emotions

Typically, in philosophy, emotions are regarded as intentional states, bodily feelings or a combination of both. There is also a tendency in the philosophical literature to focus on a fairly standard inventory of emotions moods, including anger, sadness, fear, joy, grief, jealousy, guilt, and so on. As a result, a range of other emotional states, many of which do not have established names, have been neglected. Although the category of ‘neglected emotional states’ is not itself phenomenologically homogeneous, many of these neglected phenomena do have something in common. They are not intentional states, directed at objects, and they are not feelings of the body or some part of it. Instead, they amount to a felt sense of belonging to the world (Ratcliffe, 2012a: 23-24). Ratcliffe calls this type of emotional experience ‘existential feeling’ (2005, 2008) or ‘pre-intentional emotion’ (2010). Ratcliffe uses these different terms in different articles while seeing them apparently as synonyms (see Ratcliffe, 2010: 605). Pre-intentional emotions determine what kind of intentional state it is possible to have. They provide us a range of possibilities while excluding another range. They are emotions that are characterized by a person’s sense of what is possible (Ratcliffe, 2010). These emotions are typically unacknowledged as elements that make a contribution to one’s life. The importance of them to our world-view typically becomes only obvious in the case of emotional disruptions in our life. It is when we are presented with the loss, exaggeration or

distortion of emotional aspects that we ordinary overlook that they become salient (see Ratcliffe, 2008, cited in 2012b: 483).⁴

5. Easy and hard cases of understanding another person

5.1. Introduction

In our daily life we often have a good idea as to what is going on in the mind of another person. When we see someone running towards a bus waving, many of us will understand her mental state. In some veridical sense we might experience her sense of urgency. Empathetic understanding of a person rushing after a bus is easily achieved because it takes place against the backdrop of a shared cultural world. Because the person who is waving and running towards a bus shares a world with us characterized by certain norms and artifact functions, it is possible for us to quickly and automatically have a sense of the inner state of the other person (see Ratcliffe 2012b: 477).

An *absence* of a shared cultural world can problematize empathetic understanding, but it need not do so. Much depends upon the content of the experience with which we are attempting to empathise. For example, when presented with a sobbing face clutched in a person's hands, cultural differences might not interfere at all with the ability to appreciate the person's sadness (in some cases, at least). When the content of the experience is not just 'B is sad' but 'B is sad about *p*', matters are different. As Ratcliffe makes clear, some people will find it harder than others to empathise with B's being sad about the fact that no one ever visits his Facebook page (Ratcliffe, 2012b: 478).

⁴ I am not entirely sure what exactly Ratcliffe wants to argue for here. To use an example of a depressed person Ratcliffe could mean to say that up until the time of his depression the person had no idea *that* there was a default state of emotional experience in him. I would say that most probably Ratcliffe wants to argue for this claim. Probably Ratcliffe also wants to say that before his depression the depressed person had no idea about the *nature* of this default state of emotional experience.

5.2. The depressed person

In some circumstances many of us will have a really hard time understanding the mental experience of another person. Take the case of severe depression or a severe burn-out. Many of us, while reading some sort of article about depression, may fail to understand the first person claims of a depressed person that are cited in the article. We might wonder, e.g. how to understand the claim of depressed Jim that the world has no practical significance for him. We might get some sort of idea as to what Jim means, while having the feeling that we do not fully understand what Jim says.

Suppose we do not get Jim fully. We may come to understand him better when being in his presence. A personal encounter may help us understand his first-person claims. While such a personal encounter may be helpful to come closer to an understanding of the mental state of Jim, it might be the case that a maximally good understanding of his mental world (as well as Jim's first-person claims resulting from it) is only possible once we have experienced a severe form of depression ourselves. Without a personal experience of depression the propositional content of Jim's claim might not ring a real bell. To use the metaphor, the propositional content might make noise, but might fail to do so in a deep way that reflects Jim's experience.

5.3. Understanding psychopaths: profound experiential difficulties

Just as it most probably is difficult for most of us to understand what it means to live in a world where all practical significance has disappeared as in the case of Jim, so it most probably is difficult for many of us to understand the claims and beliefs of psychopaths and psychopaths' internal world that accompanies them. It might be impossible for many of us -

beings with a certain temperament and say, relatively successfully socialized - to experience the world as a psychopath experiences it.

According to Robert Hare it is “close to impossible” to imagine the world of psychopaths (Hare, 1993: 78). There is a variety of specific substantive ways we could understand that claim. First, we ‘typical human agents’ may not be able to understand psychopaths’ pervasive sense of egoism.⁵ For psychopaths, egoism is an all-encompassing way of being. Typical human beings may not know what that is like in terms of experience. Because we typical human agents do not have exclusively egoistic experiences, we may lack an understanding of how it is to be an absolute egoist at all times. It sounds plausible to me that our own experience implies limits for understanding the mental world of psychopaths.

Secondly, typical human beings may not be able to understand what it is like to be indifferent to the suffering of others. Here as well, typical human beings’ own experiences could imply limits for understanding the mental world of psychopaths. Indeed, that is what I would want to suggest.

Thirdly, typical human beings may not understand what it is like to lead a life without an internal control mechanism. I believe Hare is right in claiming that in typical human agents the inner voice of conscience and the internalized norms and rules of society act as an ‘inner policeman’, regulating our behaviour even in the absence of the many *external* controls, such as laws, our perceptions of what others expect of us, and real-life policemen (Hare, 1993: 75). In psychopaths’ psychology there is no such internal policeman. Psychopaths carry out their evaluation of a situation - what they will get out of it and at what cost - without the usual anxieties, doubts, and concerns about being humiliated, causing others pain (Hare, 1993: 78). Without the shackles of a nagging conscience and other inhibiting mental mechanisms, psychopaths feel free to satisfy their needs and wants and do

⁵ For the sake of avoiding unnecessary complexity as to the point at issue, I group all non-psychopaths in one class and refer to them as ‘typical human agents’. Outside the current context, dependent on the situation, (a) referring to non-psychopaths as ‘typical human agents’ and (b) putting them in one class as if there were a homogenous group, might not make sense. Indeed, as to (b), this will often make very little sense.

whatever they think they can get away with. Any anti-social act, from petty theft to bloody murder, becomes possible (see Hare, 1993: 76).

My suggestion would be that we typical human beings do not have and cannot have a full-fledged idea as to what it is like to live a life without internal policeman, because there is no way in which our own experience can lead us to a deep grasp of that feature of psychopaths' psychology.

Fourthly, typical human agents are likely to be unable to understand what it is like to act in a mental atmosphere of strong primitive egoistic and sexual drives that strongly direct one's behaviour. This also as a result of their own experience.

In terms of ways of having difficulties understanding the mental life of psychopaths, I would like to present one further suggestion related to the normative views of psychopaths, if they indeed have such things. In chapter 1 I have explained that psychopaths use other human beings as objects of self-gratification. Without being inhibited by typical mental barriers such as a conscience, psychopaths use others to fulfil their own egoistic needs. Psychopaths hold a selfish, instrumental world view, and prey upon and exploit others, using aggression and manipulative conversational skills as weapons (see Hancock, Woodworth & Porter, 2011). Psychopaths have no difficulty in making use of people whenever they think using others serves their interest. Their motives are to manipulate and take, ruthlessly and without remorse (Hare, 1993: 145). So psychopaths *use* others as 'objects of exploitation'. Some, if not all psychopaths however, also *see* their used objects of exploitation as *legitimate* objects of exploitation. As Babiak and Hare (2006: 46) argue, it is unlikely that for psychopaths people exist in any other way than objects, targets and obstacles (see Babiak & Hare, 2006: 46). Psychopaths view people as little more than objects to be used for their own gratification (Hare, 1993: 44). The weak and the vulnerable - whom they mock, rather than pity - are favourite targets (Hare, 1993: 44).

While all psychopaths somehow seem to view people as little more than objects to be used for their own gratification, some seem to have and possibly only *some* have a real

normative rationale related to that view. For those psychopaths who can be said to have a normative rationale for their view, I take it that it is the case that they hold the normative view that people who are weak deserve to be exploited *because they are weak*.⁶ As Hare cites psychologist Robert Rieber: “There is no such thing, in the psychopathic universe, as the merely weak”. Whoever is weak is also a sucker; that is, someone who demands to be exploited”⁷ (Hare, 1993: 44). As non-psychopaths, we may not be able to see even some sort of minimal intelligibility in such a normative view. I do not want to suggest that those views are intelligible. More generally, in this section, I do not want to commit myself to any sort of normative view about the intelligibility of the view under consideration. It is a real option that his view of psychopaths is not intelligible in a normative sense (and therefore we would have good reason to not find them intelligible). Rather than commenting on that, in this section, I ‘just’ want to argue that we as non-psychopaths might not understand the intelligibility of the view mentioned. And I want to claim that supplemented by the suggestion that we might not understand its intelligibility because we lack the possibility to experience the experiential nature of psychopaths’ conviction. Granting for the sake of the argument that there are indeed some amongst the people in the class of non-psychopaths who cannot in any way understand the intelligibility of the psychopaths’ view, I assume that if those people were able to experience the *sine qua non* feeling and conviction coefficient that Hume regards as part of ‘belief’, they would come to understand the psychopath’s view as intelligible. Probably one thorough experience would be enough. A working affective memory would do the rest.

Comparing the inner world of psychopaths and non-psychopaths in a rather general all-encompassing way of experience, I believe that psychopaths and ‘more typical human

⁶In his (1993: 107) Hare relates this attitude/worldview to a thought from John Grambling whom Hare regards as a white-collar psychopath: “In Grambling’s mind, anyone who is stupid enough to trust or believe him deserves the consequences”.

⁷Rieber, R (1997). *Manufacturing Social Distress: the Psychopathy of Everyday life*. New York: Plenum [cited in Hare, 1993: 223; I have adapted the reference to the reference system I generally use in this study].

beings' do not share the same 'modal space' (see Ratcliffe, e.g. 2012b: 486). That is, I assume that psychopaths and 'more typical human beings' do not have the same phenomenological range and nature of possibilities of belonging to the world. Indeed a very different phenomenological range and nature.⁸

6. Bio-psychological conditions for not taking the interests of others into account

6.1. Introductory remarks

Psychopathy is typically regarded as a syndrome. That syndrome comes with a number of features of psychopaths' mental life and behaviour. For each feature an empirical scientist could try to give a causal explanation. Note that a behavioural feature or character trait characteristic of psychopathy can be causally explained *in different ways* as well as *on different levels*. As an empirical scientist one can for example try to identify some *genes* that as a *rock-bottom fundamental cause* can be given causal credits for the manifestation of a particular character trait or a particular type of behaviour. Additional or alternative to the fundamental level of genes, on a *less fundamental level* of explanation, one could try to identify a cause of a certain type of behaviour characteristic of psychopathy by identifying some sort of psychological mechanism that causes it. Or, to focus on different *types* of explanation (rather than levels), rather than attempting to give a genetic explanation, one could want to offer a neural explanation of particular character trait or type of behaviour that characterizes psychopathy.

As explained in chapter 1, current empirical research suggests that there is a genetic and not a social ultimate cause to the core of the syndrome (*cf* Blair et al., 2006). In the remainder of this section, rather than focusing on the ultimate roots of psychopathy from an

⁸ How flexible these ranges are is a very difficult empirical problem. It is also an issue for which it might be the case that there is no one single answer that counts for all who by some criterion qualify as psychopaths.

empirical perspective, I present several conceptual ideas. These ideas reflect a philosophical explanation of the core of the psychopathy syndrome. Importantly, they also reflect a characteristic feature of psychopaths: their psychological impossibility to take the interests of their fellow human beings into account in any form of evaluative judgement about practical life. Following the dominant empirical view, the conceptual ideas I present concern the causal level of *pre-dispositions*.

The reason I focus on psychopaths' alleged psychological impossibility to take the interests of their fellow beings into account in their practical decision-making (and for some probably in what count as genuinely normative judgements—that have a sufficient rationale) is that I appeal to the object of this alleged psychological impossibility in both of my psychopathy based arguments as to their being limits to the objective nature of moral objectivity (recall that there is another argument based in Hume research). Roughly speaking, I do so by assuming that conceptually speaking the forms of moral objectivity contemporary moral realists and rationalists defend take the interests of others into account in some way.⁹

⁹ It might even be the case that conceptually speaking all forms of moral objectivity are other-regarding in some way, but one should be rather careful to assume this, i.e. *even if* one has a very minimalistic conception of 'other-regarding', *not* meaning '*pro-social*'. The reason is that a conception of moral objectivity as necessarily 'other-regarding' might illegitimately rule out ethical egoism as a form of 'moral objectivity'.

In philosophy and psychology the term 'morality' is used in a variety of ways. One popular use is a normative one about it reflecting objective standards of rightness and wrongness in regard to states of affairs, courses of action et cetera. However, there are also descriptive definitions of morality, especially in psychology. I take it that if and when the mentioned normative definition of morality is used by academics it is meant to apply to a domain of the world where other human beings are around. 'Morality' here then one could define as being 'the standard of right and wrong in our dealings with other persons' as Thomas Nagel (1999) does. Nagel might be happy to use the just mentioned definition also for the term '*moral objectivity*'. In any case, the advantage of the mentioned Nagelian definition of morality, i.e. if we apply it to moral objectivity, seems to be that it is open to ethical egoism. This possibility seems to be excluded when one claims that 'all moral objectivity conceptually speaking is necessarily other-regarding'. Assuming for the sake of the argument that ethical egoism is excluded once the mentioned conceptualization of moral objectivity is adopted, that might be problematic. It might be problematic because (a) it might be both a necessary and *sufficient* condition that moral objectivity consists in standards of rightness and wrongness in dealings with other persons; given (b) that the underlined phrase might accurately refer to what can normatively count as 'the moral domain' and given that it seems plausible to see ethical egoism as a philosophical position defending a moral standard (where the adjective 'moral' here at least matches the underlined bit from an independent conceptual perspective, regardless how the particular author who intends to defend a form of ethical egoism intends to use the term. Additionally (or alternatively for the author) the adjective 'moral' as applicable to the normative standards of an ethical egoist might refer to standards about 'the good life').

While I think it is an intelligible conceptual option to see all moral objectivity as necessarily other regarding, while - for the reason mentioned above- I also think that care should be exercised in assuming that all

6.2. Three types of missing seeds

In this section I present three conceptual ideas that causally and conceptually explain psychopaths' alleged psychological impossibility to take the interests of others seriously (and more broadly explains core traits and behaviour of psychopaths as presented in chapter 1). A psychological impossibility I assume to be *long-term* for all psychopaths, if not *permanent* for all. I assume it to be *permanent* at least *for some*. And at least for some of *those* psychopaths, I assume that the presumed psychological impossibility is due a *constitutional incapacity* to take account of others. In their union, the three ideas I present below can be taken as *one type of comprehensive conceptual suggestion* that causally explains psychopaths' alleged psychological impossibility to take the interests of others seriously, with another type of conceptual suggestion to follow afterwards (NB: the first (comprehensive type) of suggestion assumes that psychopaths have an *animate* perception of other human beings, while the second assumes that psychopaths have an *inanimate* perception of other human beings.)

The first idea I would like to offer as an ultimate conceptual explanation of the core of psychopathy, including psychopaths' alleged psychological impossibility to take the interests of others seriously, is the following one. I assume that fundamentally psychopaths do not have an internal feedback mechanism in terms of pain and distress of others the roots of which I assume to be innate. They lack the possibility for empathic distress. I would want to suggest that empathic distress is largely an innate feedback mechanism freely given to practical agents and not under typical reflective control. That does not mean that I believe that once that mechanism is in place, it cannot be destroyed by others or the agent herself. I

forms of moral objectivity are other-regarding, one thing seems beyond dispute. That is, it seems beyond dispute that factually speaking almost all, if not all, contemporary forms of moral realism and moral rationalism are other-regarding in some way. It seems to be the case that most if not all contemporary meta-ethicists who call themselves moral realists and moral rationalists subscribe to a form of moral objectivity that in its standards in one way or another takes the interests of others into account.

rather argue that typical practical agents can't help not having the slightest inclination not to be unsympathetic towards this feedback mechanism.¹⁰

All three conceptual suggestions I provide are somewhat tentative in the sense of them being *suggestions*. But they are *strong* suggestions. More tentatively than what applies to the broad suggestion offered and the two upcoming ones, I suggest that the assumed constitutional incapacity for empathic distress addressed above, in very young infants, manifests itself in an inability to mimic other people's facial expressions and possibly also in an inability to respond to the cries of fellow infants. As capacities, they might be precursors for a full-fledged capacity to vicariously experience other people's mental states, including their distress.

My second philosophical suggestion as to an ultimate cause that can explain psychopaths' alleged psychological impossibility to take the interests of others seriously, is the following one. I assume that some, if not all, psychopaths are constitutionally incapable of developing an 'other-regarding conscience'. If not *constitutionally (utterly) incapable*, they may be in a constitutionally disadvantageous position, because of which they have a *profound difficulty* in developing one.

Above, I have underlined the term 'other-regarding' for an important reason. As we have seen in chapter 1, it is claimed that psychopaths have no conscience.¹¹ I doubt that this is an accurate conceptual claim in the sense of it being applicable to the mental world of psychopaths as we know it on basis of current scientific evidence. I doubt this because I am inclined to think that the concept 'conscience' cannot be identified with 'moral conscience'. I share the dominant opinion in empirical science that psychopaths lack *such* a type of conscience. But I doubt that it follows from that both conceptually and conceptually-cum-empirically that therefore psychopaths necessarily lack a (i.e. all types of) conscience. Said

¹⁰ If they get to the point of introspectively recognizing that it is there.

¹¹ Indicative is the title of the seminal work from Robert Hare: *Without conscience, the disturbing world of the psychopaths amongst us*.

slightly differently, I tentatively assume that at least some psychopaths have something in their psychology that deserves the label 'conscience'. That is, I find it plausible to think that in (current PCL-R diagnosed) psychopaths, there are some who are subject to an internal tribunal, to use Kant's term. My suggestion would be that in contrast to typical human agents, in (some) psychopaths there is a voice that does utter its disapproval, but in a different way as is the case for most other human beings. I presume that in psychopaths, the internal tribunal utters its disapproval once the acting agent misses out on opportunities for self-gratification in a way he can be held responsible for.

My third and final suggestion as to the roots of the core of psychopathy, including the psychological impossibility to take account of the interests of others in practical judgement is connected to the second. That is, in the sense that I would like to suggest that the constitutional impossibility to develop an other-regarding conscience comes with a constitutional impossibility for the development of a dialectical self that allows for reflective disapproval of actions that negatively affect other practical agents. I assume that psychopaths generally do not have an intra-personal relationship of communicating with oneself that allows for significant evaluative gaps between two aspects of a practical agent's self: one who desires, intends and does and one who keeps a critical normative eye on what this desiring, intending and doing self desires, intends and does.

As to the abovementioned third suggestion, my tentative proposal is that due to some seeds being absent, young 'turn-out-to-be psychopaths' have not had the opportunity to develop a particular type of intra-personal relationship of communicating with oneself that allows for significant normative gaps between the two types of selves mentioned. I assume that in *typical* human agents some egoistic inclinations that are generally present in human infants are perfectly apt for becoming less dominant and even vanish as a result of being equipped with a variety of constitutional pre-dispositions, including empathic and pro-social ones. My hunch would be that the early developmental route of cases of adult psychopathy is

characterized by the empathic and pro-social aspects not developing over time resulting in a small band width of a dialectical self.

All three suggestions above are meant to be some sort of partial and *ultimate* explanation as to why it is psychopaths develop the traits and behaviour that are considered to be core to psychopathy, including the feature of not taking the interests of others seriously in practical deliberation and judgement. There seems to be a difference however between the first suggestion (on the one hand) and the second and third (on the other hand). That is, in the sense that it seems to be somewhat more sensible to think about the first suggestion in terms of psychopaths 'lacking seeds' for the feature (in this case experience of empathic distress) than for the second and third (conscience and particular dialectical self, respectively). That might well be because the conceptual suggestions under the second and third point do not organically conceptually connect with the developmental capacities of a newborn baby. For the first suggestion it is the case that one can make good attempts to conceptualize it in such a way that it fits the developmental capacities of a baby. That is not quite the case for the second and third suggestion. The capacities I point to there and about which I claim that psychopaths are missing them in their actual psychology are advanced capacities that do not map onto the developmental state of a baby. Therefore the most one might be able to say about them in terms of predispositions is that psychopaths do not have predispositions for developing them. That seems a somewhat empty claim however. It might be better not to phrase the second and third suggestion in terms of 'predispositions'. Rather one may want to regard them as an ultimate conceptual explanation of psychopaths' psychological impossibility to take account of others, an ultimate conceptual explanation that as such has (i) a causal and (ii) etiological component to it. That is, in terms of it being a conceptual suggestion that involves (i) a causal claim about how the assumed mental impossibility comes about. And this while (ii), in that context of a causal claim being made, an appeal is made to the psychological development of children who later on as adults turn out to present with psychopathy.

6.3. Psychopaths and the perception of fellow human beings

In the previous section I have offered three conceptual suggestions that can explain psychopaths' (alleged) psychological impossibility to take the interests of others into account in evaluative judgement. An assumption, as yet unstated, about all three ways is that they come with an *animate* perception of human beings. In this section I present one conceptual suggestion in regard to psychopaths' psychological impossibility to take the interests of others into account in evaluative judgement that relates to an *inanimate* perception of human beings.

In chapter 1 and above I have explained that psychopaths *use* others as 'objects of exploitation'. I have explained as well that some, if not all psychopaths *see* their used objects of exploitation as *legitimate* objects of exploitation. When it comes to psychopaths' alleged perception of other human beings as (legitimate) 'objects of exploitation', it is the case that there is an ambiguity in the compound term 'objects of exploitation'. The simple term 'object' here, could be used just in a grammatical sense, while the emphasis is on *exploitation*. I take it that this is indeed what is typically meant when some claim is made about psychopaths using/regarding others as 'objects of exploitation'. Be that as it seems to be, when it comes to psychopaths and 'objects of exploitation', it might also be the case that within the group of psychopaths there are some psychopaths who regard others as 'objects' somehow in the sense of '*mere things*'. I am inclined to think that the (current) class of psychopaths (as diagnosed by the PCL-R) incorporates a group of people who *not only use other people as objects of self-gratification and exploitation and see them as legitimate objects of these activities*.¹² Amongst the psychopaths, I am inclined to think, there is also a smaller or larger group who have a special perception of human beings. Some sort of inanimate mechanical perception. Be that my careful hypothesis, I would like to point out that currently, the psychopathy

¹² Either justified by just the view that self-gratification at whatever cost is perfectly alright and/or something like the view that those who are so stupid to let themselves be exploited deserve to be exploited.

literature does not support my hypothesis in some sort of direct way, meaning as well that there are currently no scientific accounts as to what, phenomenologically, neurally etc. it means for some psychopaths to regard others as ‘mere objects’.

Let me say a little more about the hypothesis just mentioned that concerns the perception of other human beings in an inanimate, mechanistic way. Let us first look at the following quotation.

I look at her, study her, praying to feel the life in her through the enveloping unreality. But she seems more a statue than ever, a manikin moved by mechanism, talking like an automaton. It is horrible, inhuman, grotesque. (Sechehaye, 1970: 38¹³; cited in Ratcliffe 2013, forthcoming)

This quotation comes from a person diagnosed with schizophrenia. What I would like to suggest is that amongst the current class of PCL-R psychopaths there may be some for whom the lack of access is primarily caused by something else than what I discussed in § 6.2. That is, rather than the psychological impossibility for taking the interests of others seriously being due to one or more factors mentioned in the previous section, it might primarily/most ultimately be caused by the perception of human beings as inanimate. On a slightly deeper explanatory level, my hypothesis is that amongst the class of PCL-R psychopaths there are some who have a very mechanistic self-perception. I assume that there is an inverse relationship between this and the psychological possibility for taking account of others’ interests in the following way. I assume that the perception of the other as non-mechanistic (as animate) is caused by a non-conceptual feeling. Assuming that psychopaths lack this feeling, I assume they lack an automatic means to connect with another human being. That could make them regard others as objects in a ‘thing’ –sense of the term.

¹³ Sechehaye, M. (1970). *Autobiography of a Schizophrenic Girl*. New York: Signet.

7. Intermezzo. FUL: not completely empirically free?

In § 2, I have focused on a specific application variant of Kant's *FUL*: *CCT*. I have argued that Kant seems to succeed in keeping *CCT* impression –and thereby empirical-free, while at least and probably solely outside the context of Kant's philosophy, it seems to come at the cost of *FUL* not being a moral justification at all. In this section I approach *FUL* differently. That is, I focus on the concepts of universalizability/universality that characterize *FUL*. My purpose in this section is to suggest that despite what Kant intended, there may be an empirical element to *FUL*, although –as far as I can see- not to *CCT* as an application variant of *FUL*. So, my suggestion will be that there may be an empirical element to *FUL* as a supreme principle of morality that has *CCT* and *CWT* as application variants, while there is no empirical element to *CCT* narrowly construed (leaving out any thoughts about implicit Kantian ideas about Vernunft as an essential aspect of the self as discussed in § 2.2).

An essential aspect of *FUL* as a principle with just 'form' is the idea of universalizability. While any moral justification seems to be in need of a *scope of addressees and beneficiaries*, *FUL* with its concept of universalizability essentially, by implication, appeals to a scope (while –from a non-exegetical perspective- this need not necessarily be a cosmopolitan one, which it however is for Kant). On basis of what I have argued for in regard to 'pre-intentional emotions' in this chapter, in regard to *FUL*, I would want to make the following claims.

First, a psychological decision who is and who is not included in the universal range of *FUL* could differ in case of rock-bottom radically different background states of agency. Secondly, this seems to mean that a theoretical decision as to who is and who is not included in the universal range of *FUL* could be dependent on background states of agency. Let me explain this.

For Kant it is a capacity for Vernunft that is a criterion for determining who is in the ballpark of *FUL* and who is not. If you have such a capacity, you are in. If not, you are out.

Having said that, what should we say from an exegetical point of view about how that decision is taken? There seem to be two plausible exegetical perspectives. Either *Vernunft alone* in the form of an intellectual power of the soul can take the decision that all those with *Vernunft* are in the ballpark of *FUL*. Alternatively it means that the decision can be taken completely *externally* while the decision is that all those who are in the ballpark of *FUL* are those with *Vernunft*.

Evaluating both from a non-exegetical perspective, I do not consider it plausible that intellect alone can do the trick. Let us recall that psychopaths have rather intact intellectual faculties. For the remaining I cannot think of any other evidence for such a hypothesis.

How about a full-blown externalist claim? My inclination would be to say that the theoretical decision as to who is in the ballpark and who is not cannot be taken completely externally. Decisions need to be made by minds and minds take some perspective. The decision will come from some perspective –there is no view from nowhere. Within the framework of a particular background state of agency theorists will get a long way to externalist decision making. They can point to certain descriptive features of human beings that function as criteria for the ballpark. However, I think the fact remains that there is nothing in the world that can turn features of flesh and blood human beings theorists can descriptively pick out into normative claims about the range and nature of the ballpark. Assuming for the sake of the argument that there is a biological class of human beings that are part of an externally real world, I would suggest that there is nothing in the world that compels us to classify them similarly in the moral domain. Perceptual classification of them as fellow beings I assume is contingent and extremely occasionally someone biologically classed as a human being may carve up the world fundamentally differently from the typical case. Rock-bottom, background states of (embodied) minds might determine which descriptive features present themselves as normatively relevant in regard to the decision who is in the ballpark and who is not.

8. A shared argumentative building block for a second and third criterion

In the beginning of this chapter I have focused on Hume's philosophy of reason. I have done so to end up with a specific (fully developed) criterion for there being a non-realist component to moral objectivity. The argument as it arises from Hume's philosophy of reason is built on two fundamental assumptions. First, the assumption that any theoretical decision as to which actions can be identified as good and which ones as bad must necessarily rely on a (subjective) *judgements* made by some mind(s) that necessarily occupies/occupy a perspective. The second assumption on which the argument relies is that there is an important difference between the mental process of reasoning (on the one hand) and the mental act of judging which results in mental product of judgement (on the other hand).

Relying on these two assumptions, the argument then holds that without there being an object of mattering, there is no way in which a mind capable of the mental process of reasoning can get to the psychological state of holding a judgemental outcome. A mind capable of reasoning but without having an object of mattering would be clueless as to what to judge, not knowing which direction to go in terms of making a decision.

In this final chapter, I have also recalled and conceptually built on the science of psychopathy. Furthermore, I have focused on research from Matthew Ratcliffe. The combination of this delivers a comprehensive building consisting of two ideas for two possible further criteria for there being a non-realist component to moral objectivity. These two possible further criteria I will conceptually develop in the general conclusion. The two ideas that together constitute the comprehensive building block are:

- (a) Within the class of human practical agents there is a class of agents who do not take the interests of others seriously in practical deliberation and judgement.

(b) Without there being certain elements in her constitutional make-up, elements given for free by the author of nature, (whoever or whatever that is) a human practical agent will not be able to have psychological experiences that include sincere moral reasoning and moral decision-making.

Note that while the first criterion presented is an independent criterion, there is a relationship that can be drawn between the building block above and the first criterion. While the first criterion holds that there is an *open judgemental end* for a mind with just intellectual reasoning activities (assuming that any theoretical decisions must reflect a *judgement*), the comprehensive building block/psychopathy research suggests that *judgemental outcomes can be essentially anti-moral*. Thereby it supports my Hume-based suggestion that a deliberative¹⁴ intellect alone cannot do the trick in arriving at theoretical judgements as reflecting genuinely better and worse courses of practical action. Having said that, the first criterion based in Hume's philosophy of reason consists in a necessarily *open ended outcome* for minds with an exclusive intellectual capacity for inferential reasoning; it is not about the possibility of anti-moral judgemental outcomes.

¹⁴ As to 'intuitions', I believe that affect contributes to all moral intuitions (albeit in different ways and strengths dependent on the idiosyncratic psychological nature of the agent and typically also the sort of event/action about which an intuitive judgement is made (some propositional content –e.g. strangling innocent babies- will in most human cases push emotional buttons more strongly and in different ways than other sorts of propositional content (e.g. ones about the goodness of promise-keeping).

In closing

[A]ntirealists, ...stress the dependence of the ordinary world on us, our minds and categories. (Blackburn, 1988: 371)

I cannot see how to refute the arguments for the subjectivity of ethical values, but I find myself incapable of believing that all that is wrong with wanton cruelty is that I don't like it. (Russell, *RoE*: 165¹/*Papers* 11: 310-11²)

Blackburn is right. Anti-realists stress the dependence of the ordinary world on our minds. However, at the same time, many anti-realists, including Blackburn himself, want to depart from the extreme relativist view that anything goes in the moral domain. Said differently, Blackburn and some of his anti-realist colleagues want to defend the existence of moral objectivity; they want to defend that somehow there are genuinely better and worse courses of action within the domain of morality.

The challenge for anti-realists as Blackburn, i.e. anti-realists who are sympathetic to the existence of a certain genuine form of moral objectivity, is to find:

- 1) a theoretical criterion by means of which better and worse courses of action can be identified, without...
- 2) this criterion reflecting forms of relativism that are incompatible with their commitment to the existence of moral objectivity³ and while...

¹ Russell, B. *Russell on Ethics*. [anthology of Russell's writings] (1999). Charles Pigden (ed.). London: Routledge; referenced in Pigden, 2008.

² Russell, B. *Last Philosophical Testament, 1943-68, vol.11* [collected papers]. London, New York: Routledge, 1997; referenced in Pigden, 2008.

³ It is easy to get confused when talking about relativism in ethics. Below I make some distinctions that sometimes get blurred.

- 3) upholding their essential anti-realists commitment that morality is somehow dependent on human minds rather than morality being built into or otherwise essentially part of the fabric of the world.

That is a real challenge and one very much worth philosophizing about. It is also a type of challenge however I have *not* been concerned with in this study.

Above I have talked about a challenge for a specific group of anti-realists: those who want to defend the existence of some form of moral objectivity. More generally, the intellectual challenge for every anti-realist is either to (a) justify that really anything goes in the moral domain or else to (b) find a theoretical criterion by means of which a form of moral objectivity can be identified, a form that is compatible with the anti-realist commitment that morality is not part of the fabric of the world.

Thus far I have talked about anti-realists. I now turn to the intellectual challenges for moral realists.

Realists face the challenge either to (1) make plausible the view that moral objectivity (as something that reflects better and worse courses of action) is mind-*in*dependent. Or else to (2) find a convincing argument that incorporates the idea that that which counts as moral

Descriptive ethical relativism: the empirical claim that certain groups differ in their views about what is right or good (or just or virtuous).

Response-dependent metaphysical ethical relativism: what is right or good (or just or virtuous) is metaphysically dependent on certain fundamental characteristics of practical agents.

Normative ethical relativism is the claim that at least sometimes when evaluating moral actions, it is true that no genuinely better and worse courses of action can be identified. In its most radical form this form of relativism consists in the idea that anything goes in the moral domain. Normative ethical relativism can primarily make a claim about the equal normative value of actions (telling an inconvenient truth to a friend is equally good as not telling it because it would cause harm). Normative relativism can also primarily make a claim about the moral opinions of particular individuals or groups. (the moral opinions of people of this tribe in the north of Africa are no less true or false than the moral opinions of people of this tribe in the south of Africa). If so, then this kind of claim will often be mixed with another sort of normative claim about ethical truth. That is, the claim that the moral opinions of people in the north of Africa are true *for them*, while the moral opinions of people in the south of Africa are true for the people in the south.

There are difficult intellectual problems when it comes to the relationship between especially response-dependent metaphysical ethical relativism and forms of normative relativism.

objectivity is mind-dependent, but in such a way that excludes an anti-realist idea of the mind-dependent nature of moral objectivity. That is, the idea that there is an important qualitative difference between certain real things in the world (if they exist) and morality even if morality somehow reflects genuinely better and worse courses of action - a qualitative difference one must recognize when as a theorist one wants to give an accurate description of the nature of morality.

As mentioned, one way for a moral realist to express his commitment to moral objectivity in the form of there being 'moral reality', is to defend the existence of mind-independent moral objectivity. A moral realist who wants to embark on this argumentative route has a difficult task. He needs to explain how the mind-independent substantive bits of moral reality that according to the mind-independent realist exist in our world have come to occupy and can continue to occupy a place in our universe. Here then one can reasonably wonder whether a plausible explanation can ever be given; a plausible explanation without any appeal to human minds.

The mind-independent moral realist has a hard task defending his view, given that it seems rather implausible that morality is completely mind-independent. Assuming that it is rather implausible that morality is completely mind-independent, that gives reason to wonder about the intelligibility of putting effort in defending the view that morality is completely mind-independent. Having said that, from some perspective it is intelligible to make an attempt to argue for the existence of mind-independent moral objectivity. That is to say, (complete) mind-independence means that morality is free from all forms of subjectivity. That rules out some big threats. Namely, subjectivism as an essential feature of moral objectivity cannot only imply a form of response-dependent metaphysical ethical relativism (what is right or just or virtuous or good is metaphysically dependent on certain fundamental characteristics of practical agents) that is defended by many moral realists. It also runs the risk of lapsing into a form of (i) anti-realism (substantive bits of moral objectivity if they exist at all cannot qualify as somehow being essentially part of the universe) and (ii) (anti-realist)

relativism (there is no normative difference to be made between some or all courses of action) by necessarily having to appeal to a response-dependent form of subjectivity.

In this study I have wanted to explore where the limits of moral objectivity - as understood in terms of objective morality being somehow significantly anti-realist - lie. That is, if there is something anti-realist to objective morality at all. I have wanted to deliver something theoretically realistic, non-preposterous and secure by providing some anti-realist boundaries -if there are such things- within which moral objectivity might exist. Note that up until the point of anti-realist counterevidence arising I have kept the option open that there are no anti-realist limits to the nature of moral objectivity (which explains my engagement with Kant's *FUL* that on a (reflective) first impression seems to have a component of mind-independence while it also seems to be a moral principle that is strictly objective).

The outcomes of the investigation conducted in this study are three lines of reasoning in support of the view that moral objectivity, if it exists, is essentially non-realist. The first line I have presented already (at the end of chapter 5). It revolves around Hume's conception of reason. I will recall this line of reasoning shortly. The second and third line of reasoning will be laid out in the remainder of this conclusion. The second line of reasoning consists in an advanced argument that, although advanced, needs some extra empirical investigation and conceptual reflection in future research. This due to the second line of reasoning being dependent on a controversial empirical assumption that needs more empirical investigation and due various conceptual issues about possible worlds of human nature that - to make it more problematic - also argumentatively interact with controversial empirical matters about the same topic.

The third line of reasoning consists in what is best called 'an argumentative foundation for a potential criterion that identifies why morality is essentially non-realist'. This line of reasoning is in need of more development than the second line. This due to conceptual matters about constitutional luck and freedom of will that exist alongside

problems and interact with issues about possible worlds of human nature that have both conceptual and empirical aspects.

Both the second and third line of reasoning build on the following comprehensive argumentative building block provided at the end of the previous chapter.

(a) Within the class of human practical agents there is a class of agents who do not take the interests of others seriously in practical deliberation and judgement.

(b) Without there being certain elements in her constitutional make-up, elements given for free by the author of nature, (whoever or whatever that is) a human practical agent will not be able to have psychological experiences that include sincere moral reasoning and moral decision-making.

I point out that the three (potential) arguments I offer are *not meant to be exhaustive* in terms of fundamental criteria for regarding the nature of morality as non-realist rather than realist. There may be more (potential) criteria for there being a non-realist component to moral objectivity than my study identifies/provides building blocks for. I will now recall the first line of reasoning before I present the second and third line of reasoning.

The first line of reasoning has arisen out of a comprehensive investigation of Hume's conception of reason. The argument is built on two fundamental assumptions. First, the assumption that any theoretical decision as to which actions can be identified as good and which ones as bad must necessarily rely on a (subjective) *judgement* made by some mind(s) that necessarily occupies/occupy a perspective. The second assumption on which the argument relies is that there is an important difference between the mental process of reasoning (on the one hand) and the mental act of judging which results in mental product of judgement (on the other hand).

By means of a comprehensive investigation of Hume's philosophy of reason, taking into account different areas of his work, I have argued that without there being an object of mattering, there is no way in which a mind capable of the mental process of reasoning can get to the psychological state of holding a judgemental outcome. A mind capable of reasoning but without having an object of mattering would be clueless as to what to judge, not knowing which direction to go in terms of making a decision.

I have discussed the first criterion in the context of Kant's philosophy. In my study Kant's *Formula of Universal Law* proved to be an *indirect illustration* of the force of the first criterion.⁴ What the interaction between Hume's conception of reason and *FUL* shows is that if one tries to come up with a moral justification that is maximally free from empirical influences while grounded in the intellectual reasoning capacities of practical agents, one runs the risk of having no justification at all. I have argued that the *Contradiction in Conception* variant of *FUL* can respond successfully to Hume's concern about the limitations of the powers of reason. If that counts as a success, the problem seems to be that outside the context of Kant's philosophy *FUL* runs the risk of being no more than a philosophical tool that can answer Hume's concern about reason. That is, while it can do that, *FUL*, in its *CCT* variant, might fail to be any form of moral justification. Being a successful tool to answer Hume's concern is one thing. Being a moral justification is quite another.

I have just discussed the fully developed criterion for a non-realist nature of morality my study offers. As mentioned, my study offers also an argument that needs some extra reflection in future research and an argumentative building block for a potential criterion as I expect it to arise from future research that builds upon the argumentative contribution for this criterion I offer in this study.

⁴ Recall that the initial two reasons (individually as well in their combination) for a focus on *FUL* were the following ones. First, *FUL* is an object of challenge for any theory about the limits of moral objectivity because a (reflective) first impression points to *FUL* somehow having a mind-independent nature. Secondly, *FUL* seems grounded in the reasoning capacities of agents. Generally such a feature meets the requirement that any plausible account of objective morality has a connection with the natural world (this however turned out not to be the case in Kant's work, or at least not in any straightforward sense).

I now lay out the second line of reasoning. It revolves around a radical form of normative relativism about practical ways of living.

I start by recalling (as presented at the end of the previous chapter) a fundamental assumption about the kind of moral justifications contemporary moral realists and moral rationalists defend on which my argument relies. The fundamental conceptual assumption, on which my argument relies is the idea that the moral justifications of contemporary moral realists and moral rationalists are 'other-regarding' in one way or another. Note that in a footnote, I have suggested that it might even be the case that all that can rightfully be called moral objectivity is necessarily other-regarding. I also recall, once more, the shared argumentative foundation on which both the second and third line of reasoning are fundamentally built:

(a) Within the class of human practical agents there is a class of agents who do not take the interests of others seriously in practical deliberation and judgement.

(b) Without there being certain elements in her constitutional make-up, elements given for free by the author of nature, (whoever or whatever that is) a human practical agent will not be able to have psychological experiences that include sincere moral reasoning and moral decision-making.

What is implied by point (a) above is the idea that in our actual world there is radical conceptual variation as to substantive subjective decisions in regard to practical living. Psychopaths, across the board of many different life situations, make radically different practical decisions –without any consideration for others– than 'typical' human agents do. Other human beings can be justifiably exploited if and when this serves the needs of a psychopath. Or so psychopaths believe. They believe that self-gratification at other people's expense is perfectly alright, a view that is reflected by all practical decisions they make in

daily life. In psychopaths' phenomenological world there is no space for the interests of others.

In line with the scientific claim that the core of psychopathy –including a complete disregard to the interests of other human beings - is *not* primarily caused by environmental influences, in my study I have developed two conceptual suggestions that each reflect a proposal as to how it is possible to have a background state of agency that has a range of possibilities that excludes moral decision-making. The first of these two conceptual suggestions identifies three factors that on a deep etiological level explain why and how it is human beings with psychopathy lack mental access to the realm of other-regarding thought. The three factors are: (i) a lack of predispositions for the experience of empathic distress; (ii) the absence of a possibility for the development of a conscience and (iii) the absence of a possibility for the development of a particular dialectical self⁵. The second suggestion I have offered reflects the idea that some psychopaths may have an inanimate perception of human beings due to a particular perception of themselves.

As argued, in psychopaths' phenomenological world, there is no space for others. Because of this being dependent on psychopaths' constitutional make-up (or so I suggest), deliberating and believing out of line with what I have argued to be a feature of all contemporary moral realist and moral rationalist positions seems biologically determined in a handful of human agents. My assumption is that a necessary other side of the psychological coin of a full-fledged phenomenological exclusion of others in thought is full-fledged egoism. If that is true and if some or all psychopaths can be said to have a genuine justification for their egoistic practical decisions, then thereby there seems to arise a counter argument against the existence of moral reality by means of there being radically anti-moral practical

⁵ As to the lack of predispositions for the experience of empathic distress, recall Blair's empirical suggestion (chapter 1, § 9) that genetic anomalies reduce the salience of punishment information (perhaps as a result of noradrenergic disturbance). As explained in chapter 1, the suggestion Blair offers is that this noradrenergic disturbance has an effect on various aspects of amygdala function, most importantly the ability to form stimulus–punishment associations, including stimulus punishment associations related to punishment by a victim's distress (see Blair, 2006).

thought amongst the class of human practical agents; radically anti-moral thought that fundamentally depends on constitutional features, where qualitatively similar, but substantively different constitutional features constitute *moral* thought. If an appeal to the latter is featuring in a response-dependent form of moral realism, meaning that morality as a normative institution is metaphysically dependent on fundamental features of practical agent's psychology, then there seems no reason to theoretically regard the constitutional features that stand in a fundamental causal relationship with psychopaths' radically egoistic world-view as of less normative importance than the world-view of agents who think morally. From that it might follow that there is no theoretical criterion to conceive of psychopaths' worldview as a normatively less good view about practical ways of living than the world-view of agents who think morally. That on its turn would imply a fundamental form of normative relativism about practical ways of living and that speaks against the existence of mind-independent and mind-dependent forms of moral reality.

Assuming for the sake of the argument and reality that there is radically anti-moral practical thought, I assume that this radically anti-moral practical thought can be seen as reflective of a criterion for there being a non-realist component to moral objectivity only once at least some psychopaths can be said to have a *genuine rationale* for their daily practical decisions that never take the interests of others into account. This positive claim brings me to a caveat about the argument just discussed. I would want to argue that the first potential criterion will not get off the ground in case a largely empirical assumption on which it relies is false. An assumption that is controversial. The controversial assumption concerns the hypothesis that at least some psychopaths can be said to hold a genuine normative rationale for their view that there is no need to take the interests of others into account. As discussed in chapters 1 and 5, *possibly* it is the case that some of those who are currently classified as psychopaths (by means of the PCL-R) do indeed have such a genuine rationale for their view. As discussed in chapters 1 and 5, on basis of scientific research we

would have to assume that such psychopaths justify their practical decisions by claiming that those who are so stupid/weak to allow for being exploited deserve to be exploited.

When it comes to the just discussed first argumentative contribution for there being a non-realist component to moral objectivity there are argumentative issues related to possible world problems that should be investigated in future research.

Most importantly, there are issues about *metaphysically possible and metaphysically plausible* worlds that are different from our actual world. These issues are especially and perhaps exclusively of relevance in case a moral realist wants to argue that a handful of psychopathic outliers amongst the class of human practical agents should not be able to threaten the existence of moral reality. Apart from the fact that one counterexample coming from our actual world may suffice to challenge a moral realist claim about the existence of moral reality that is meant to *apply to our actual world* (that include psychopaths), there is the following issue. Within the boundaries of the current natural laws of science, it is possible to have a possible world of practical judges that is highly significantly different from the one we live in now. If somehow we were to actively wipe out a large number of virtuous moral agent from the world (alternatively for some reason they die organically from some sort of virus), while letting psychopaths actively reproduce, what is our actual world now, in all likelihood would be very different in terms of the type of practical agents it inhabits twenty years later. Future philosophical research should shed light on what sorts of possible and plausible worlds that fit within the current scientific framework affect the argument I have offered and how they do so. Does it matter, and if so how, that 20 or 45 per cent of the world's population *could have consisted of/could come to consist of* human agents with the characteristics I have ascribed to psychopaths due to some *contingencies of the world*? This is a matter about possibility and contingency. Another question to explore would be how my argument would be affected by an *actual empirical fact* of psychopathy percentage X, Y and Z in our world. In regard to both questions (and possibly others), attention should be paid to scientific research about gene-based evolutionary criminology (recall chapter 1, § 7.6) in

terms of realistically possible minimal and especially also maximal prevalence rates of psychopaths. There should be reflection on how such scientific theories deserve their place in any philosophical reflection on different possible worlds of human nature with different prevalence rates of psychopathy.

Additional to considerations about metaphysically possible worlds, (at least some initial) reflective attention should be given to considerations about logically possible worlds (and metaphysically possible worlds that are somehow out of line with the current laws of science, if such worlds can exist as metaphysically possible worlds). While attention should be given to the worlds mentioned, a wakeful eye should be kept on exotic/esoteric theorizing. Only when it serves an intellectual goal that matters in terms of understanding and solving the problem of the limits of moral objectivity should reflections on logically possible and rather exotic metaphysically possible worlds be continued.

I now turn to the third line of reasoning, or, one could say, the second way in which my study of psychopathy contributes to the debate about the nature of moral objectivity. Future research should develop this second type of argumentative contribution into a full-fledged argument. The argument as it stands has the potential to develop into a fully developed argument in favour of there being a non-realist component to moral objectivity (if it exists) by using psychopathy as a negative test case. That is, the argument under consideration relies on the idea that a good understanding of the syndrome of psychopathy in combination with knowledge about the etiology of psychopathy can be used to learn something about the way typical human practical agents make moral judgements. Such teachings about 'us', I would want to argue, can result in an argument revolving around the idea of there being some form of involuntariness in moral judgement making as a result of constitutional luck. I suggest that if (subjective token) moral judgements are fundamentally dependent on constitutional luck, this makes it plausible (though not demonstratively provable) that moral objectivity as some sort of institution consisting in/representing the best moral normativity does have a non-realist component.

As mentioned, the third line of reasoning also fundamentally arises from the following two ideas that also underlie the argument about normative relativism just discussed:

(a) Within the class of human practical agents there is a class of agents who do not take the interests of others seriously in practical deliberation and judgement.

(b) Without there being certain elements in her constitutional make-up, elements given for free by the author of nature, (whoever or whatever that is) a human practical agent will not be able to have psychological experiences that include sincere moral reasoning and moral decision-making.

Furthermore, the third line of reasoning is built on the following assumption:

*All principles, judgements, actions, traits and events in the domain of practical action that deserve to be called 'moral' and qualify as genuinely objective **take account of others**, minimally by requiring a genuine consideration of their interests in the justificatory process fundamentally grounding the objective nature of a moral principle, action, trait or event.*

Assuming the above, in my study and in relation to idea (b), recall that in line with the scientific claim that the core of psychopathy is *not* primarily caused by environmental influences, I have developed the two conceptual suggestions mentioned earlier in this conclusion (and in chapter 5) that each reflect a proposal as to how it is possible to have a background state of agency that has a range of possibilities that excludes moral decision-making. What the first conceptual suggestions as to the possibility for other-regarding thought and judgement tells us and what seems plausible about the second as well is that nature has given us some constitutional presents at birth, the absence of which would mean that there is no way in which human beings can come to take the interest seriously and

therefore no way to opt for a moral life as a reasonable practical life of living. And that is likely to mean that moral objectivity is not part of the fabric of the world, something which on its turn seems to imply some form of non-realism about morality. That is, in case the object of challenge is a mind-dependent form of realism and unless what I take to be an implausible view in the presence of an acceptance of my naturalistic conclusions is true. The implausible view I take to be that morality is somehow mind-independent or mind-dependently essentially part of the world while it is possible that no human being has epistemic access to it.

As mentioned, in the argument reflected by the third line of reasoning- psychopaths are used as a negative test case. They do the job of giving us information about what the constitution and psychology of typical human beings capable of making moral judgements is like. Here then it is important to note the following. Once there are/were no psychopaths in our world that does *not* seem to mean that the yet underdetermined final conclusions for the argument as should arise from future research cannot be true, whatever these conclusions, after careful philosophical research, turn out to be. Neither does an absence of psychopaths in our actual world seem to mean that those final conclusions future research is supposed to generate and that have something to do with constitutional luck cannot be reached by other means than using psychopaths as a negative test case. An absence of psychopaths in our actual world in regard to this argument only seems to mean that there disappears a possibility for philosophers (or others) to use the presence of psychopaths as a means to get an argument off the ground about something that apparently has to do with there being an element to moral decision-making a practical agent is inevitably subject to and cannot be held responsible for.

One thing that should be investigated in further research is how exactly this alleged free aid functions in an argument about the limits of moral objectivity. That is, what happens to the argument if there is such a thing as a practical agent's *power to overrule* what an agent gets for free at birth? This question for future research should be split into at least the

following questions: (a) ‘What is the argumentative relevance of *having a (possibly latent) power of the mind* to overrule nature’s help?’; (b) ‘What is the argumentative relevance of an agent *making use of this power*/what is the argumentative relevance of him *actually overruling* nature’s help? (in situations X, Y and/or Z)’ ; (c) ‘What is the relevance of an *involuntary absence of a willingness* to overrule, while an actual power to overrule is present?’

Thus far I have pointed to complexities having to do with the relationship between constitutional luck and reflective control over freely given elements. This is an issue that concerns necessity, contingency and possibility for change of a dominant if not fixed influential factor on one’s moral decisions-making. There are further complexities in regard to issues having to do with necessity, contingency and ‘possibilities for change’, issues that have a relationship with technological development. In a future possible world, genetic equipment and/or actual psychological capacities may be an object of technological interference in the sense that some technique might be able to add, change and/or eliminate factors that now naturally/organically contribute to practical evaluative decision-making in typical human agents. Certain genetic, hormonal or neural factors that affect normative decision-making and that cannot be temporarily or permanently changed now, might be a realistically possible object of change at some point in future due to technological development. Think for example about the development of a pill by a group of scientists that somehow influences our inclinations for moral decision-making. A pill practical agents can voluntarily decide or refuse to take in a similar way as they decide whether to take a paracetamol when they have some sort of physical pain. For ethical reasons as well as for reasons of technological complexity it is not so likely that within the next five decades there will be a pill on the market that can bring about *radical* changes in moral decision-making of the type ‘the psycho-phenomenological world of psychopaths can change into one of typical

moral beings and/or the other way around'.⁶ Having said that, scientists are looking into the possibility of implanting chips in psychopaths' brains that, if successful, would transform their mental life rather radically.

I have just talked about (relatively nearby and at the same time still quite distant) possible worlds where there are possibilities to change practical decisions due to technological developments. While this necessarily is, or at least can reasonable considered as an issue about other possible worlds than our actual one, the starting point here has still been that the judges we are concerned with are *human* beings operating within certain neuro-bio-psychological scientific laws. Now, it might be the case that the argumentative problems for completion of this second argument increase by invoking considerations about non-human practical judges, either as some sort of actual metaphysical possibility or as a merely logical possibility. Perhaps there is reason to invoke such beings in one way or another when thinking about the second criterion as it arises from research about psychopathy. And perhaps some of these considerations justifiably problematize the argument after a longer time of philosophical reflection. Perhaps (leaving aside an initial call upon them) neither of them is at issue, which would probably mean that when debating the issue of the limits of moral objectivity, exotic theorizing is of no help to understand and/or solve the problem and perhaps even destructive to understanding and/or solving the problem of moral objectivity.

The only real moral judges we currently know of are human judges. But right now or in future there might be judges of another type –on some distant planet for example- whose

⁶ However, less drastic forms of practical engineering issues are not unlikely to be somehow of influence within the next five decades. Think for example about the availability of oxytocin⁶ pills that we can take after a day of hard work. Oxytocin pills that make it more easy not to shout at our beloved three-year old because of stress-related irritation caused by hindering lego all over the floor or to be angry with our beloved 16-year old, because of him having left his muddy trainers on the white carpet.

Many people have a gut response of aversion to such forms of engineering (including me), but it does not seem to particularly easy to get this gut responses justified. If chemical things like paracetamol are justified for a headache, why would an oxytocin pill not be justified after a day of hard work? What are the normatively relevant differences between holding a baby as justified as a means to increase our oxytocin levels, and taking an oxytocin pill if it is claimed that the former is allowed as a means to a relaxed state of mind and the latter is not?

psychology is somehow wildly dissimilar to the psychology of human beings and/or does not operate within the natural laws of say, neuroscience. What would the existence of such creatures mean for an argument about the limits of moral objectivity as grounded in conclusions about freedom of evaluative judgement? Perhaps even their logical possibility means something. Does it? Also in regard to this third line of reasoning I would want to argue that while at least at an initial stage consideration should be given to the worlds mentioned, a wakeful eye should be kept on exotic/esoteric theorizing. Only when it serves an intellectual goal that matters in terms of understanding and/or solving the problem of the limits of moral objectivity should reflections on logically possible and rather exotic metaphysically possible worlds be continued.⁷ When it comes to the usefulness of esoteric theorizing, my hunch is that the relevance of it depends on the scope of the realist claim about moral reality in terms of our actual world, nearby possibly world (specified in way X and or Y etc.) and worlds further away. That scope will dictate the kind of counterexamples as consisting in certain practical agents with a certain constitutional make-up or psychology that are relevant; counterexamples that dependent on the scope are or are not allowed to appeal to features of human constitution and psychology that are out of line with what we know about them currently from our scientific inquiries, phenomenological introspection or other methods relevant and applicable to our actual world.

⁷Kant seems to have believed that there is intelligent life on other planets. Louden argues:

“In the *Critique of Pure Reason*, Kant states confidently that he is “ready to bet everything [alles]” (A 825/B 853) he has in defense of the proposition that intelligent life does exist on other planets, and in his early work *Universal Natural History and Theory of the Heavens* (1755) he announces that “most of the planets are certainly inhabited [gewis bewohnt]” (1: 354) and that “human nature . . . occupies exactly the middle rung” on the ladder between “the most sublime classes of rational creatures,” who inhabit Jupiter and Saturn, and the less intelligent ones, who live on Venus and Mercury (1: 359). So it is clear that Kant, like “many eminent philosophers—among others Aristotle, Nicolas of Cusa, Giordano Bruno, Gassendi, Locke, Lambert, . . . and William Whewell—believed that there is extraterrestrial life.”[fn omitted] But in his more empirically sober anthropological writings he acknowledges that we have no reliable evidence for this claim. Nevertheless, the fact that Kant clearly does believe in intelligent extraterrestrial life also indicates that he does not subscribe to “the fantasy of human exceptionalism,”[fn omitted] a fantasy allegedly fueled by our own narcissism. Kant is not in humanist despair over giving up “the specialness of being human” [fn omitted] because he does not think we humans know for sure that we are special. There may be others out there like us.” (Louden, 2011:xx)

In this conclusion, I have discussed one criterion that challenges the realist nature of moral objectivity as arising from an investigation of Hume's conception of reason. I have also presented two further argumentative contributions that challenge the realist nature of moral objectivity. I have presented one argument that arises from the study of psychopathy that needs some further reflection in future research and that centres on a radical form of normative relativism. I have also presented an argumentative building block building centering on constitutional luck and freedom that has the potential to develop into argument that demonstrates a necessary non-realist component to moral objectivity. Note that empirical/empirically informed philosophical conclusions about the etiology of psychopathy as I have used them in the second and third line of reasoning can organically enrich the first Hume-based line of reasoning. The Hume-based argument relies on the assumption that all theoretical criteria for whatever substantive form of moral objectivity are dependent on a theoretical decision. Argumentatively supported by views about background states of agency, it then argues that no theoretical decision about what qualifies as a genuinely morally objective principle, action or event can be made perspectiveless (there is no view from nowhere to echo the title of Nagel's book⁸). Nor can such a subjective decision be made without a mind having a preference which requires an object of mattering. Here then psychopathy research can organically enrich the first line of reasoning, indeed in such a way as visible in the second and third line of reasoning. What it suggests is that this object of mattering that is needed for any theoretical decision about what qualifies as right and wrong within the practical domain – at least when it is assumed that all moral objectivity is 'other-regarding' – is an object of mattering that is contingent, *at least* (there might be more grounds) as a result of the contingent constitutional genetic basis for a neural capacity to not be indifferent to the well-being of others.

⁸ Nagel, T. (1986) *The View from Nowhere*. New York: Oxford University Press.

As mentioned, the second and third line of reasoning are based in psychopathy research. And, as I have just mentioned, the study of psychopathy as conducted in this chapter can also enrich the first line of reasoning. There is a fourth way in which my study of psychopathy contributes to the debate about moral objectivity. Below I discuss this way. This fourth way is more of an unintended, but helpful contribution to the debate about moral objectivity as a result of a thorough empirical investigation of the phenomenon of psychopathy than an intended argumentative contribution as is the case for the first two ways.

A fourth way in which a deep investigation of the phenomenon of psychopathy as conducted in chapter 1 contributes to the debate about the limits of moral objectivity is by dismissing a particular argumentative appeal often made by moral realists to defend the existence of moral reality against the anti-realist argument that there are radical differences in normative opinions practical agents hold. The argumentative appeal I am pointing to is the appeal that human beings who do not grasp what according to moral realists are the correct moral beliefs to hold, suffer from epistemic defects due a wrong upbringing or otherwise unfortunate educational/cultural environment. Here then, an investigation of psychopathy can pull realists out of their comfort zone. Their claim that people who do not have what according to the realists are the correct moral beliefs suffer from epistemic defects due to a wrong upbringing and/or otherwise bad cultural environment becomes implausible in regard to psychopaths when there is rather strong convergence of opinion amongst scientists that psychopathic behaviour and personality do not primarily arise from (an) environmental cause(s).

My contribution to the debate about the limits of moral objectivity consists in having identified one fully-fledged criterion for it, one argument that needs some further reflection in future research and one comprehensive argumentative buildings blocks for a criterion indicating that moral objectivity has a non-realist component (if it exists at all). The fully-fledged criterion is based on a comprehensive understanding of Hume's philosophy. The two

building blocks are based in psychopathy research. Let me end this study by means of one further remark about those building blocks being based in psychopathy research.

My arguments for claims about the limits of moral objectivity are partly based in (empirically informed) conclusions about psychopathy. That is, conclusions about its actual psychology, but especially also conclusions about the developmental history of traits and behaviours that empirical scientists and I take to be characteristic of psychopathy. Empirical scientists and I might be wrong about the fundamental roots of psychopathy. I might also be wrong about related conceptual conclusions (via the philosophical tool of using psychopaths as a negative test case) about the constitutional origins for moral judgements of typical human beings.⁹ If I am wrong about one or both of these two matters, this has a drastically bad effect on the plausibility of the second argument. The first argument seems less affected in the sense that an argument about there being radical normative relativism in regard to practical ways of living may not necessarily need an appeal to constitution in order to function as an argument in favour of there being an anti-realist component to moral objectivity. As to me potentially being wrong about the strong constitutional component to psychopathy and in regard to what I have discussed in the previous paragraph, note that once psychopathy does not have the strong constitutional component I have assumed it to have, the doors open again for the moral realist to argue that psychopaths suffer from epistemic defects due to a bad educational and/or cultural environment.

In this study, I have focused a lot on issues having to do with human nature. I have done so because I believe it should become an integral component of the study of moral objectivity. I expect that integrating human nature issues in the philosophical study of moral objectivity at times is like facing a can of rather unattractive controversial and difficult empirical-cum-conceptual worms in regard to which one may be quickly inclined to look for

⁹ Research by Paul Bloom and colleagues however points to innate senses of justice (and prosocial behaviour) in babies (see Bloom 2013, 2010a, 2010b; Hamlyn, Wynn & Bloom, (2007); Hamlyn, Wynn & Bloom (2010).

the lit. That however should not prevent us from opening the can. Opening it seems much better than collective meta-ethical avoidance of issues that are no doubt difficult, but of crucial importance to both a sufficiently deep understanding and a sufficiently good solution to the problem of moral objectivity, including the problem of its limits. I hope that individually and collectively we will take Hume's attitude to heart:

Human Nature is the only science of man; and yet has been hitherto the most neglected. 'T will be sufficient for me, if I can bring it into a little more fashion. (Hume, T: 273)¹⁰

And it would be great if I myself in this study have managed to bring the science of man into a little more fashion.¹¹

¹⁰ Compare Kant. As Loudon states, Kant argues in three different texts that the question "What is the human being?" is the most fundamental question in philosophy, one that encompasses all others (*Logik* 9: 25; cf. letter to Stäudlin of May 4, 1793, 11: 429; *Politz* 28: 533–34; cited in Loudon, 2011: xvii).

¹¹ I take it that (empirically informed?) reflection on the nature (and actual?) existence of malicious beings is also of great relevance to the contemporary debate about moral projectivism. What it is the (actual? some sort of possible?) existence of Hume's malicious beings or people like psychopaths who think in a radical anti-moral way show us? Does their (actual? possible?) existence tell that these malicious human creatures are blind to some authoritative bits of moral normativity that can be said to be really there in the world (let us call these bits 'facts' while leaving open their metaphysical robustness)? Or else, is it rather the case that they demonstrate the projectivist consequence of a relational philosophy of moral distinctions by providing an example of a single thing, a moral feature, which can be judged good and bad by a human being as a result of different sensibilities without this implying a normative defect? (see Sainsbury, 1998).

Bibliography

Full referencing information about Kant and Hume's writings as I have used them in this study can be found in 'Notes to Kant and Hume's work: references and translations' in the very beginning of this study. All other bibliographical information (including references to secondary literature about Kant's and Hume's work) can be found in the bibliography below.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders. (fourth edition-text revision; DSM-IV-TR)*. Arlington: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (fifth edition; DSM-V)*. Arlington: American Psychiatric Association.

Anscombe, G.E.M. (1957). *Intention*. Oxford: Basil Blackwell.

Atkinson, R. F. (1960). 'Hume on mathematics'. *The Philosophical Quarterly*, 10 (39), pp. 127-137.

Babiak, P. & Hare, R.D. (2006). *Snakes In Suits. When Psychopaths go to Work*. New York: HarperCollins Publishers.

Babiak, P., Neumann, C.S., Hare, R.D. (2010). 'Corporate psychopathy: talking the walk'. *Behavioural Sciences and the Law*, 28, pp.174-193.

Barr, K.N & Quinsey, V.L. (2008). 'Is psychopathy pathology or a life strategy?: implications for social policy'. In Crawford, C. & Salmon, C. (eds.) *Evolutionary Psychology, Public Policy and Personal Decisions* (Taylor & Francis e-library [originally 2004, New Jersey: Lawrence Erlbaum Associates]), pp. 273-296.

Bedke, M. (2009). 'Moral judgment purposivism: saving internalism from amorality'. *Philosophical Studies*, 144, pp. 189-209.

Bedke, M. (unpublished manuscript). 'A case for besires'.

http://www.u.arizona.edu/~mbedke/Bedke/Some_Papers_files/Besires.pdf

- Belshaw, C. (1989). 'Hume and demonstrative knowledge'. *Hume Studies*, XV (1), pp. 141-162.
- Blackburn, S. (1988). 'How to be an ethical anti-realist'. *Midwest Studies in Philosophy*, 12, pp. 361-375. [page references to re-print in: Moser & Trout 2005].
- Blackburn, S. (1998). *Ruling passions*. Oxford: Oxford University Press.
- Blair, R.J.R. (1995). 'A cognitive developmental approach to morality: investigating the psychopath'. *Cognition*, 57, pp. 1-29.
- Blair, R.J.R. (2003). 'Neurobiological basis of psychopathy'. *The British Journal of Psychiatry*, 182, pp. 5-7.
- Blair, R.J.R. (2005). 'Applying a cognitive neuroscience perspective to the disorder of psychopathy'. *Development and Psychopathology*, 17, pp. 865–891.
- Blair, R.J.R. (2006). 'The emergence of psychopathy'. *Cognition*, 101, pp. 414-442.
- Blair, J. Mitchell, D. & Blair, K. (2005). *The Psychopath: Emotion and the Brain*. Malden, Oxford, Carlton: Blackwell Publishing.
- Blair, R.J.R., Peschardt, K.S., Budhani, S., Mitchell, D.G.V., Pine, D.S. (2006). 'The development of psychopathy'. *Journal of Child Psychology and Psychiatry*, 47 (3/4), pp. 262-275.
- Bloom, P. (2010a). 'The moral life of babies'. *New York Times Magazine*, May.
<http://www.nytimes.com/2010/05/09/magazine/09babies-t.html>
- Bloom, P. (2010b). 'Infant origins of human kindness'. *New Scientist*. October.
- Bloom, P. (2013). *Just Babies: The Origins of Good and Evil*. New York: Crown.
- Cleckley, H. (1988). *The Mask of Sanity (fifth edition)*.
http://www.cassiopaea.org/cass/sanity_1.PdF
- Cohen, M.F. (1990). 'Obligation and human nature in Hume's philosophy'. *The Philosophical Quarterly*, 40 (160), pp. 316-341.
- Cohon, R. (1997). 'The common point of view in Hume's ethics'. *Philosophy and*

- Phenomenological Research*, 57 (4), pp. 827-850.
- Cohon, R. (2008). *Hume's Morality. Feeling and Fabrication*. New York: Oxford University Press.
- Cohon, R. (Fall 2010). 'Hume's moral philosophy'. In Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* URL = <http://plato.stanford.edu/archives/fall2010/entries/hume-moral/>.
- Cohon, R. & Owen, D. (unpublished manuscript). 'Hume on representation, reason and motivation, pp. 1-23.
<http://davidowencourses.arizona.edu/David's%20Papers/papersind.htm>
- Coid, J., Yang, M., Ullrich, S., Roberts, A., Moran, P., Bebbington, P., Brugh, T., Jenkins, R., Farrell, M., Lewis, G., Singleton, N., Hare, R. (2009). 'Psychopathy among prisoners in England and Wales'. *International Journal of Law and Psychiatry*, 32, pp. 134-141.
- Cole, S.W. (2009). 'Social regulation of human gene expression'. *Current Directions in Psychological Science*, 18 (3), pp. 132-137.
- Cooke, D.J., Michie, C. (1999). 'Psychopathy across cultures: North America and Scotland compared'. *Journal of Abnormal Psychology*, 108 (1), pp. 58-68.
- Cooke, D.J., Michie, C., Hart, S.D., Clark, D. (2005a). 'Assessing psychopathy in the UK: concerns about cross-cultural generalisability'. *British Journal of Psychiatry*, 186, pp. 335-341.
- Cooke, D.J., Michie, C., Hart, S.D., Clark, D.A. (2005b). Searching for the pan-cultural core of psychopathic personality disorder. *Personality and Individual Differences*, 39, pp. 238-259.
- Davidson, D. (1963). 'Actions, Reasons, and Causes' *The Journal of Philosophy*, 60, (23), pp. 685-700.
- Doyle, J. (2000). 'Moral rationalism and moral commitment'. *Philosophy and Phenomenological Research*, 60 (1), pp. 1-22.

- Drosdowski, G. & Grebe, P. (1963). *Der Große Duden - Etymologie. Herkunftswörterbuch der Deutschen Sprache*. Mannheim: Bibliographisches Institut.
- Dupré, J. (1998). 'Normal people', *Social Research*, 65, pp. 221-248.
- Dupré, J. (2003). 'On human nature'. *Human Affairs, Journal of the Slovakian Academy of Sciences*, 13, pp. 109-122.
- Edens, J.F. (2001). 'Misuses of the Hare Psychopathy Checklist-Revised in court: two case examples'. *Journal of Interpersonal Violence*, 16, pp. 1082-1093.
- Eisler, R. (1930). *Kant-Lexicon*. <http://www.textlog.de/kant-lexikon.html>
- Ellis, L. & Walsh, A. (1997). 'Gene-based evolutionary theories in criminology'. *Criminology*, 35 (2), pp. 229-276.
- Falkenstein, L. (1997). 'Naturalism, normativity, and scepticism in Hume's account of belief'. *Hume Studies*, XXIII (1), pp. 29-72.
- Finger, E.C., Marsh, A.A., Mitchell, D.G., Reid, M.E., Sims, C., Budhani, S., Kosson, D.S., Chen, G., Towbin, K.E., Leibenluft, E., Pine, D.S., Blair, J.R. (2008). 'Abnormal prefrontal cortex function in children with psychopathic traits during reversal learning'. *Archives of General Psychiatry*, 65 (5), pp. 586-594.
- Finlay, S. (2007). 'Four faces of moral realism'. *Philosophy Compass* 2 (6), pp. 820-849.
- Fisher, A. & Kirchin, S. (2006). 'Introduction'. In Fisher, A. & Kirchin, S.T. (eds.) *Arguing About Metaethics* (Oxon: Routledge), pp. 1-22.
- Fitzpatrick, W.J. (2009). 'Recent work on ethical realism'. *Analysis Reviews*, 69 (4), pp. 746-760.
- Flew, A. (1976). 'Three questions about justice in Hume's Treatise'. *The Philosophical Quarterly*, 26 (102), pp.1-13.
- Foot, P. (2001). *Natural Goodness*. Oxford: Oxford University Press.
- Forth, A.E. (2005). 'Hare psychopathy checklist: youth version', In Grisso, T. Vincent, G. Seagrave, D. (eds.) *Mental Health Screening and Assessment in Juvenile Justice* (New York: The Guilford Press), pp. 324-338.

- Gazzaniga, M.S. & Heatherton, T.F. (2003). *Psychological science: mind, brain and behavior*. New York: W.W. Norton & Company.
- Gorman, M.M. (1993). 'Hume's theory of belief'. *Hume Studies*, XIX (1), pp. 89-102.
- Gotterbarn, D. (1974). 'Kant, Hume and analyticity'. *Kant-Studien*, 65, pp. 274-283.
- Gunter, T.D., Vaughn, M.G., Philibert, R.A. (2010). 'Behavioral genetics in antisocial spectrum disorders and psychopathy: a review of the recent literature'. *Behavioral Sciences and the Law*, 28, pp. 148-173.
- Hamlyn, K. J., Wynn, K., Bloom, P. (2007). 'Social evaluation by preverbal infants'. *Nature*, letters, 450, 22 November.
- Hamlyn, K. J., Wynn, K., Bloom, P. (2010). 'Three-months-olds show a negativity bias in their social evaluations'. *Developmental Science*, pp. 1-7.
- Hancock, J.T, Woodworth, M.T., Porter, S. (2011). 'Hungry like the wolf: A word-pattern analysis of the language of psychopaths'. *Legal and Criminological Psychology*, 1-13.
- Hare, R.D. (1993). *Without Conscience. The Disturbing World of the Psychopaths Among Us*. New York, London: The Guilford Press.
- Hare, R.D. (1996). 'Psychopathy and antisocial personality disorder: a case of diagnostic Confusion'. *Psychiatric Times*, 13 (2), February 1.
- Hare, R. D. (2003). *Manual for the Hare Psychopathy Checklist - Revised* (2nd ed.). Toronto, ON: Multi-Health Systems.
- Hare, R.D., Harpur, T.J., Hakstian, A.R., Forth, A.E., Hart, S.D., Newman, J.P. (1990). 'The Revised Psychopathy Checklist: reliability, and factor structure'. *Psychological Assessment*, 2 (3) pp. 338-41.
- Hare, R.D. & Neumann, C.S. (2008). 'Psychopathy as a clinical and empirical construct'. *Annual Review of Clinical Psychology*, 4, pp. 217-246.
- Hare, R.D. & Neumann, C.S. (2010). 'Comment and reply. The role of antisociality in the psychopathy construct: comment on Skeem and Cooke'. *Psychological Assessment*, 22 (2), pp. 446-454.

- Harpur, T. J., Hakstian, A. R., & Hare, R. D. (1988). 'Factor structure of the Psychopathy Checklist'. *Journal of Consulting and Clinical Psychology*, 56, pp. 741-747.
- Harpur, T. J., Hare, R. D., & Hakstian, A. R. (1989). 'Two-factor conceptualization of psychopathy: construct validity and assessment implications'. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1, pp. 6-11.
- Hastings, M.E., Tangney, J.P., Stuewig, J. (2008). 'Psychopathy and identification of facial expressions of emotion'. *Personality and Individual Differences*, 44 (7), pp. 1474–1483.
- Hobson, J. & Shine, J. (1998). 'Measurement of psychopathy in a UK prison population referred for long-term psychotherapy'. *British Journal of Criminology*, 38 (3), pp. 504–515.
- Horgan, T. and Timmons, M. (2007). 'Morphological rationalism and the psychology of moral judgment'. *Ethical theory and moral Practice*, 10, pp.279–295.
- Humberstone, L. (1992). 'Direction of Fit'. *Mind*, 101 (401), pp. 59-83.
- Jacobs, K. A. & Walter, S. (2011). 'Natürlich defekt? Zweck und Natur in Moralphilosophie und Krankheitstheorie'. In Schlicht, T. (ed.) *Zweck und Natur*. (München: Fink), pp. 289-307. <http://www.philosophy-online.de/pdf/natuerlichdefekt.pdf>.
- Kail, P. J.E. (2007). *Projection and Realism in Hume's Philosophy*. Oxford: Oxford University Press.
- Karpman, B. (1948). 'Conscience in the psychopath: another version'. *American Journal of Orthopsychiatry*, 18, 455–491.
- Kiehl, K.A. (2006). 'A cognitive neuroscience perspective on psychopathy: evidence for paralimbic system dysfunction'. *Psychiatry Research*, 15, pp. 107-128.
- Kiehl, K.A., Smith, A.M., Hare, R.D., Mendrek, A., Forster, B.B., Brink, J & Liddle, P.F.

- (2001). 'Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging'. *Biological Psychiatry*, 50, pp. 677–684.
- Korsgaard, C. M. (1997). 'The general point of view: love and moral approval in Hume's ethics'. *Hume Studies*, XXV (1 & 2), pp. 3-42.
- Kreis, M.K.F. (2009). *Psychopathy in Women: A Multi-Method Exploration of the Construct Using the Comprehensive Assessment of Psychopathic Personality (CAPP)*. Doctoral dissertation.
- Kreis, M.K.F. & Cooke, D.J. (2011). 'Capturing the psychopathic female: a prototypicality analysis of the comprehensive assessment of psychopathic personality (CAPP) across gender'. *Behavioral Sciences and the Law*, 29, pp. 634–648.
- Laird, J. (1939). 'Hume's account of sensitive belief'. *Mind*, 48 (192), pp. 427-445.
- Landy, D. (2006). 'Hume's impression/idea distinction'. *Hume Studies*, 32 (1), pp. 119-139.
- Louden, R.B. (2000). *Kant's Impure Ethics. From Rational Beings to Human Beings*. Oxford: Oxford University Press.
- Louden, R.B. (2011). *Kant's Human Being. Essays on his Theory of Human Nature*. Oxford: Oxford University Press.
- Lynch, M.P. (1996). 'Hume and the limits of reason'. *Hume Studies*, XXII (1), pp. 89-104.
- Magri, T. (1996). 'Natural obligation and normative motivation in Hume's Treatise'. *Hume Studies*, XXII (2), pp. 231-254.
- Mackie, J. L., (1977). *Ethics: Inventing Right and Wrong*. London: Penguin.
- McCrae, R.R., & Costa, P.T. jr (1999). 'A five-factor theory of personality'. In John, O.P., Robins, R.W. & Pervin, L.A. (eds). *Handbook of personality: Theory and research* (2nd ed. New York: Guilford Press), pp. 139-153.
- McCrae, R.R., & Costa, P.T. jr, Ostendorf, F., Angleitner, A., Hrebickova, M., Avia, M.D.,

- et al. (2000). 'Nature over nurture: temperament, personality, and life span development'. *Journal of personality and Social Psychology*, 78, pp. 173-186.
- McDowell, J. (1995). 'Might there be external reasons'. In Altham, J.E.J. & Harrison, R. (eds.) *World, Mind, and Ethics, Essays on the Ethical Philosophy of Bernard Williams* (Cambridge: Cambridge University Press), pp. 68-85.
- McIntyre, J.L. (2000). 'Hume's passions: direct and indirect'. *Hume Studies*, XXVI (1), pp. 77-86.
- Mealey, L. (1995). 'The sociobiology of sociopathy: An integrated evolutionary model'. *Behavioral and Brain Sciences*, 18, pp. 523-541.
- Meeker, K. (2007). 'Hume on knowledge, certainty and probability: anticipating the disintegration of the analytic/synthetic divide?' *Pacific Philosophical Quarterly*, 88, pp. 226 – 242.
- Meloy, J.R. & Reavis, J.A. (2007). 'Dangerous cases: when treatment is not an option'. In: Luyn van, B. Akhtar, S. Livesley, J. (eds.) *Severe personality disorders* (New York: Cambridge University Press), pp. 181-195.
- Michaud, Y. (1987). 'Hume's naturalized philosophy'. *Hume Studies*, XIII (2), pp. 360 -380.
- Neto, J.R.M. (1991). 'Hume and Pascal: pyrrhonism vs. nature', *Hume Studies*, XVII (1), pp. 41-50.
- Miller, A. (Spring 2012). 'Realism'. In Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* URL = <<http://plato.stanford.edu/archives/spr2012/entries/realism/>>.
- Shapiro, I. (2005). 'Human nature'. In Craig, E. (ed.) *The Shorter Routledge Encyclopedia of Philosophy* (Oxon, New York: Routledge), pp. 390-95.
- Milliken, J. (2008). 'In a fitter direction: moving beyond the direction of fit picture of belief and desire'. *Ethical Theory and Moral Practice*, 11, pp. 563-571.
- Moser, P.K. & Trout, J.D. (2005). *Contemporary Materialism: A Reader*. Taylor and Francis e-library [Originally published in 1995. New York, London: Routledge].
- Murphy, J.G. (1972) 'Moral death: a Kantian essay on psychopathy'. *Ethics*, 82, (4), pp.

284-298.

- Murphy, J.M. (1976). 'Psychiatric labelling in cross-cultural perspective'. *Science*, 141 (4231), pp. 1019-1028.
- Nagel, T. (1999). 'One-to-one' [Book review T.M. Scanlon's *What we Owe to Each other*]. *London Review of Books*, 21 (3/4), pp. 3-6.
- Neumann, C.S, Hare, R.D., Newman, C.P. (2007). 'The super-ordinate structure of the Psychopathy Checklist-Revised'. *Journal of Personality Disorders*, 21, pp. 102-107.
- Nichols, S. (2002). 'How psychopaths threaten moral rationalism: is it irrational to be amoral?'. *Monist*, 85 (2), pp. 285-303.
- Nuyen, A.T. (1984). 'David Hume on reason, passions and morals'. *Hume Studies*, X (1), pp. 26-45.
- Nuyen, A.T. (1988). 'The role of reason in Hume's theory of belief'. *Hume Studies*, XIV (2), pp. 372-389.
- Oakley, B. (2008). *Evil genes: Why Rome Fell, Hitler Rose, Enron Failed, and my Sister Stole my Mother's Boyfriend*. New York: Prometheus Books.
- O'Neill, O. (2004). 'Rationality as practical reason'. In Mele, A.R. & Rawling, P. (eds.) *The Oxford Handbook of Rationality* (New York: Oxford University Press), pp. 93-109.
- Owen, D. (1992). 'Hume and the Lockean background: induction and the uniformity principle'. *Hume Studies*, XVIII (2), pp. 199-208.
- Owen, D. (1994). 'Reason, reflections, and reductios'. *Hume Studies*, XX (2), pp. 195-210.
- Paulhus, D.L. & Williams, K.M. (2002). 'The dark triad of personality: narcissism Machiavellianism, and psychopathy'. *Journal of research in personality*, 36, pp. 556-563.
- Persson, I. (1997). 'Hume - not a Humean about motivation'. *History of Philosophy Quarterly*, 14 (2), pp. 189-206.
- Pettit, P. (1987). 'Humeans, anti-Humeans, and motivation'. *Mind*, 96, (384), pp. 530-533.

- Pettit, P. (1998) 'Desire'. In Craig, E. (ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge. Retrieved June 09, 2012, from <http://www.rep.routledge.com/article/V009>
- Platt, M. (1979). *Ways of meaning*, London: Routledge and Kegan Paul.
- Ratcliffe, M. (2005). 'The feeling of being'. *Journal of Consciousness Studies*, 12 (8-10), pp. 45-63.
- Ratcliffe, M. (2008). *Feelings of Being: Phenomenology, Psychiatry and the Sense of Reality*. Oxford: Oxford University Press.
- Ratcliffe, M. (2010). 'Depression, guilt and emotional depth'. *Inquiry* 53, pp. 602-626.
- Ratcliffe, M. (2012a). 'The phenomenology of existential feeling'. In Marienberg, J. & Fingerhut, S. (eds.) *The Feeling of Being Alive* (Berlin: de Gruyter), pp. 23-54.
- Ratcliffe, M. (2012b). 'Phenomenology as a form of empathy'. *Inquiry*, 55 (5), pp. 473-495.
- Ratcliffe, M. (forthcoming 2013). 'The structure of interpersonal experience'. In Jensen, R.T. & Moran, D. (eds.) *Phenomenology of Embodied Subjectivity* (Dordrecht: Springer).
- Rooijen van, M. (2010). 'Moral rationalism and rational amorality'. *Ethics*, 120 (3), pp. 495-525.
- Rosati, C.S. (Fall 2008). 'Moral Motivation'. In Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* URL = <<http://plato.stanford.edu/archives/fall2008/entries/moral-motivation/>>.
- Russell, B. (Spring 2010), 'A Priori Justification and Knowledge'. In Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* URL =<
<<http://plato.stanford.edu/archives/spr2010/entries/apriori/>>.
- Sainsbury, R.M. (1998). 'Projections and relations'. *Monist*, 81, pp. 133-160.
- Scanlon, T. (1998). *What We Owe to Each Other*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Schroeder, T. (Winter 2009). 'Desire'. In Zalta, E.N. (ed.) *The Stanford Encyclopedia of*

- Philosophy*, URL = <<http://plato.stanford.edu/archives/win2009/entries/desire/>>.
- Schroeder, M. (2010). *Noncognitivism In Ethics*. Oxon: Routledge.
- Searle, J. R. (1979). *Expression and Meaning*. Cambridge: Cambridge University Press.
- Searle, J.R. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Shafer-Landau, R. (2003). *Moral Realism-A defense*. New York: Oxford University Press.
- Shafer- Landau, R. (2006). 'Rationalism in Ethics' [addendum]. In Borchert, D.M. (ed.) *Encyclopedia of Philosophy (2nd edition)*, (Farmington Hills: Thomson Gale), vol. 8, pp. 251-253.
- Smith, M. (1987), 'The Humean theory of motivation', *Mind*, 96 (381), pp. 36-61.
- Smith, M. (1988). 'On Humeans, anti-Humeans, and motivation: a reply to Pettit', *Mind*, 97 (388), pp. 589-595.
- Smith, M. (1991), 'Moral Realism'. In Singer, P. (ed.) *A Companion to Ethics* (Blackwell: Oxford), pp. 399-410.
- Smith, Michael (1994). *The Moral Problem*. Oxford: Blackwell Publishers.
- Stich, S & Ravenscroft, I. (1994). 'What is folk psychology?' *Cognition*, 50, pp. 447-468.
- Steiner, M. (1987). 'Kant's misrepresentations of Hume's philosophy of mathematics in the Prolegomena'. *Hume Studies*, XIII (2), pp. 400 -410.
- Sullivan, R.J. (1989). *Immanuel Kant's Moral Theory*. Cambridge: Cambridge University Press.
- Taylor, J. (1998). 'Justice and the foundations of social morality in Hume's Treatise'. *Hume Studies*, XXIV (1), pp. 5-30.
- Vernon, P, Villani V.C., Vickers, L. Aitken Harris, J. (2007). 'A Behavioral genetic investigation of the dark triad and the big 5'. *Personality and Individual Differences*, 44, pp. 445-452.
- Voorhoeve, A. (2009). *Conversations on Ethics*. Oxford: Oxford University Press.
- Vitale, J.E. & Newman, J.P. (2001). 'Using the Psychopathy Checklist-Revised with

- female samples: reliability, validity, and implications for clinical utility.' *Clinical Psychology: Science and Practice*, 8 (1) pp. 117-132.
- Watkins Tate, M. (2005) 'Obligation, justice, and the will in Hume's moral philosophy'. *Hume Studies*, 31 (1), pp. 93-122.
- Weber, S., Habel, U., Amunts, K., Schneider, F. (2008). 'Structural brain abnormalities in psychopaths—a review'. *Behavioral Sciences and the Law*, 26, pp. 7–28.
- Westphal, K. (2010). 'From 'convention' to 'ethical life': Hume's theory of justice in post-Kantian perspective'. *Journal of Moral Philosophy*, 7, pp. 105–132.
- Whelan, F.G. (1994). 'Hume and contractarianism'. *Polity*, 27 (2), pp. 201-224.
- Williams, B. (1981). 'Internal and External reasons', In: *Moral Luck*, Cambridge: Cambridge University Press.
- Williams, B. (1995a). 'Internal reasons and the obscurity of blame'. In *Making Sense of Humanity and other Philosophical Papers*, Cambridge: Cambridge University Press, pp. 35-45.
- Williams, B. (1995b). 'Replies'. In Altham, J.E.J. & Harrison, R. (eds.) *World, Mind, and Ethics, Essays on the Ethical Philosophy of Bernard Williams* (Cambridge: Cambridge University Press), pp. 68-85.
- Williams, B. (2001). 'Postscript: some further notes on internal and external reasons. In Zangwill, N. (2008), 'Besires and the Motivation Debate'. *Theoria*, 74, pp. 50-59.